J Math Model Algor (2007) 6:47–75
DOI 10.1007/s10852-006-9049-y

Optimal Solutions in the Multi-location Inventory System with Transshipments

Lars Magne Nonås · Kurt Jörnsten

Received: 3 October 2004 / Accepted: 7 October 2005 /
Published online: 27 October 2006

Abstract We consider a single-period multi-location inventory system where inventory choices at each location are centrally coordinated. Transshipments are allowed as recourse actions in order to reduce the cost of shortage or surplus inventory after demands are realized. This problem has not been solved to optimality before for more than two locations with general cost parameters. In this paper we present a simple and intuitive model that enables us to characterize optimal inventory and transshipment policies for three and four locations as well. The insight gained from these analytical results leads us to examine the optimality conditions of a greedy transshipment policy. We show that this policy will be optimal for two and three locations. For the n location model we characterize the necessary and sufficient conditions on the cost structure for which the greedy transshipment policy will be optimal.

Key words multi-location inventory system · transshipment · cost parameter.

Mathematics Subject Classifications (2000) 90B06 · 90B15

1 Introduction

In the last couple of decades, the number of products offered to the market has generally exploded. At the same time, the product life-time has decreased drastically. The combination of these two trends leads to increased inaccuracy of the demand forecasts, leading to firms facing an increased demand uncertainty. Further, as a response to higher pressure on cost reductions, firms tend to source more from low cost countries in regions such as the Far East, resulting in longer lead times. An effect

L. M. Nonås (✉) · K. Jörnsten
Department of Finance and Management Science, Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway
e-mail: Lars.Nonas@nhh.no

of this is that firms are less responsive to the demand uncertainty. Correspondingly, one of the major challenges in many industries is making supply meet demand (see [6]). Several strategies and initiatives to achieve this have gained increasing popularity with firms. This paper studies one such strategy, namely transshipments.

Consider two retailers that are both owned and operated by the same firm. A central planner decides the inventory of both retailers before the start of the season. Due to long lead time the planner does not have the opportunity to replenish additional inventory during the season. Consider the case when the supply of one retailer exceeds its demand, and, conversely, the demand of the other retailer exceeds its supply. Then, if the benefit exceeds the incurred costs of doing so, it would be beneficial, from a system point of view, to have the former retailer transfer (some of) its inventory to the latter retailer. This practice is called *transshipment*, and is routinely performed in a variety of industries. We have recently witnessed an increasing use of transshipment, mainly as a result of better integration of the information systems of the retailers participating in a distribution network. Also, the availability of faster, flexible and more reliable freight-providers such as DHL and UPS has facilitated this development.

The decision of transshipment, in terms of how much to transship between what locations at what time, is made after the firm has acquired improved demand information. What makes the decision problem inherently difficult is that the practice of transshipment strongly affects the optimal inventory quantities decided upon prior to the season. As an illustration, consider the following example. A firm offering a high service level determines inventory levels at each retailer without taking the possibility of transshipment into account. The resulting inventory levels for each location will be relatively high to ensure that the supply at each location with high probability will meet the demand at the corresponding locations. While this, in expectation, will lead to a large number of units available for transshipment (i.e., a majority of the retailers will have excessive supply), it will be unlikely that there is a large need for transshipment (i.e., few retailers will have insufficient supply). Hence, to reap the full benefit of transshipment the firm will need to take the transshipment option into account when making its *quantity* decisions.

1.1 Literature Review

The most commonly used model when analyzing transshipment in connection with ordering quantities in inventory models is the classical newsvendor model. All the papers mentioned below are based on the newsvendor concept. The optimal solution in a newsvendor model balances the expected cost of understocking against the expected cost of overstocking. For a detailed treatment of the newsvendor model see Porteus [16]. Among the earliest published works on transshipment as a cost reducing policy yielding better customer service is Krishnan and Rao [14]. They look at the case of identical cost parameters and identical independent demands. Robinson [17] generalizes this to general cost parameters and demand distributions. Krishnan and Rao [14] and Robinson [17] assume that transshipments take place after the true demand has been realized,but before customer orders has to be fulfilled. Some papers ([12, 13]) assume that transshipments take place before demand is fully realized, leading to models in which demand is realized only partially. There are also papers which assume that transshipments occur routinely instead of when a stockout is imminent ([4, 11]). We will employ the scenario given by Krishnan and Rao [14] and Robinson [17]. While some papers also consider a multi-period transshipment model, these models are often basically reduced to a set of single period problems, i.e., a myopic order-up-to policy is optimal ([9, 17]).

The literature on transshipments has generally either been concerned with analytical results for the two location model ([18, 19]) or heuristics for the n location model ([13, 17]). Tagaras [19] defines a set of assumptions that when satisfied means that complete pooling is an optimal strategy (see Section 2.1) in the two location model. Similar assumptions are often used in the literature ([7, 17]) and will be used in our model. While both Rudi et al. [18] and Robinson [17] provide analytical solutions for the two location model, Robinson [17] also suggests a heuristic for the n location model. By discretizing the demand distribution Robinson [17] sets up a large linear programming problem to approximate the optimal solution. However, the size of this linear program is an obstacle with respect to both memory problems and solution time. For a similar problem Tayur [20] used a gradient based approach that is based on discretizing the demand distribution for an n location model. Tayur uses a nested reoptimization technique for approximation of the optimal solution which works well for medium sized problems. Most of the literature mentioned above considers models where all decisions are centralized in a "parent-firm." Dong and Rudi [5] and Rudi et al. [18] study models with a more decentralized decision structure.

The contribution of this paper is twofold. Firstly, we formulate a model that is simpler and more intuitive than the model of Robinson [17] although it incorporates all of the complexities of that model. Secondly, due to the simplicity of the model we are able to gain analytical insight into problems of higher dimensions than has been achieved earlier, i.e., problems with more than two locations and general cost parameters. We show how to characterize the optimal order quantities for problems with three and four locations. For two and three locations we prove that a greedy transshipment policy is optimal always. For the four location model we characterize, in an exhaustive way, the conditions for when a greedy transshipment policy is optimal, depending on the cost structure as well as the order quantity and demand. The corresponding conditions for the cost structure alone has also been characterized for the n location model.

1.2 Organization of Paper

This paper is organized as follows: in Section 2 we formulate the transshipment problem as a two stage stochastic program where the second stage is modeled as a classical Hitchcock–Koopmans transportation problem, hence a specially structured linear programming problem. Utilizing the structure of the dual of this linear programming problem, we are able to determine the gradient for the first stage objective function. In Section 3 we study the simplest version of the problem where transshipments can be of use, that is the two location model. Section 4 contains the analysis for the case with three locations. The optimal transshipment policy is formulated, and the optimal order quantities are characterized. In Section 5 we analyze the case with four locations. It is shown why the complexity of the problem increases significantly as the number of locations increases. The main reason is that the optimal allocation of transshipments can no longer be determined using a greedy algorithm. We also formulate the optimal ex-post transshipment policy, which enables us to characterize the optimal order quantities. In Section 6 we consider the n location model. We show that a greedy allocation of transshipments will be an optimal transshipment policy for a specific cost structure. Finally, in Sections 7 and 8, we give some ideas for future directions for research on transshipment problems and summarize the results obtained in this paper.

2 Model Formulation

Consider the following real life problem where we have n stores selling a seasonal product. Before the season starts and long before the stores know much about the future demand, store i has to order large quantities, Q_i, of the product in order for the store to be able to meet the future unknown demand, D_i. We assume that the joint distribution of demand is known and continuous.

Store i sells at unit revenue cost r_i. The stores procure the product at unit ordering cost c_i ($r_i > c_i$). If store i has not managed to sell all their products, ($D_i < Q_i$), at the end of the season, the surplus inventory will have a per unit salvage value of $s_i > 0$ for store i. There will be an opportunity to sell it back to the factory, or they can let it go on sale for a sales price ($s_i < c_i$) after the season has finished. This might lead to increased storage-expenses, but this can be included in the per unit salvage value. If store i can not satisfy their entire demand, they are penalized a unit cost p_i (the cost of a customers dissatisfaction).

When the season has started, and store j has sold all the products in their warehouse, ($D_j > Q_j$), it will be possible to transship products from another store i with a surplus inventory of the product ($D_i < Q_i$), in order to satisfy the demand at store j. The transshipment cost per unit is denoted by τ_{ij}. We will assume that the customers are willing to wait for the transshipment T_{ij} i.e. the lead time is considered negligible. This will be a natural assumption in the case where the locations are in close proximity to each other or when an overnight delivery service is used for a greater area. Otherwise the loss of goodwill due to the delay can be included in the transshipment cost $\tau_{ij} > 0$. Furthermore we assume negligible fixed transshipment costs in our model formulation. To see the effect of fixed and joint ordering costs on a two location model formulation see Herer and Rashit [7]. Transshipments will be considered as a recourse action occurring after demand realization, but before demand must be satisfied in order to optimize profit.

2.1 Parameter Assumptions

The object of our model is to determine the optimal ordering and transshipment policies that maximizes aggregate profit. In our model we will employ a transshipment policy known as *complete pooling*. This transshipment policy can be described as follows ([8]): the amount transshipped from one location to another will be the minimum between (a) the surplus inventory of the sending location and (b) the shortage inventory at the receiving location. Accordingly, transshipments will take place until all remaining locations for which demand has not been completely fulfilled must either have a surplus inventory or they all have to have a shortage inventory.

The optimality of the complete pooling policy is ensured under the so-called triangle inequalities (which we will denote as the complete pooling assumptions).

$$r_j + p_j - \tau_{ij} \geq s_i \; i, j = 1, \ldots, n. \tag{1}$$

$$r_i + p_i \geq r_j + p_j - \tau_{ij} \; i, j = 1, \ldots, n. \tag{2}$$

$$s_i \geq s_j - \tau_{ij} \; i, j = 1, \ldots, n. \tag{3}$$

Equation 1 implies that it is always beneficial to transship from a location with excess inventory to a location with shortage inventory. This is so since the revenue value and the saved penalty cost at the receiving location minus the transshipment cost, $r_j + p_j - \tau_{ij}$, out-weighs the salvage value, s_i, at the shipping location. Further, it is neither preferable to transship between two shortage locations by Eq. 2, nor between two surplus locations by Eq. 3. In addition, to ensure that it is not beneficial to order indirectly from another location (instead of directly from the factory) we consider only the cases where

$$c_i + \tau_{ij} \geq c_j \quad i, j = 1, \ldots, n.$$

These assumptions are common in the literature on transshipments (e.g. [7, 17, 19]) and seem to be justified in practice.

2.2 Objective Function

In this section we formalize the problem. We consider the case where the choices of order quantities in each location are centrally coordinated. Retail stores all owned by the same company can be forced to cooperate since it is in the best interest for the company as a whole to maximize total aggregate profit. We can write the maximum aggregate profit for a company with retailers on n locations as

$$\max_{Q_1,\ldots,Q_n} \pi = \max_{Q_1,\ldots,Q_n} \left\{ \sum_{i=1}^{n} -c_i Q_i + E\bar{K}(\mathbf{Q}, \mathbf{D}) \right\} \tag{4}$$

where $\bar{K}$ is the maximum income given order quantities and realized demands. Notation in boldface indicates the corresponding vector/matrix. Note that for notational convenience we will use the "transshipment" variable T_{ii} as the amount sold at location i from own inventory at location i. Thus it would be natural to set $\tau_{ii} = 0$.

Due to the complete pooling policy, all transshipments are sold at the receiving location. This allows us to write the maximum income $\bar{K}$ as

$$\bar{K}(\mathbf{Q}, \mathbf{D})$$

$$= \max_{T_{ij}} \sum_{i=1}^{n} \left[\sum_{j=1}^{n} r_j T_{ij} - \sum_{j=1}^{n} \tau_{ij} T_{ij} + s_i \left(Q_i - \sum_{j=1}^{n} T_{ij} \right) - p_i \left(D_i - \sum_{j=1}^{n} T_{ji} \right) \right] \tag{5}$$

$$\text{subject to} \sum_{j=1}^{n} T_{ij} \leq Q_i \quad , \forall i = 1, \ldots, n. \tag{6}$$

$$\sum_{j=1}^{n} T_{ji} \leq D_i \quad , \forall i = 1, \ldots, n. \tag{7}$$

$$T_{ij} \geq 0 \quad , \forall i, j = 1, \ldots, n.$$

The first term on the right hand side of Eq. 5 can be recognized as the income from all that is sent from location i and sold at location j. The second term is the corresponding transshipment costs. The third term is the salvage value from the surplus inventory at location i, while the fourth term is the penalty cost for not meeting the demand at location i. Constraints (6) and (7) say that you can not sell more than you have at hand, nor can you sell more than the demand at the location. By extracting $\sum_i^n s_i Q_i$ and $\sum_i^n p_i D_i$ from $\bar{K}$, program (4) can be reformulated as

$$\max_{Q_1,\ldots,Q_n} \pi = \max_{Q_1,\ldots,Q_n} \left\{ -\sum_{i=1}^{n}((c_i - s_i)Q_i + p_i E(D_i)) + EK(\mathbf{Q},\mathbf{D}) \right\} \tag{8}$$

where

$$\begin{aligned} K(\mathbf{Q},\mathbf{D}) = \max_{T_{ij}} &\sum_{i=1}^{n}\sum_{j=1}^{n} \left(r_j + p_j - \tau_{ij} - s_i\right) T_{ij} \\ \text{subject to} &\sum_{j=1}^{n} T_{ij} \le Q_i \quad , \forall i = 1,\ldots,n. && (9) \\ &\sum_{j=1}^{n} T_{ji} \le D_i \quad , \forall i = 1,\ldots,n. && (10) \\ &T_{ij} \ge 0 \quad , \forall i, j = 1,\ldots,n. \end{aligned}$$

Not surprisingly, we can see from Eq. 8 that the optimal order quantity can be seen as a balance between the procurement cost minus the salvage value of the total order quantity, and the expected profit from the total amount transshipped in the second stage program K.

By using the result that the optimal value of a linear program is a concave polyhedral function of its right hand side vector (see [3] pp. 697), it can easily be shown that the π program is jointly concave in the decision variables $\mathbf{Q}$. Thus the first order conditions give an optimal solution which allows us to determine the optimal order quantity.

2.3 Dual Formulation of the K Program

In order to determine the first order conditions of π we will first look at the dual formulation of the K-problem. We will in this section rewrite the linear programming problem K in terms of the extreme points of the feasible region of the dual problem. Each of these extreme points will be optimal for some polyhedral cone of $\mathbf{D}$, allowing us to compute the expectation of K and its partial derivatives with respect to $\mathbf{Q}$. To solve the first order conditions of Eq. 11, we use the simple fact that for a linear program, the dual value of a constraint is the derivative of the objective function with respect to the right hand side of that constraint.

$$\frac{\partial \pi}{\partial Q_k} = -(c_k - s_k) + \frac{\partial EK(\mathbf{Q},\mathbf{D})}{\partial Q_k} \qquad , k = 1,\ldots,n. \tag{11}$$

We will first define the dual problem $\tilde{K}$ of K. Let α and β be the dual variables associated with Eqs. 9 and 10. Then we can write the dual problem $\tilde{K}$ as:

$$\begin{aligned} \tilde{K}(\mathbf{Q},\mathbf{D}) = \min \quad & \mathbf{Q}\alpha + \mathbf{D}\beta \\ \text{subject to} \quad & \alpha_i + \beta_j \geq r_j + p_j - \tau_{ij} - s_i \qquad i,j = 1,\ldots,n. \\ & \alpha,\beta \geq \mathbf{0} \end{aligned}$$

Let s be the number of feasible extreme points in $\tilde{K}$, and let $(\alpha_\mathbf{l}, \beta_\mathbf{l})$ describe these extreme points. We then get,

$$\tilde{K}(\mathbf{Q},\mathbf{D}) = \min_{l=1,\ldots,s} \{\mathbf{Q}\alpha_\mathbf{l} + \mathbf{D}\beta_\mathbf{l}\}$$

Denote the region of $\mathbf{D}$ where $(\alpha_\mathbf{l}, \beta_\mathbf{l})$ is optimal, for a given $\mathbf{Q}$, as $\omega_l \in \mathbb{R}^n_+$. The optimality regions ω_l, for a given $\mathbf{Q}$, can be characterized as

$$\omega_\mathbf{l}(\mathbf{Q}) = \{\mathbf{D} | \mathbf{Q}\alpha_\mathbf{l} + \mathbf{D}\beta_\mathbf{l} \leq \mathbf{Q}\alpha_\mathbf{i} + \mathbf{D}\beta_\mathbf{i} \quad, i = 1,\ldots,s\} \qquad, l = 1,\ldots,s.$$

The expectation of $\tilde{K}$ can now be written as the sum of the s products of optimal $(\alpha_\mathbf{l}, \beta_\mathbf{l})$ and the probability of a random $\mathbf{D}$-vector lying in the respective region ω_l.

$$E\tilde{K}(\mathbf{Q},\mathbf{D}) = \sum_{l=1}^{s} \Pr(\mathbf{D} \in \omega_\mathbf{l})\,(\mathbf{Q}\alpha_\mathbf{l} + \mathbf{D}\beta_\mathbf{l})$$

From which it follows that

$$\frac{\partial E\tilde{K}}{\partial \mathbf{Q}} = \sum_{l=1}^{s} \Pr(\mathbf{D} \in \omega_\mathbf{l})\,\alpha_\mathbf{l}$$

Using linear programming duality yields

$$\frac{\partial EK}{\partial \mathbf{Q}} = \frac{\partial E\tilde{K}}{\partial \mathbf{Q}} = \sum_{l=1}^{s} \Pr(\mathbf{D} \in \omega_\mathbf{l})\,\alpha_\mathbf{l}$$

$\Pr(\mathbf{D} \in \omega_l)$ can be found by integrating the demand density function over ω_l. Thus we can write the partial derivative of the stochastic program π as

$$\frac{\partial \pi}{\partial Q_k} = -(c_k - s_k) + \sum_{l=1}^{s} \alpha_l \int \cdots \int_{\omega_l} f(\mathbf{D})\, d\mathbf{D} \qquad, k = 1,\ldots,n. \tag{12}$$

With a characterization of α_l and ω_l for $l = 1,\ldots,s$, we can characterize the gradient of π *exactly* if we are able to calculate the integral of Eq. 12 exactly. However, calculating these probability integrals analytically is usually either too difficult or computationally too expensive. Also, the number of extreme points increases exponentially as the number of locations increases. This makes the required calculations for determining the optimal order quantity enormous as n increases. Regardless of the complexity, we will in the next few sections show how to characterize the optimal order quantities in the two, three and four location model, respectively, in an analytical way.

Note that whether we find the optimality regions and corresponding dual in terms of the primal (K) or dual ($\tilde{K}$) problem, is a matter of taste. While the dual problem gives a direct insight into the partial derivatives, the primal problem is clearer and easier to understand. Since we are interested in characterizing analytical solutions for an already complex stochastic linear programming problem, we will for clarity reasons use the primal problem K to find the optimality conditions in this paper.

3 Transshipments with Two Locations

We will now look at the basic two location model as an introduction to more complex versions of the transshipment problem.

In general, we want to map any demand realization onto a corresponding transshipment matrix **T**, given cost parameters and order quantities, in order to determine the expectation of the K program. To achieve this, we would like to have a general way of solving the linear program (by considering different cases of the cost parameters) instead of solving large sets of similar linear programming problems.

Due to the optimality of the complete pooling policy, the transshipments will be completely determined by the order quantity and realized demand for a two location model. From Tagaras [19] we have that the formal expressions of the optimal transshipments T_{ij} $(i \neq j)$, can be characterized as follows

$$\text{if} \quad \forall i \quad D_i \leq Q_i \quad \text{then} \quad T_{ij} = 0 \tag{13}$$

$$\text{if} \quad \forall i \quad D_i \geq Q_i \quad \text{then} \quad T_{ij} = 0 \tag{14}$$

$$\begin{aligned} &\text{if} \quad D_i < Q_i \quad \text{and} \quad D_j > Q_j \\ &\text{then} \quad T_{ij} = \min(Q_i - D_i, D_j - Q_j) \quad \text{and} \quad T_{ji} = 0 \end{aligned} \tag{15}$$

While T_{ij} $(i \neq j)$ depends on the cost parameters for more than two locations, T_{ii} does not. The amount sold of own inventory, T_{ii}, will of course always equal the minimum of the demand and the order quantity at the location.

$$T_{ii} = \min(D_i, Q_i) \qquad i = 1, \ldots, n$$

For given order quantities, Eqs. 13, 14, and 15 divides the possible demand realizations into six different regions, each of which corresponds to an optimal characterization of transshipments for that region. The graphical representation of the six optimal regions (E_i ,$i = 1, \ldots, 6$) resulting from Eqs. 13, 14, and 15, are depicted in Figure 1. Note that for simplicity, the events defined in any section of this paper are specified only for the model of that section. The expectation of the K program can now be found by weighting the expectation of the transshipments by the probability that the demand realization will fall into the corresponding region.

By deriving the partial derivative of expected transshipments and determining the first order conditions, the optimal order quantity can be characterized by

$$\begin{aligned} &\Pr(D_i < Q_i) - \frac{(r_j + p_j - \tau_{ij} - s_i)}{r_i + p_i - s_i} \Pr(D_i < Q_i < D_i + D_j - Q_j) \\ &+ \frac{(r_i + p_i - \tau_{ji} - s_j)}{r_i + p_i - s_i} \Pr(D_i + D_j - Q_i < Q_j < D_j) = \frac{r_i + p_i - c_i}{r_i + p_i - s_i} \\ &\text{for } i, j = 1, 2; \; i \neq j \end{aligned}$$

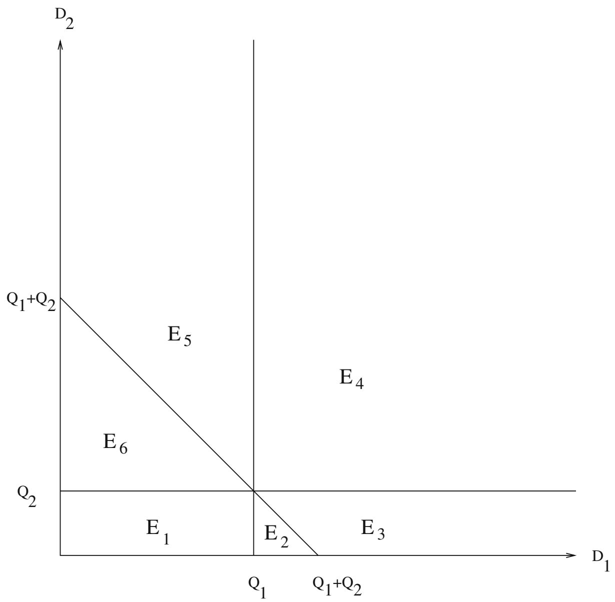


Figure 1 Optimality regions of a two location model given order quantity **Q**.

(see [18]). In words, Eq. 16 shows how the optimal order quantity Q_i trades off the expected marginal benefit with the marginal cost (r.h.s.) at location i, while adjusting for the possibility of transshipping inventory to location j (second term l.h.s.) or receiving inventory from location j (third term l.h.s.).

Robinson [17] shows how to find the optimal order quantity for the case of two locations and non-identical cost parameters. However, he also claims that this can not be done for more than two locations. The motivation for this paper is mainly to show how to find the optimal order quantities for three and four locations as well.

4 Transshipments with Three Locations

The problem of how much to transship in a general multiple location model with non-identical cost parameters will be determined by the cost parameters as well as the inventory level and realized demand. We define the optimal transshipment policy for three locations in Section 4.1. In Section 4.2 we find the formal expressions for the transshipment policy, while we in Section 4.3 use these expressions to find a characterization of the optimal inventory choices.

4.1 Optimal Transshipment Policy

Let us name the event of $D_i < Q_i$ as $+$ (surplus inventory) and the event of $D_i > Q_i$ as $-$ (shortage inventory). The following table illustrates the possible different transshipment structures, given order quantities and demand realizations, for three locations. The table also shows the corresponding positive transshipments that may occur.

Structure	Q_i	Q_j	Q_k	T_{ii}	T_{ij}	T_{ik}	T_{ji}	T_{jj}	T_{jk}	T_{ki}	T_{kj}	T_{kk}
1	$-$	$-$	$-$	T_{ii}	0	0	0	T_{jj}	0	0	0	T_{kk}
2	$+$	$-$	$-$	T_{ii}	T_{ij}	T_{ik}	0	T_{jj}	0	0	0	T_{kk}
3	$+$	$+$	$-$	T_{ii}	0	T_{ik}	0	T_{jj}	T_{jk}	0	0	T_{kk}
4	$+$	$+$	$+$	T_{ii}	0	0	0	T_{jj}	0	0	0	T_{kk}

For each of the different transshipment structures we want to characterize the optimal transshipments given any order quantity and demand realization. Note that all of the locations in structure 1 have a shortage inventory, while they all have a surplus inventory in structure 4. This means that there will be no transshipments between the locations for these transshipment structures (this follows from the complete pooling assumptions). For transshipment structures 2 and 3 there is a maximum of two positive transshipments T_{ij} $(i \neq j)$ in an optimal allocation. These transshipments will depend on the cost parameters. For notational convenience, let a_{ij} be defined as the net income from transshipping from location i to j: $a_{ij} = r_j + p_j - \tau_{ij} - s_i$, $1 \leq i, j \leq n$. The following lemma will help us determine the optimal transshipment policy for the three location model. Note that the term greedy in this paper will be addressed to the cost parameters a_{ij}. For each greedy choice of a_{ij} it is assumed that transshipment T_{ij} either will yield the surplus inventory at location i to become zero, or be equal to the shortage inventory at location j.

Lemma 1 Consider a given order quantity $\mathbf{Q} \in \mathbb{R}^n_+$ and demand realization $\mathbf{D} \in \mathbb{R}^n_+$ that results in a transshipment structure containing maximally one shortage *or* maximally one surplus inventory location. The optimal transshipment policy of the corresponding $K(\mathbf{Q}, \mathbf{D})$ program will then be in the form of a greedy allocation.

Proof Due to the complete pooling assumptions it follows that it is neither profitable to transship between two shortage inventory locations nor between two surplus inventory locations. However, it will always be profitable to transship from a surplus to a shortage inventory location. The proof follows then trivially due to the structure of the problem. □

Proposition 1 The optimal transshipment policy in a two and three location model is in the form of a greedy allocation.

Proof Follows from Lemma 1 and the fact that any given order quantity and demand realization in the two and three location model results in a transshipment structure with maximum one shortage *or* maximum one surplus inventory location. □

4.2 Formal Expressions of the Transshipment Policy

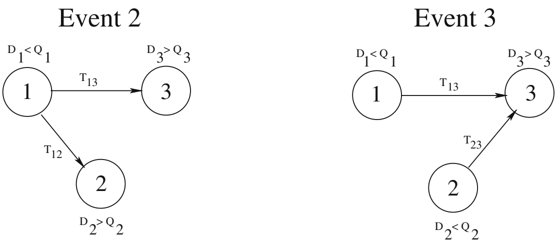
Given cost parameters and an order quantity vector $\mathbf{Q}$, we can now define a mapping G from the set of arbitrary demand vectors $\{\mathbf{D}\}$ to the set of transshipment matrices $\{\mathbf{T}\}$, $G : \mathbb{R}^3_+ \to \mathbb{R}^{3x3}_+$. Define also the mapping g from demand $\{\mathbf{D}\}$ to a

Table I The relevant cost structures for T_{13} in Events 2 and 3

Event 2	Event3
$a_{13} > a_{12}$	$a_{13} > a_{23}$
$a_{13} < a_{12}$	$a_{13} < a_{23}$

general transshipment $\{T_{13}\}$, $g : \mathbb{R}^3_+ \to \mathbb{R}_+$. For ease of analysis and without loss of generality, only the mapping g will be explicitly characterized in our further analysis.

In order for T_{13} not to be zero we must have $D_1 < Q_1$ and $D_3 > Q_3$. This leaves us with the events of $D_2 > Q_2$ and $D_2 < Q_2$ which corresponds to transshipment structures 2 and 3, and are defined as Events 2 and 3, respectively.



In both of the events there will be two cost structures of relevance for T_{13} which is shown in Table I. The mapping g can then be characterized as follows

Event 2: $D_1 < Q_1,\ D_2 > Q_2,\ D_3 > Q_3$

$$a_{13} > a_{12}: \qquad T_{13} = \min(Q_1 - D_1,\ D_3 - Q_3) \tag{16}$$

$$a_{13} < a_{12}: \qquad T_{13} = \min((Q_1 + Q_2 - D_1 - D_2)^+,\ D_3 - Q_3) \tag{17}$$

Event 3: $D_1 < Q_1,\ D_2 < Q_2,\ D_3 > Q_3$

$$a_{13} > a_{23}: \qquad T_{13} = \min(Q_1 - D_1,\ D_3 - Q_3) \tag{18}$$

$$a_{13} < a_{23}: \qquad T_{13} = \min((D_2 + D_3 - Q_2 - Q_3)^+,\ Q_1 - D_1) \tag{19}$$

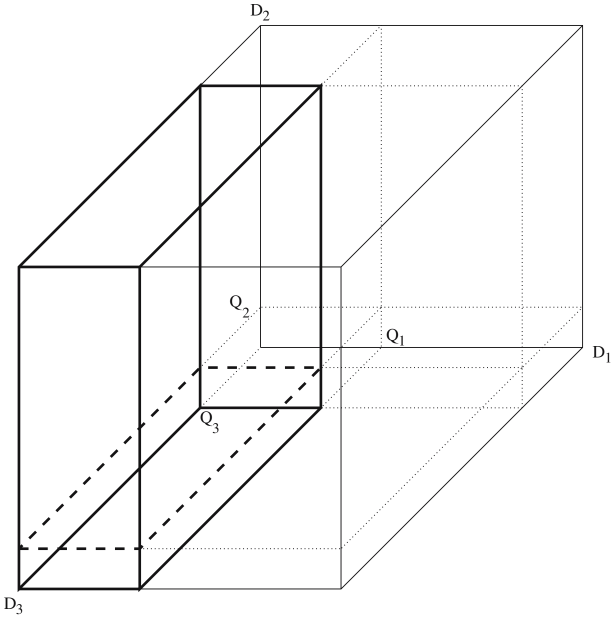
For other order quantities and demand realizations there will be no transshipments between location 1 and 3, $T_{13} = 0$, regardless of cost structures.

Figure 2 divides the demand realizations into regions of shortage and surplus inventory for each of the three locations. Correspondingly, Figure 3 divides the region of demand realizations into total shortage and total surplus inventory for any two locations, and for all three locations. The graphical representation from combining the three planes of Figure 2 and the four planes of Figure 3 will characterize the different optimality regions of demand realizations $\mathbf{D}$ that maps into a corresponding transshipment matrix $\mathbf{T}$. in such a way that they cover $\mathbb{R}^3_+$ without overlap. Note that the only interesting region for the mapping g is the region above and below the bold dashed rectangle in Figure 2. These are namely the regions of Events 2 and 3.

4.3 Optimal Inventory Choices

In order to characterize the first order conditions of the π program we find the expectations of the transshipments, and then derive the partial derivatives of the ex-

Figure 2 Three of seven planes of significance in the mapping $G : \mathbb{R}^3_+ \rightarrow \mathbb{R}^{3x3}_+$.



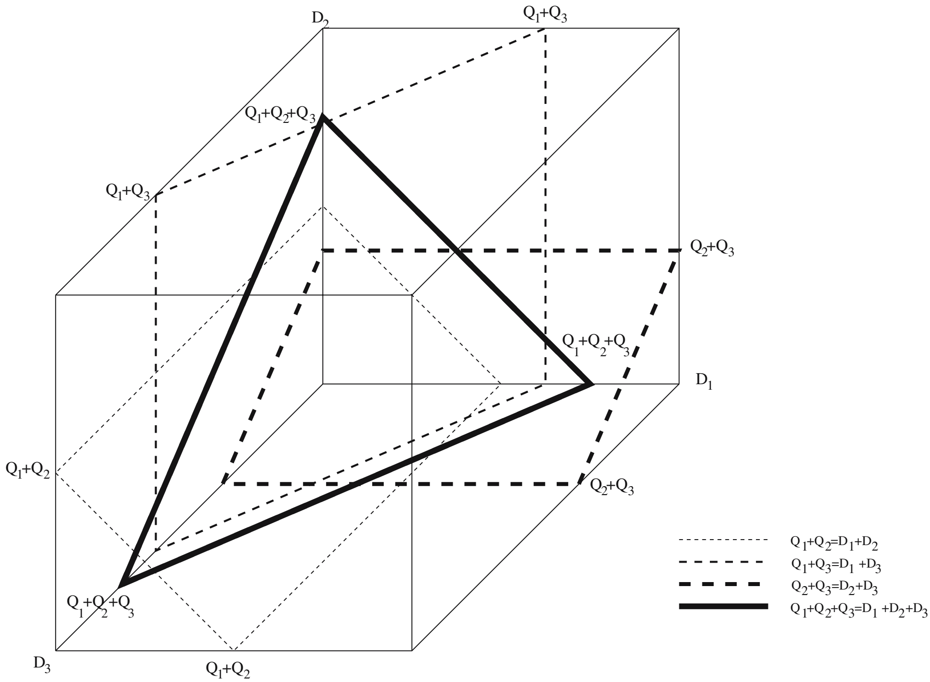


Figure 3 Four of seven planes of significance in the mapping $G : \mathbb{R}^3_+ \rightarrow \mathbb{R}^{3x3}_+$.

pectations of the transshipments. These expressions will then be used to characterize the conditions for the optimal order quantity.

4.3.1 Expectations and Derivatives of Transshipments

The expression for an optimal transshipment T_{13} is (as shown in the previous section) dependent on the cost structure, the order quantity and the realization of demand. Define $\hat{T}_{ij}$ as the expected value of T_{ij} given optimal allocations of transshipments in the second stage program K, $\hat{T}_{ij} = E(T_{ij})$. The different cost structures that are relevant to $\hat{T}_{13}$, are found by combining the cost structures for T_{13} from Events 2 and 3. In order to characterize $\hat{T}_{13}$, we divide our analysis into the following four possible cases of cost structures.

Cost structure	Event 2	Event 3
γ_1	$a_{13} > a_{12}$	$a_{13} > a_{23}$
γ_2	$a_{13} > a_{12}$	$a_{13} < a_{23}$
γ_3	$a_{13} < a_{12}$	$a_{13} > a_{23}$
γ_4	$a_{13} < a_{12}$	$a_{13} < a_{23}$

We find $\hat{T}_{13}$ by summing the expectations of the different possible optimal transshipments of T_{13}, weighted with the corresponding probability for a demand realization to fall into the respective regions, ω_l, which makes the different possible transshipment of T_{13} optimal, in other words

$$\hat{T}_{13} = \sum_{l=1}^{s} \Pr(\mathbf{D} \in \omega_l) E(T_{13} | \mathbf{D} \in \omega_l)$$

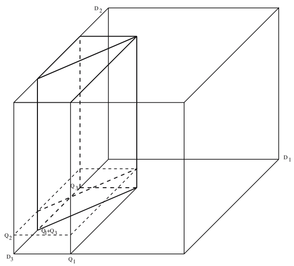
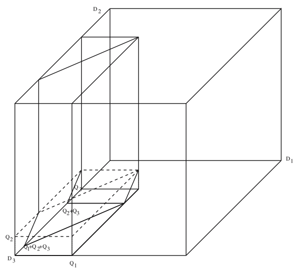
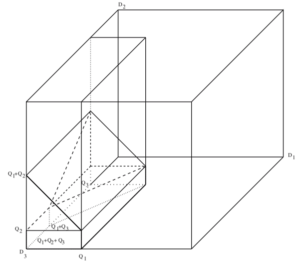
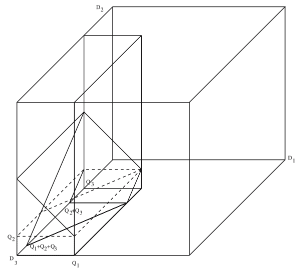
While the optimal transshipments are characterized in Eqs. 16, 17, 18, and 19, Table II characterizes the corresponding regions of optimality. The events of Table II are all subevents of Events 2 and 3 from Section 4.2 and is derived from Eqs. 16, 17, 18 and 19. To simplify notation in further analysis let $\bar{Q}_i = Q_i - D_i$ and $\bar{D}_i = D_i - Q_i$.

$\hat{T}_{13}$ and its partial derivatives for cost structures γ_i $(1 \le i \le 4)$ are put into Table III. The optimality regions for T_{13} for each of the cost structures are also shown in Table III. In order to describe how the expressions in Table III were found, we will look closer at the analysis for cost structure γ_2. Due to the complexity of the figures

Table II Events used to determine the expectations of T_{13} for three locations

Event	Description
$E_{2.1}$	$D_1 + D_3 < Q_1 + Q_3,\ D_1 < Q_1,\ D_2 > Q_2,\ D_3 > Q_3$
$E_{2.2}$	$D_1 + D_3 > Q_1 + Q_3,\ D_1 < Q_1,\ D_2 > Q_2,\ D_3 > Q_3$
$E_{2.3}$	$D_1 + D_2 + D_3 < Q_1 + Q_2 + Q_3,\ D_1 < Q_1,\ D_2 > Q_2,\ D_3 > Q_3$
$E_{2.4}$	$D_1 + D_2 + D_3 > Q_1 + Q_2 + Q_3,\ D_1 + D_2 < Q_1 + Q_2,\ D_1 < Q_1,\ D_2 > Q_2,\ D_3 > Q_3$
$E_{3.1}$	$D_1 + D_3 < Q_1 + Q_3,\ D_1 < Q_1,\ D_2 < Q_2,\ D_3 > Q_3$
$E_{3.2}$	$D_1 + D_3 > Q_1 + Q_3,\ D_1 < Q_1,\ D_2 < Q_2,\ D_3 > Q_3$
$E_{3.3}$	$D_1 + D_2 + D_3 < Q_1 + Q_2 + Q_3,\ D_2 + D_3 > Q_2 + Q_3,\ D_1 < Q_1,\ D_2 < Q_2,\ D_3 > Q_3$
$E_{3.4}$	$D_1 + D_2 + D_3 > Q_1 + Q_2 + Q_3,\ D_1 < Q_1,\ D_2 < Q_2,\ D_3 > Q_3$

Table III Expectations of T_{13} and partial derivatives of $\hat{T}_{13}$

				
	$\gamma_1 : a_{13} > a_{12}$ & $a_{13} > a_{23}$	$\gamma_2 : a_{13} > a_{12}$ & $a_{13} < a_{23}$	$\gamma_3 : a_{13} < a_{12}$ & $a_{13} > a_{23}$	$\gamma_4 : a_{13} < a_{12}$ & $a_{13} < a_{23}$
$\hat{T}_{13}$	$\Pr(E_{2.1})E(\bar{D}_3\|\mathbf{D} \in E_{2.1})$ $+\Pr(E_{2.2})E(\bar{Q}_1\|\mathbf{D} \in E_{2.2})$ $+\Pr(E_{3.1})E(\bar{D}_3\|\mathbf{D} \in E_{3.1})$ $+\Pr(E_{3.2})E(\bar{Q}_1\|\mathbf{D} \in E_{3.2})$	$\Pr(E_{2.1})E(\bar{D}_3\|\mathbf{D} \in E_{2.1})$ $+\Pr(E_{2.2})E(\bar{Q}_1\|\mathbf{D} \in E_{2.2})$ $+\Pr(E_{3.3})E(\bar{D}_2 + \bar{D}_3\|\mathbf{D} \in E_{3.3})$ $+\Pr(E_{3.4})E(\bar{Q}_1\|\mathbf{D} \in E_{3.4})$	$\Pr(E_{2.3})E(\bar{D}_3\|\mathbf{D} \in E_{2.3})$ $+\Pr(E_{2.4})E(\bar{Q}_1 + \bar{Q}_2\|\mathbf{D} \in E_{2.4})$ $+\Pr(E_{3.1})E(\bar{D}_3\|\mathbf{D} \in E_{3.1})$ $+\Pr(E_{3.2})E(\bar{Q}_1\|\mathbf{D} \in E_{2.2})$	$\Pr(E_{2.3})E(\bar{D}_3\|\mathbf{D} \in E_{2.3})$ $+\Pr(E_{2.4})E(\bar{Q}_1 + \bar{Q}_2\|\mathbf{D} \in E_{2.4})$ $+\Pr(E_{3.3})E(\bar{D}_2 + \bar{D}_3\|\mathbf{D} \in E_{3.3})$ $+\Pr(E_{3.4})E(\bar{Q}_1\|\mathbf{D} \in E_{3.4})$
$\frac{\partial \hat{T}_{13}}{\partial Q_1}$	$Pr(E_{2.2}) + Pr(E_{3.2})$	$Pr(E_{2.2}) + Pr(E_{3.4})$	$Pr(E_{2.4}) + Pr(E_{3.2})$	$Pr(E_{2.4}) + Pr(E_{3.4})$
$\frac{\partial \hat{T}_{13}}{\partial Q_2}$	0	$-Pr(E_{3.3})$	$Pr(E_{2.4})$	$Pr(E_{2.4}) - Pr(E_{3.3})$
$\frac{\partial \hat{T}_{13}}{\partial Q_3}$	$-Pr(E_{2.1}) - Pr(E_{3.1})$	$-Pr(E_{2.1}) - Pr(E_{3.3})$	$-Pr(E_{2.3}) - Pr(E_{3.1})$	$-Pr(E_{2.3}) - Pr(E_{3.3})$

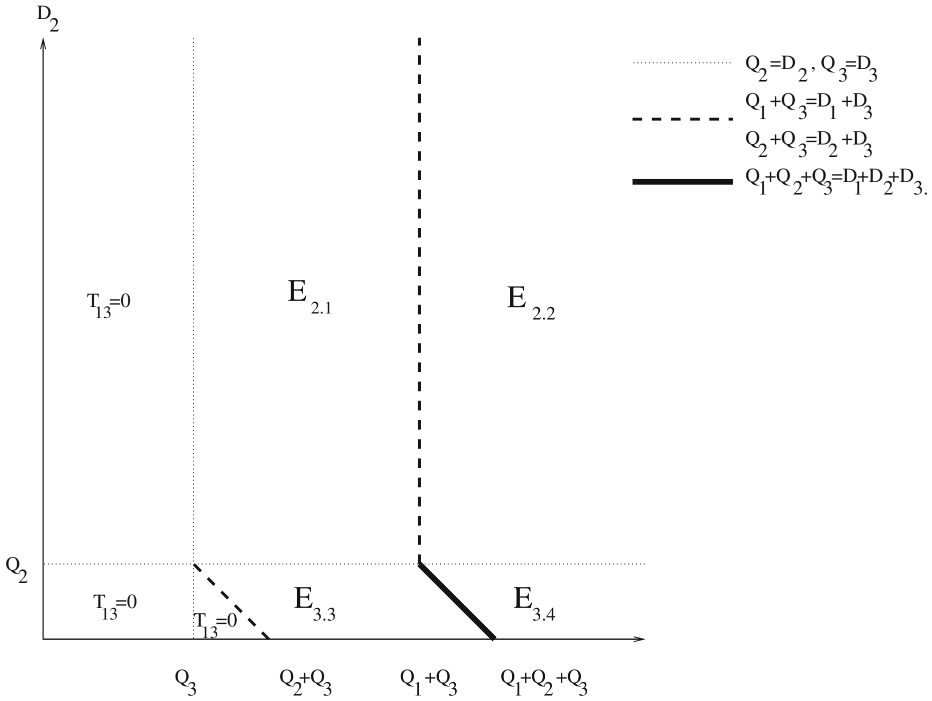


Figure 4 Regions of optimality for transshipment T_{13} given cost structure $\gamma_2 : a_{13} > a_{12}, a_{13} < a_{23}$ illustrated in the two dimensional plane (D_2, D_3) with $D_1 = 0$.

in Table III we illustrate the case γ_2 graphically in two dimensions (D_2, D_3) with $D_1 = 0$ in Figure 4 (see Table III for a three dimensional figure). There are only four events where $T_{13} > 0$ for cost structure γ_2, and these are depicted in Figure 4. The optimal transshipments from the events of Figure 4 can be found from Eqs. 16 and 19. We can then write $\hat{T}_{13}$ for cost structure γ_2 as

$$\hat{T}_{13} = \Pr(E_{2.1})E(\bar{D}_3|\mathbf{D} \in E_{2.1}) + \Pr(E_{2.2})E(\bar{Q}_1|\mathbf{D} \in E_{2.2}) \\ + \Pr(E_{3.3})E(\bar{D}_2 + \bar{D}_3|\mathbf{D} \in E_{3.3}) + \Pr(E_{3.4})E(\bar{Q}_1|\mathbf{D} \in E_{3.4}) \quad (20)$$

The partial derivative of $\hat{T}_{13}$ with respect to Q_1 can then be expressed as

$$\frac{\partial \hat{T}_{13}}{\partial Q_1} = \Pr(E_{2.2}) + \Pr(E_{3.4})$$

An increase in Q_1 affects $\hat{T}_{13}$ under cost structure γ_2 by increasing T_{13} by the same amount under events $E_{2.2}$ and $E_{3.4}$ and leaving T_{13} unchanged under all other events, including $E_{2.1}$ and $E_{3.3}$.

From Table III we can see, as would be expected, that $\frac{\partial \hat{T}_{13}}{\partial Q_1}$ and $\frac{\partial \hat{T}_{13}}{\partial Q_3}$ are strictly positive and negative respectively regardless of cost parameters. $\frac{\partial \hat{T}_{13}}{\partial Q_2}$ is very much depending on the cost parameters. Table III shows that $\frac{\partial \hat{T}_{13}}{\partial Q_2}$ varies from strictly positive to strictly negative depending on the cost parameters.

4.3.2 Characterization of Optimal Inventory Choices

The partial derivative of the π program can be written as

$$\frac{\partial \pi}{\partial Q_k} = -c_k + s_k + (r_k + p_k - s_k)\Pr(D_k > Q_k) + \sum_{i=1}^{3}\sum_{j=1, i\neq j}^{3} (r_j + p_j - \tau_{ij} - s_i)\frac{\partial \hat{T}_{ij}}{\partial Q_k} \quad \text{for} \quad k = 1, 2, 3 \tag{21}$$

where $\frac{\partial \hat{T}_{ij}}{\partial Q_k}$ $(i,\ j, k = 1, 2, 3)$ depends on the given cost parameters and can be found by generalizing the partial derivatives expressions of Table III. By rearranging and equating Eq. 21 to zero, we can now characterize the conditions for the optimal order quantity of the π program as

$$\Pr(D_k < Q_k) - \sum_{i=1}^{3}\sum_{j=1, i\neq j}^{3} \frac{(r_j + p_j - \tau_{ij} - s_i)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{ij}}{\partial Q_k} = \frac{r_k + p_k - c_k}{r_k + p_k - s_k} \quad \text{for } k = 1, 2, 3 \tag{22}$$

Note that there is a parallel between Eq. 22 and the well known newsvendor model. In the latter model there are no transshipments allowed and the optimal order quantity for location i can be expressed as Eq. 23.

$$\Pr(D_k < Q_k) = \frac{r_k + p_k - c_k}{r_k + p_k - s_k} \tag{23}$$

In order to make some comparison remarks between Eqs. 22 and 23, we will first rewrite Eq. 22 in full length.

$$\begin{aligned}\Pr(D_k < Q_k) \; & - \frac{(r_i + p_i - \tau_{ki} - s_k)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{ki}}{\partial Q_k} - \frac{(r_j + p_j - \tau_{kj} - s_k)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{kj}}{\partial Q_k} \\ & - \frac{(r_k + p_k - \tau_{ik} - s_i)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{ik}}{\partial Q_k} - \frac{(r_k + p_k - \tau_{jk} - s_j)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{jk}}{\partial Q_k} \\ & - \frac{(r_j + p_j - \tau_{ij} - s_i)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{ij}}{\partial Q_k} - \frac{(r_i + p_i - \tau_{ji} - s_j)}{r_k + p_k - s_k}\frac{\partial \hat{T}_{ji}}{\partial Q_k} \\ & = \frac{r_k + p_k - c_k}{r_k + p_k - s_k} \quad \text{for } i,\ j, k = 1, 2, 3 \quad i \neq j \neq k \end{aligned} \tag{24}$$

When we compare Eqs. 23 and 24, we can see that Eq. 24 is only an adjustment of Eq. 23. By generalizing the partial derivatives of Table III, it is clear that the second and third term of Eq. 24 will be strictly positive regardless of cost parameters. This means that Q_k will be adjusted up due to the possibility of sending transshipments to location i and j. Likewise, the fourth and fifth term on the left hand side of Eq. 24 will be strictly negative, thus adjusting Q_k down due to the possibility of receiving transshipments from locations i and j. The sixth and seventh term will either have no affect on Q_k or adjust Q_k up or down depending on the cost structure. These latter terms will reflect the impact on Q_k due to the possibility of making transshipments between locations i and j. For a cost structure corresponding to γ_4 the sign of the sixth and seventh term will be dependent on the relative size of the order quantities as well.

The resulting inventory level from our transshipment model will normally be closer to the expected demand realization compared to the newsvendor model. This is due to the possibility of transshipments between the locations. Instead of each location depending solely on their own warehouse, with the extra costs this incurs (more often surplus and shortage inventory), they all have a common pool to draw upon.

5 Transshipments with Four Locations

We will now take a further step and look at the case of four locations where the inventory choices at each location are centrally coordinated. The motivation for this analysis is to provide insight into the analytical solution of more general systems where the optimal allocation of transshipments is not necessarily in a greedy form.

We define the optimal transshipment policy of the four location model in Section 5.1 and look at the optimality of a greedy allocation of transshipments. In Section 5.2 we define the formal expressions of the transshipment policy, while we in Section 5.3 make some remarks on how to characterize the conditions of the optimal order quantity.

5.1 Optimal Transshipment Policy

The table below illustrates the different transshipment structures that may arise in the four location model for any given order quantity and demand realization.

Structure	Q_i	Q_j	Q_k	Q_l
1	−	−	−	−
2	+	−	−	−
3	+	+	−	−
4	+	+	+	−
5	+	+	+	+

Note that there are no surplus inventory (+) locations for transshipment structure 1, likewise there are no shortage inventory (−) locations for transshipment structure 5. Also note that there is only one location with a surplus inventory for transshipment structure 2, while there is only one location with a shortage inventory for transshipment structure 4. It then follows from Lemma 1 that a greedy allocation of transshipments will be optimal for transshipment structures 1, 2, 4 and 5.

We will concentrate on transshipment structure 3 where a greedy allocation of transshipments is not necessarily optimal. In this structure we have two surplus and two shortage inventory locations. The K program have 16 unknown transshipments. When considering transshipment structure 3, this can be reduced to four unknowns. Without loss of generality, let $\mathbf{Q}$ be determined and $\mathbf{D}$ realized such that $D_1 <$

Q_1, $D_2 < Q_2$, $D_3 > Q_3$ and $D_4 > Q_4$. The related K program can then be written as

$$\max_{T_{ij}} \sum_{i=1}^{2}\sum_{j=3}^{4} a_{ij}T_{ij}$$

subject to

$$T_{13} + T_{14} \leq \bar{Q}_1 \tag{25}$$

$$T_{23} + T_{24} \leq \bar{Q}_2 \tag{26}$$

$$T_{13} + T_{23} \leq \bar{D}_3 \tag{27}$$

$$T_{14} + T_{24} \leq \bar{D}_4 \tag{28}$$

$$T_{ij} \geq 0 \quad , i = 1, 2 \quad j = 3, 4$$

where $\bar{Q}_i$ and $\bar{D}_i$ is defined as in Section 4.3. The K program has a very useful characteristic that will be presented without formal proof in the following lemma due to its simplicity.

Lemma 2 The constraints (9) of the K program that are bounded above by the order quantity $\mathbf{Q}$, will be active in optimum in the event of total shortage inventory $\left(\sum_{i=1}^{4} D_i > \sum_{i=1}^{4} Q_i\right)$. Likewise, the constraints (10) that are bounded above by the demand realizations $\mathbf{D}$, will be active in optimum in the event of total surplus inventory $\left(\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i\right)$.

Proof From the complete pooling assumptions it follows that it will be optimal to make transshipments until either there are no shortage or no surplus inventory locations left. □

This observation will be very useful in order to determine the optimal transshipment policy for transshipment structure 3. The idea is to first characterize one of the transshipments in an optimal allocation. Then we can characterize an adjacent transshipment as well based on Eqs. 25, 26, 27 and 28 and Lemma 2. This will reduce the problem to a three location problem.

Note that the only difference between the analysis of the total surplus and total shortage inventory, is which constraints that are binding in optimum. Therefore we will in our further analysis regarding transshipment structure 3 only focus on the event of total surplus inventory. This is because the same approach can be used to characterize similar results in the case of total shortage inventory.

Proposition 2 Given order quantities and demand realizations $D_i < Q_i, D_j < Q_j, D_k > Q_k, D_l > Q_l$ and $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$ with the corresponding cost structure $|a_{ik} - a_{jk}| > |a_{il} - a_{jl}|$ and $a_{jk} > a_{ik}$, we will then in the optimum of the related K program have $T_{jk} = \min(\bar{Q}_j, \bar{D}_k)$ $(i, j, k, l = 1, \ldots, 4;\;\; i \neq j \neq k \neq l)$.

Proof (by contradiction) Assume, without loss of generality, the following order quantities and demand realizations: $D_1 < Q_1, D_2 < Q_2, D_3 > Q_3, D_4 > Q_4$ and

$\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$ with the following cost structure: $|a_{13} - a_{23}| > |a_{14} - a_{24}|$ and $a_{23} > a_{13}$. Assume then that z is an optimal solution to K with $T_{23} < \min(\bar{Q}_2, \bar{D}_3)$.

$$z = \sum_{i=1}^{2} \sum_{j=3}^{4} a_{ij} T_{ij} \quad , \text{ where } \quad T_{23} < \min(\bar{Q}_2, \bar{D}_3)$$

Consider then the feasible solution $\bar{z}$ constructed from z

$$\bar{z} = \sum_{i=1}^{2} \sum_{j=3}^{4} a_{ij} \bar{T}_{ij} \quad , \text{ where } \quad \bar{T}_{23} = \min(\bar{Q}_2, \bar{D}_3)$$

We want to construct $\bar{z}$ such that $\bar{z} > z$ thus contradicting our assumption. In order to maintain the feasibility of $\bar{z}$ when we increase T_{23} by δ_1 we have to reduce T_{13} by δ_1. Also to ensure feasibility we have to reduce T_{24} by δ_2 and increase T_{14} by δ_2. Due to the notation that the constraints bounded above by the demand realization $\mathbf{D}$ will be active in optimum in the event of total surplus inventory, we can write

$$\begin{aligned} \bar{z} - z &= -a_{13}\delta_1 + a_{23}\delta_1 + a_{14}\delta_2 - a_{24}\delta_2 \\ &= \delta_1(a_{23} - a_{13}) + \delta_2(a_{14} - a_{24}) \quad , -\delta_1 \leq \delta_2 \leq \delta_1 > 0 \end{aligned} \tag{29}$$

Note that we have $\delta_1 = \bar{T}_{23} - T_{23} > 0$, while we in order to maintain feasibility have $-\delta_1 \leq \delta_2 \leq \delta_1$. Since $|a_{13} - a_{23}| > |a_{14} - a_{24}|$, $a_{23} > a_{13}$ and $-\delta_1 \leq \delta_2 \leq \delta_1$ it follows that $\bar{z} \geq z$, which implies that we will have $T_{23} = \min(\bar{Q}_2, \bar{D}_3)$ in an optimal solution. □

Assume $D_i < Q_i$, $D_j < Q_j$, $D_k > Q_k$, $D_l > Q_l$, $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$ and we were to transship one unit from location i to location k. This transshipment would result in an income of a_{ik}, but also in an opportunity loss of a_{jk} since the demand at location k will be reduced by one unit. Define this difference in potential income due to a shortage inventory at location k as $d_k = |a_{ik} - a_{jk}|$ and the corresponding difference for location l as $d_l = |a_{il} - a_{jl}|$. Note that d_k and d_l is defined as the differences between the corresponding cost parameters of the respective binding constraints, $T_{ik} + T_{jk} = D_k$ and $T_{il} + T_{jl} = D_l$, of the related K program. In words, Proposition 2 then says that there exists a solution of the related K program where it is optimal to transship as much as possible ($\min(\bar{Q}_j, \bar{D}_k)$) in the transshipment corresponding to the greatest cost parameter ($a_{jk} > a_{ik}$) in the greatest difference of potential income ($d_k > d_l$).

Proposition 3 An optimal transshipment policy for transshipment structure 3 of the four location model (given a total surplus inventory), is to use Proposition 2 to determine one of the transshipments, and then determine the remaining transshipments in a greedy manner.

Proof Proposition 2 will characterize an optimal transshipment of the form $\min(Q_j, D_k)$. The related K program can then be reduced to a three location model where a greedy solution is always optimal (follows from Proposition 1). □

5.1.1 Optimality of a Greedy Allocation of Transshipments

The complexity of the π program increases significantly when the number of locations increases. For more than four locations the optimal transshipment policy is not known, thus one has to use some kind of heuristic. Herer et al. [9] reports a solution time of their heuristic of between two and three hours for up to seven locations. This heuristic is a gradient search based heuristic where they use Infinitesimal Perturbation Analysis (IPA) in order to estimate the gradient in each step. Basically this means that they are solving a *huge* number of transshipment problems (corresponding to our K program) in *every* gradient step. By characterizing the conditions on the cost structure for when a greedy allocation of transshipments is optimal, the solution time would be dramatically reduced for these cases. In fact, every transshipment problem in every gradient step can then be solved in linear time by sorting the cost parameters *once*. Note also that in the cases where the cost structure does not satisfy the necessary conditions, a greedy allocation might still be optimal depending on the order quantities and demand realizations. We are thus interested in determining the optimality of a greedy allocation. The analytical insight obtained could then be used to develop efficient heuristics for large size problems. First we introduce some notation.

Definition 1 If $a_{jk} > a_{ik}$ and $a_{jl} > a_{il}$, location j are said to *dominate* location i in the event of $D_i < Q_i$, $D_j < Q_j$, $D_k > Q_k$, $D_l > Q_l$ and $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$.

Given $D_i < Q_i$, $D_j < Q_j$, $D_k > Q_k$, $D_l > Q_l$ and $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$, consider the following cost structures

Case ϕ: Neither of the surplus inventory locations dominates.
Case χ: One of the surplus inventory locations dominates.
The shortage inventory location k with the greatest difference of potential income, $d_k > d_l$, also have the greatest cost parameter of the related K program, $a_{ik} \vee a_{jk} = \max\{a_{ik}, a_{il}, a_{jl}, a_{jk}\}$.

The characteristics of the two different cost structures and the terminology introduced can be illustrated as shown in Figure 5. In cost structure ϕ, lines between cost parameters corresponding to the same *surplus* inventory location are crossing. This implies no domination ($a_{jk} > a_{ik}$, $a_{jl} < a_{il}$). Accordingly for cost structure χ, lines do not cross, which implies domination ($a_{jk} > a_{ik}$, $a_{jl} > a_{il}$). Also note for case χ, that both the greater difference of potential income, d_k, and the greatest cost parameter, a_{jk}, correspond to the same *shortage* inventory location k.

Proposition 4 Given $D_i < Q_i$, $D_j < Q_j$, $D_k > Q_k$, $D_l > Q_l$ and $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$, a greedy allocation is then optimal if and only if we have cost structures ϕ *or* χ *or* there exists a dominating surplus inventory location that can satisfy the shortage inventory of both location k and l.

Proof Assume, without loss of generality, the following order quantities and demand realizations: $D_i < Q_i$, $D_j < Q_j$, $D_k > Q_k$, $D_l > Q_l$ and $\sum_{i=1}^{4} D_i < \sum_{i=1}^{4} Q_i$ with the corresponding cost structure: $|a_{ik} - a_{jk}| > |a_{il} - a_{jl}|$ and $a_{jk} > a_{ik}$. For cost structure

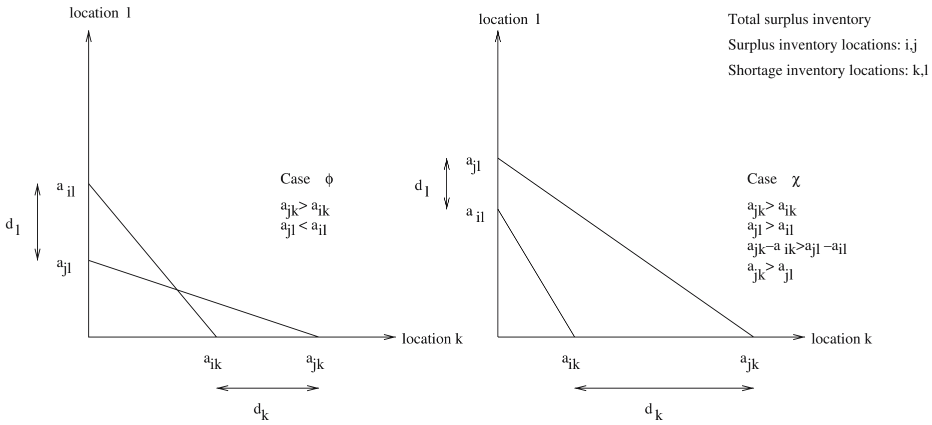


Figure 5 Graphical illustration of terminology.

ϕ we will have $a_{il} > a_{jl}$, else j will be dominating. Two cases can here be considered, either $a_{jk} > a_{il}$ or $a_{jk} < a_{il}$. For $a_{jk} > a_{il}$ we have that $a_{jk} = \max\{a_{ik}, a_{il}, a_{jl}, a_{jk}\}$. Then it follows directly from Propositions 2 and 3 that a greedy allocation is optimal. For $a_{jk} < a_{il}$ we want to show that there exists an optimal solution with $T_{il} = \min(\bar{Q}_i, \bar{D}_l)$. We can then see from Propositions 2 and 3 and the independence between T_{il} and T_{jk} due to no dominating location that a greedy solution will be optimal. Note that due to Proposition 2 we can reduce the related K program into a three location problem. If $\bar{Q}_j > \bar{D}_k$ we will have $T_{il} = \min(\bar{Q}_i, \bar{D}_l)$ since $a_{il} > a_{jl}$. If $\bar{Q}_j < \bar{D}_k$ we will have $T_{il} = \min(\bar{Q}_i, \bar{D}_l) = \bar{D}_l$ due to the binding constraint $T_{il} + T_{jl} = \bar{D}_l$.

For cost structure χ we have $a_{jl} > a_{il}$ and $a_{jk} > a_{jl}$. Since T_{jk} will be the most beneficial transshipment, it is trivial to see from Propositions 2 and 3 that a greedy allocation is optimal for this cost structure.

For a dominating surplus inventory location j we have $a_{jl} > a_{il}$, and to rule out cost structures χ we have $a_{jk} < a_{jl}$. A greedy allocation will be optimal as long as the surplus inventory in location j covers the shortage inventory at locations k and l, $\bar{Q}_j \geq \bar{D}_k + \bar{D}_l$, since T_{jk} and T_{jl} are the most beneficial transshipments for this cost structure. For $\bar{Q}_j < \bar{D}_k + \bar{D}_l$ we will have $T_{jl} < \bar{D}_l$, thus a greedy allocation will not be optimal. □

Note that by generalizing Proposition 4 and the corresponding result for the total shortage inventory case, we have the necessary and sufficient conditions for a greedy transshipment policy to be optimal in the four location model depending on the cost structure as well as the order quantity and demand.

5.2 Formal Expressions of the Transshipment Policy

Given an order quantity $\mathbf{Q} \in \mathbb{R}^4_+$, we can characterize a mapping from an arbitrary demand vector $\mathbf{D} \in \mathbb{R}^4_+$ to a transshipment matrix $\mathbf{T^{4x4}}$. Define $x^+ = \max(x, 0)$. From Propositions 2 and 3 and the corresponding results for the case of total shortage inventory, the optimal transshipment policy can be expressed as follows for the various cost structures of relevance.

(a) If $\forall i \quad D_i \leq Q_i$ then $\quad T_{ij} = 0 \quad$ i,j=1,2,3,4; $\quad i \neq j$
(b) If $\forall i \quad D_i \geq Q_i$ then $\quad T_{ij} = 0 \quad$ i,j=1,2,3,4; $\quad i \neq j$
(c) If $D_i < Q_i, D_j > Q_j, D_k > Q_k$ and $D_l > Q_l$ then for $a_{ij} > a_{ik} > a_{il}$

$$T_{ij} = \min(\bar{Q}_i, \bar{D}_j)$$
$$T_{ik} = \min(\bar{Q}_i - T_{ij}, \bar{D}_k)$$
$$T_{il} = \min(\bar{Q}_i - T_{ij} - T_{ik}, \bar{D}_l)$$

(d) If $D_i < Q_i, D_j < Q_j, D_k > Q_k, D_l > Q_l$ and $\sum_{t=1}^{4} D_t < \sum_{t=1}^{4} Q_t$ then for $a_{ik} > a_{jk}$ and $a_{il} < a_{jl}$

$$T_{ik} = \min(\bar{Q}_i, \bar{D}_k)$$
$$T_{jk} = (\bar{D}_k - Q_i)^+$$
$$T_{jl} = \min(\bar{Q}_j, \bar{D}_l)$$
$$T_{il} = (\bar{D}_l - Q_j)^+$$

and for $a_{ik} > a_{jk}, a_{il} > a_{jl}$ and $|a_{ik} - a_{jk}| > |a_{il} - a_{jl}|$

$$T_{ik} = \min(\bar{Q}_i, \bar{D}_k)$$
$$T_{jk} = (\bar{D}_k - Q_i)^+$$
$$T_{il} = \min(\bar{Q}_i - T_{ik}, \bar{D}_l)$$
$$T_{jl} = \bar{D}_l - T_{il}$$

(e) If $D_i < Q_i, D_j < Q_j, D_k > Q_k, D_l > Q_l$ and $\sum_{t=1}^{4} D_t > \sum_{t=1}^{4} Q_t$ then for $a_{ik} > a_{il}$ and $a_{jk} < a_{jl}$

$$T_{ik} = \min(\bar{Q}_i, \bar{D}_k)$$
$$T_{il} = (\bar{Q}_i - D_k)^+$$
$$T_{jl} = \min(\bar{Q}_j, \bar{D}_l)$$
$$T_{jk} = (\bar{Q}_j - D_l)^+$$

and for $a_{ik} > a_{il}, a_{jk} > a_{jl}$ and $|a_{ik} - a_{il}| > |a_{jk} - a_{jl}|$

$$T_{ik} = \min(\bar{Q}_i, \bar{D}_k)$$
$$T_{il} = (\bar{Q}_i - D_k)^+$$
$$T_{jk} = \min(\bar{Q}_j, \bar{D}_k - T_{ik})$$
$$T_{jl} = \bar{Q}_l - T_{jk}$$

(f) If $D_i < Q_i$, $D_j < Q_j$, $D_k < Q_k$ and $D_l > Q_l$ then for $a_{il} > a_{jl} > a_{kl}$

$$T_{il} = \min(\bar{Q}_i, \bar{D}_l)$$
$$T_{jl} = \min(\bar{Q}_j, \bar{D}_l - T_{il})$$
$$T_{kl} = \min(\bar{Q}_k, \bar{D}_l - T_{il} - T_{jl})$$

For (c)–(f) we have indexes $i, j, k, l = 1, 2, 3, 4; \quad i \neq j \neq k \neq l$.

The transshipments that are not explicitly defined in the various events are equal to zero. There are no positive transshipment T_{ij} $(i \neq j)$ in (a) and (b) because of the complete pooling assumptions. Note that a greedy allocation will be optimal in (c) and (f). For (d) and (e) the optimality of a greedy allocation will depend on the given cost parameters, the order quantities and the demand realizations. Note also that there are two different cost structures of relevance for determining the optimal transshipments in both (d) and (e).

5.3 Optimal Inventory Choices

The optimal order quantity can be determined by solving the first order conditions of the π program. In order to find the first order conditions of the π program we can find the expectations of the transshipments and the corresponding gradients with respect to $\mathbf{Q}$ as we did for the three location model. To get an insight of the complexity for the four location model, Table IV shows the main events and corresponding cost structures of relevance for $\hat{T}_{13}$. Note that each event must be divided into sub-events in order to characterize the unique optimality regions of interest. However, due to the large number of calculations required we will not go any further here, only refer to the three location model regarding how to characterize the optimal order quantity.

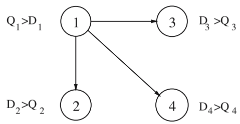
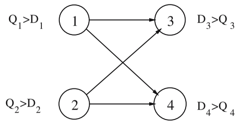
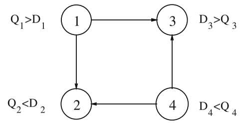
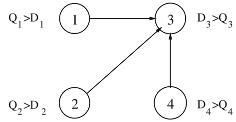
6 Transshipments with *N* Locations

By determining the optimal transshipment policy we have showed how to characterize the optimal order quantity in an analytical manner for up to four locations (given that we can calculate the probability integrals exactly). We have not been able to determine the optimal transshipment policy for the most general setting of n locations with unidentical cost parameters. In this general case an optimal transshipment policy can be determined for some specific cost structures which will be outlined in Section 6.1.

6.1 Restrictions on the Cost Parameters

In Proposition 5 below, we state the necessary and sufficient conditions on the cost structure for a greedy algorithm to be optimal in the transshipment problem. This result is particularly attractive because it will enable us to characterize the optimal order quantity for some specific cost structure for the n location model, which has not been possible before. Also, we will be able to identify for which cases the linear programming problems in the gradient based heuristic can be solved in linear time.

Table IV Events and corresponding cost structures of relevance for $\hat{T}_{13}$

 Event 1	 Event 2	 Event 3	 Event 4
	$E_{2.1}$ (surplus): $\sum_{t=1}^{4} D_t < \sum_{t=1}^{4} Q_t$	$E_{3.1}$ (surplus): $\sum_{t=1}^{4} D_t < \sum_{t=1}^{4} Q_t$	
$a_{13} > a_{14}$ & $a_{13} > a_{12}$	$\lvert a_{13} - a_{23}\rvert > \lvert a_{14} - a_{24}\rvert$ & $a_{13} > a_{23}$	$\lvert a_{13} - a_{43}\rvert > \lvert a_{12} - a_{42}\rvert$ & $a_{13} > a_{43}$	$a_{13} > a_{23}$ & $a_{13} > a_{43}$
$a_{13} > a_{14}$ & $a_{13} < a_{12}$	$\lvert a_{13} - a_{23}\rvert > \lvert a_{14} - a_{24}\rvert$ & $a_{13} < a_{23}$	$\lvert a_{13} - a_{43}\rvert > \lvert a_{12} - a_{42}\rvert$ & $a_{13} < a_{43}$	$a_{13} > a_{23}$ & $a_{13} < a_{43}$
$a_{13} < a_{14}$ & $a_{13} > a_{12}$	$\lvert a_{13} - a_{23}\rvert < \lvert a_{14} - a_{24}\rvert$ & $a_{14} > a_{24}$ & $a_{13} > a_{23}$	$\lvert a_{13} - a_{43}\rvert < \lvert a_{12} - a_{42}\rvert$ & $a_{12} > a_{42}$ & $a_{13} > a_{43}$	$a_{13} < a_{23}$ & $a_{13} > a_{43}$
$a_{13} < a_{14}$ & $a_{13} < a_{12}$	$\lvert a_{13} - a_{23}\rvert < \lvert a_{14} - a_{24}\rvert$ & $a_{14} > a_{24}$ & $a_{13} < a_{23}$	$\lvert a_{13} - a_{43}\rvert < \lvert a_{12} - a_{42}\rvert$ & $a_{12} > a_{42}$ & $a_{13} < a_{43}$	$a_{13} < a_{23}$ & $a_{13} < a_{43}$
	$\lvert a_{13} - a_{23}\rvert < \lvert a_{14} - a_{24}\rvert$ & $a_{14} < a_{24}$ & $a_{13} > a_{23}$	$\lvert a_{13} - a_{43}\rvert < \lvert a_{12} - a_{42}\rvert$ & $a_{12} < a_{42}$ & $a_{13} > a_{43}$	
	$\lvert a_{13} - a_{23}\rvert < \lvert a_{14} - a_{24}\rvert$ & $a_{14} < a_{24}$ & $a_{13} < a_{23}$	$\lvert a_{13} - a_{43}\rvert < \lvert a_{12} - a_{42}\rvert$ & $a_{12} < a_{42}$ & $a_{13} < a_{43}$	
	$E_{2.2}$ (shortage): $\sum_{t=1}^{4} D_t > \sum_{t=1}^{4} Q_t$	$E_{3.2}$ (shortage): $\sum_{t=1}^{4} D_t > \sum_{t=1}^{4} Q_t$	
	$\lvert a_{13} - a_{14}\rvert > \lvert a_{23} - a_{24}\rvert$ & $a_{13} > a_{14}$	$\lvert a_{13} - a_{12}\rvert > \lvert a_{43} - a_{42}\rvert$ & $a_{13} > a_{12}$	
	$\lvert a_{13} - a_{14}\rvert > \lvert a_{23} - a_{24}\rvert$ & $a_{13} < a_{14}$	$\lvert a_{13} - a_{12}\rvert > \lvert a_{43} - a_{42}\rvert$ & $a_{13} < a_{12}$	
	$\lvert a_{13} - a_{14}\rvert < \lvert a_{23} - a_{24}\rvert$ & $a_{23} > a_{24}$ & $a_{13} > a_{14}$	$\lvert a_{13} - a_{12}\rvert < \lvert a_{43} - a_{42}\rvert$ & $a_{43} > a_{42}$ & $a_{13} > a_{12}$	
	$\lvert a_{13} - a_{14}\rvert < \lvert a_{23} - a_{24}\rvert$ & $a_{23} > a_{24}$ & $a_{13} < a_{14}$	$\lvert a_{13} - a_{12}\rvert < \lvert a_{43} - a_{42}\rvert$ & $a_{43} > a_{42}$ & $a_{13} < a_{12}$	
	$\lvert a_{13} - a_{14}\rvert < \lvert a_{23} - a_{24}\rvert$ & $a_{23} < a_{24}$ & $a_{13} > a_{14}$	$\lvert a_{13} - a_{12}\rvert < \lvert a_{43} - a_{42}\rvert$ & $a_{43} < a_{42}$ & $a_{13} > a_{12}$	
	$\lvert a_{13} - a_{14}\rvert < \lvert a_{23} - a_{24}\rvert$ & $a_{23} < a_{24}$ & $a_{13} < a_{14}$	$\lvert a_{13} - a_{12}\rvert < \lvert a_{43} - a_{42}\rvert$ & $a_{43} < a_{42}$ & $a_{13} < a_{12}$	

This means that there will exist order quantum and demand realizations such that a greedy algorithm is not optimal given that the corresponding cost structure does not satisfy the necessary and sufficient conditions.

The essential idea behind Proposition 5 stems from Monge [15] and was used by Hoffman [10] to determine the corresponding conditions for the *balanced* ($\sum_i Q_i = \sum_i D_i$) transshipment problem. We have extended these conditions to the *unbalanced* ($\sum_i Q_i \neq \sum_i D_i$) transshipment problem.

There have been some approaches in determining a fast algorithm for the unbalanced case (see Aggarwal et al. [1] but these are in different ways depending on the order quantity and demand at the locations. Blum et al. [2] have identified *sufficient* conditions for which a greedy algorithm is optimal for a related unbalanced weighted bipartite matching problem. In fact, it turns out that these conditions are equivalent to those we find to be *necessary* and *sufficient* for an unbalanced transshipment problem.

Proposition 5 The optimal transshipment policy for all order quantities and demand realizations are in the form of a greedy allocation if and only if the indices of the cost structure satisfy the following property

$$a_{ij} + a_{rs} \geq a_{is} + a_{rj} \quad \text{whenever} \quad \max\{a_{ij}, a_{rs}\} \geq \max\{a_{is}, a_{rj}\} \tag{30}$$

Proof Assume we have a solution $z = \sum_i \sum_j a_{ij} T_{ij}$ where a_{ij} is the maximum entry in matrix $\mathbf{A}$ and where the corresponding transshipment variable T_{ij} is less than the maximal feasible value i.e. $T_{ij} < \min(Q_i, D_j)$. Consider then the feasible solution z^n where $T^n_{ij} = \min(Q_i, D_j)$. We want to construct z^n from z such that $z^n \geq z$.

Note first that in the cases of total surplus and total shortage inventory there will, respectively, exist indices $r \neq i$ such that $T_{rj} = \delta_1 > 0$ and $s \neq j$ such that $T_{is} = \epsilon_1 > 0$ (otherwise we will have $T_{ij} = \min(Q_i, D_j)$).

Assume now that there exist indices $r \neq i$ and $s \neq j$ such that $\min(T_{rj}, T_{is}) > 0$ and let $\alpha_1 = \min(Q_i, D_j) - T_{ij}$. We can then construct the following feasible solution

$$T^1_{ij} = T_{ij} + \min(\delta_1, \epsilon_1, \alpha_1)$$

$$T^1_{rj} = T_{rj} - \min(\delta_1, \epsilon_1, \alpha_1)$$

$$T^1_{rs} = T_{rs} + \min(\delta_1, \epsilon_1, \alpha_1)$$

$$T^1_{is} = T_{is} - \min(\delta_1, \epsilon_1, \alpha_1)$$

where $T^1_{mn} = T_{mn}$ for all other pairs of indices (m, n). Equation 30 then ensures that $z^1 \geq z$. If $T^1_{ij} < \min(Q_i, D_j)$ and there still exist both indices $r \neq i$ such that $T^1_{rj} = \delta_2 > 0$ and $s \neq j$ such that $T^1_{is} = \epsilon_2 > 0$, we can construct z^2 in a similar way as z^1 such that $z^2 \geq z^1 \geq z$. By continuing this way one of the following three cases will occur

(*) There is a total surplus inventory and there exists no index $s \neq j$ such that $T^n_{is} > 0$, but there may exist indices $r \neq i$ such that $T^n_{rj} > 0$.

(**) There is a total shortage inventory and there exists no index $r \neq i$ such that $T^n_{rj} >= 0$, but there may exist indices $s \neq j$ such that $T^n_{is} > 0$.

(***) $T^n_{ij} = \min(Q_i, D_j)$.

We can then construct $z^n \geq z$ in (*) by iteratively reducing the variable T_{rj} ($r \neq i$), starting with the variables corresponding to the lower a_{rj}, and increase T^n_{ij} until

$T_{ij}^n = \min(Q_i, D_j)$. Keeping the other $T_{..}^n$ variables unchanged ensures that $z^n \geq z^{n-1} \geq \ldots \geq z$ (due to the second part of Eq. 30). An analogous argument can be made for (**) as well.

Thus we have shown that among the possible optimal solutions for the transshipment problem, there is one with the maximal feasible value possible on the transshipment variable corresponding to the maximal entry of the cost matrix **A**, i.e., $z^n \geq z$ where $T_{ij}^n = \min(Q_i, D_j)$ when $a_{ij} = \max_{m,n}\{a_{mn}\}$.

This will reduce our N location problem to an $N-1$ location problem. Because the conditions stated in Eq. 30 are hereditary (by deleting the entries of the "reduced" location from the cost matrix), we can use induction on N to complete the proof.

The "only if" part of Proposition 5 is trivial to see from the construction of our proof. Note that the optimality of $T_{ij} = \min(Q_i, D_j)$ when $a_{ij} = \max_{m,n}\{a_{mn}\}$ is *only* ensured if we have Eq. 30 in the unbalanced case. Otherwise a "cycle"-procedure can easily be used to show that there will exist **Q,D** such that $T_{ij} < \min(Q_i, D_j)$ in optimum. □

Think of $a_{ij} + a_{rs}$ and $a_{is} + a_{rj}$ as the larger and smaller diagonal of the corresponding 2-2 submatrix of a cost matrix **A**. In words Proposition 5 then says that the larger diagonal of any 2-2 submatrix has to contain the largest entry of that submatrix in order for a greedy algorithm to be optimal.

It can be shown that if a (rearranged) cost matrix satisfies something called the Monge property (which is the property needed for the balanced transshipment problem), the cost structure in addition has to satisfy a linear ordering of the locations in order for a greedy algorithm to be optimal. This ordering is such that all surplus inventory locations will rank the shortage inventory locations in the same order, and all shortage inventory locations will rank the surplus inventory locations in the same order. If the cost matrix has the Monge property but does not have this linear ordering among the locations, then a greedy algorithm will not necessarily be optimal. The optimality will then depend on the order quantities and demand realizations as well.

6.1.1 Special Cases

In the case of general revenue r_i, salvage s_i and penalty costs p_i but identical transshipment cost $\tau_{ij} = \tau$, we will have $a_{ij} + a_{rs} = a_{is} + a_{rj}$ as can be seen from Eqs. 31 and 32

$$\begin{aligned} a_{ij} + a_{rs} &= r_j + p_j - \tau - s_i + r_s + p_s - \tau - s_r \\ &= r_j + r_s + p_j + p_s - 2\tau - s_i - s_r \end{aligned} \tag{31}$$

$$\begin{aligned} a_{is} + a_{rj} &= r_s + p_s - \tau - s_i + r_j + p_j - \tau - s_r \\ &= r_j + r_s + p_j + p_s - 2\tau - s_i - s_r \end{aligned} \tag{32}$$

Due to Proposition 5 a greedy transshipment policy will therefore be optimal for the case of identical transshipment costs since the resulting cost structure will satisfy Eq. 30.

Cost structures with identical transshipment costs across the locations will be relevant in many real life settings. A large company with a number of stores sited in a greater city will most likely have a contract with a local delivery service where the transshipment costs can be considered as constant and therefore identical across the locations. Note that for this setting, it would probably be plausible to also assume identical revenue, penalty and salvage costs as well, due to the close proximity of the locations. However for settings where transshipments are made by an overnight delivery service to a greater area, the assumption of identical transshipment costs will be much more plausible than the assumption of identical revenue, penalty and salvage costs. Note also that the transshipment costs are generally very small compared to for instance the revenue value. In practice, the transshipment costs can therefore for some settings be considered identical across the locations.

For some settings where the transshipment costs in fact are unequal across the locations, the differences often stem from the different handling costs of physically sending or receiving the products at the store. In these settings you can model the transshipment costs as $\tau_{ij} = u_i + v_j + w$ where u_i and v_j respectively can be interpreted as the physical costs of sending and receiving at the store while w can be seen as the approximately identical cost of moving the product for an overnight delivery service. It can easily be shown that these cost structures satisfy Eq. 30 since we also for these cases will have $a_{ij} + a_{rs} = a_{is} + a_{rj}$. It follows then from Proposition 5 that a greedy transshipment policy will be optimal.

7 Future Work

A natural extension of this work is to utilize the analytical insight for small dimensions gained in this paper, to develop efficient heuristics for large size problems. Note that for large size problems you would use a gradient search procedure with a *huge* number of transshipment problems to be solved in every gradient step. Thus, the solution time would be dramatically reduced if we for instance were to use a greedy transshipment policy. This is because every transshipment problem in every gradient step could be solved in linear time by sorting the entries of the cost matrix *once*.

It would also be of interest to compare the performance of a gradient search heuristic based on a greedy allocation of transshipments in our K program versus an optimal allocation. Even though the latter heuristic in expectation would perform strictly better, it might not justify the significant increase in solution time. Also, the greedy based heuristic would be capable of solving much larger problems.

8 Conclusions

In this paper, we consider a multi-location inventory system where transshipments are allowed as recourse actions in order to reduce the costs of surplus or shortage inventory after demand are realized. We have presented a new formulation for this problem which is simpler and more intuitive while it still incorporates all of the complexities of previous models for this problem (see [17] and [18]). Our bipartite transportation network formulation has enabled us to gain analytical insight into problems of higher dimensions than has been achieved earlier. We have characterized the conditions for the optimal order quantities for three locations, which previously only has been done for the two location model in the case of general cost parameters. We have also shown how to characterize the corresponding conditions for the four location model by determining the optimal transshipment policy and using the same approach as for the three location model. For the n location model we have presented an optimal transshipment policy for some specific cost structures that will enable us to characterize the optimal order quantity for these settings as well.

The insight gained from the analytical results lead us to examine the optimality conditions of a greedy transshipment policy. We have stated that this heuristic will be optimal for the two and three location model. For the four location model we have characterized the conditions for which the heuristic is optimal. The main advantage of the greedy heuristic is its simplicity. There will be no memory problems, as opposed to using very large linear programming problems, and of course it will be much faster. The greedy heuristic can therefore be an alternative for models where a large number of locations and demand samples (to approximate the demand distribution) are considered.

Acknowledgements We would like to thank Nils Rudi for valuable suggestions and discussions, and Sigrid Lise Nonås for helpful comments. We would also like to thank Magnar Brekke for help with English grammar corrections.

References

1. Aggarwal, A., Bar-Noy, A., Khuller, S., Kravets, D., Schieber, B.: Efficient minimum cost matching using quadrangle inequality. In: Proceedings of 33rd Annual IEEE Symposium on Foundations of Computer Science, pp. 583–592. Pittsburgh, PN (1992)
2. Blum, A., Jiang, T., Li, M., Tromp, J., Yannakakis, M.: Linear approximation of shortest superstrings. In: Proceedings of the 23rd Annual ACM Symposium on the Theory of Computing, pp. 328–336. ACM, New York, NY (1991)
3. Bradley, S.P., Hax, A.C., Magnanti, T.L.: Applied Mathematical Programming. Addison-Wesley, Reading, MA (1977)
4. Chang, P., Lin, C.: On the effect of centralization of expected costs in a multi-location newsboy problem. J. Oper. Res. Soc. **42**(11), 1025–1030 (1991)
5. Dong, L., Rudi, N.: Supply chain interaction under transshipments. Working paper. University of Rochester, NY 14627, U.S.A. (2000)
6. Fisher, M.L., Hammond, J.H., Obermeyer, W.R., Raman, A.: Making supply meet demand in an uncertain world. Harvard Bus. Rev. **72**(3), 83–93 (May-June 1994)
7. Herer, Y., Rashit, A.: Lateral stock transshipments in a two-location inventory system with fixed replenishment costs. Department of Industrial Engineering, Tel Aviv University Nav. Res. Logist. **46**(5), 525–547 (1999a)
8. Herer, Y., Rashit, A.: Policies in a general two-location infinite horizon inventory system with lateral stock transshipments. Working paper. Department of Industrial Engineering, Tel Aviv University (1999b)
9. Herer, Y., Tzur, M., Yücesan, E.: The multi-location transshipment problem. Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel Nav. Res. Logist. **48**, 386–408 (2001)
10. Hoffman, A.J.: On simple linear programming problems. In: Klee, V. (ed.) Proceedings of Symposia in Pure Mathematics, Convexity, vol. 7, pp. 317–327. Am. Math. Soc., Providence, RI (1963)
11. Jönsson, H., Silver, E.A.: Analysis of a two-echelon inventory control system with complete redistribution. Manage. Sci. **33**, 215–227 (1987)
12. Karmarkar, U.S.: The multilocation multiperiod inventory problem: Bounds and approximations. Manage. Sci. **33**(1), 86–94 (1987)
13. Karmarkar, U.S., Patel, N.R.: The one-period N-location distribution problem. Nav. Res. Logist. **24**, 559–575 (1977)

14. Krishnan, K., Rao, V.: Inventory control in N warehouses. J. Ind. Eng. **16**, 212–215 (1965)
15. Monge, G.: Deblai et remblai, Memoires de l'Academie des Sciences (1781)
16. Porteus, E.L.: Stochastic inventory theory. In: Heyman, D.P., Sobel, M.J. (eds.) Handbooks in OR and MS, vol. 2, pp. 605–652. Elsevier, Amsterdam, The Netherlands (1990)
17. Robinson, L.W.: Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. Oper. Res. **38**, 278–295 (1990)
18. Rudi, N., Kapur, S., Pyke, D.: A two-location inventory model with transshipment and local decision making. Manage. Sci. **47**(12), 1668–1680 (2001)
19. Tagaras, G.: Effects of pooling on the optimization and service levels of two-location inventory systems. IIE Trans. **21**, 250–257 (1989)
20. Tayur, S.: Computing optimal stock levels for common components in an assembly system. Working paper. Carnegie Mellon University (1995)