THE INFLUENCE OF INDIVIDUAL COST FACTORS ON THE USE OF EMERGENCY TRANSSHIPMENTS

PAUL M. NEEDHAM
Industrial College of the Armed Forces, National Defense University, Washington, DC, 20319, U.S.A.

and

PHILIP T. EVERST
College of Business and Management, University of Maryland, College Park, MD, 20742, U.S.A.

(Received 11 August 1997; in revised form 8 February 1998)

Abstract—Emergency transshipments represent one way in which logistics managers can reduce inventories while simultaneously maintaining customer service levels. This paper examines the cost implications of these actions and offers a straightforward approach to analyzing their suitability. Simulation results indicate that the cost of a stockout is the primary determinant in the transshipment decision, with higher stockout cost levels generally increasing the likelihood that transshipment usage will lead to lower overall cost. A metamodel is proposed as a practical means of providing insight into when emergency transshipments should be employed. © 1998 Elsevier Science Ltd. All rights reserved

Keywords: Inventory control, metamodel, simulation, transshipments.

1. INTRODUCTION

Inventory is held to provide a specified level of product availability for either internal use or resale. However, inventory often represents a significant level of tied-up capital. If inventory can be reduced while maintaining adequate customer service levels, firms may be able to significantly lower costs and obtain a competitive advantage. One method that organizations may use to effect a reduction in inventory levels while holding service levels stable is the implementation of a transshipment policy. Transshipments can refer to a variety of different management actions that result in the sharing of inventory among locations, possibly at different levels of the supply chain. This study defines a transshipment as a sharing of items at the lowest (i.e. retail) level of the chain.

The purpose of a transshipment is to realign inventory balances to ensure the right quantities are available in the right location to satisfy either expected demand or backorders. One location may have customer demand for an item but no inventory while another location may have one or more items on hand and no demand at present nor expected in the near future. A transshipment could then be used to transfer items from the location with inventory to the location that is out of stock in order to meet demand and effectively use inventory. Indeed, the demand causing the transshipment may have actually been expected based on an aggregate forecast for items within the system as a whole, but the exact location of the demand could not be pinpointed.

Of course, there are costs associated with transshipping items between locations. The chief cost is the transportation that would not have been required otherwise. Additional costs, such as those associated with rehandling and with the information system needed to support transshipments, may also be incurred. However, these costs are at least partially offset by the reduction in required inventories and the avoidance of costs associated with reactions to a stockout—for example, lost sales, lost customers or production shutdowns.

*Author for correspondence. Fax: 001 301 405 0146; e-mail: pevers@mbs.umd.edu
This research examines the interaction of relevant costs and transshipment policies and presents a method for determining the point at which the benefits of transshipments outweigh their costs. First, the relevant literature is reviewed. Then a simulation model of a multi-echelon, multi-location inventory system is described along with its experimental design. Results of the simulation and sensitivity analysis identify the relevant costs drivers and are used to construct a decision-making tool for managers contemplating the implementation of transshipments. Finally, conclusions and managerial implications are presented.

2. LITERATURE REVIEW

Because demand at individual locations typically fluctuates and cannot be forecasted exactly, managers may employ various risk-pooling methods to reduce the inherent variation. One method of reducing demand uncertainty is through the consolidation of stock-keeping locations. Eppen and Schrage (1981) found that using supply depots to consolidate incoming shipments resulted in statistical economies of scale and, consequently, lower inventory requirements. Maister (1976) developed the square root law to show that a decrease in inventory locations resulted in lower inventory requirements. Zinn et al. (1989) introduced the portfolio effect model, a more general model than the square root law, to measure the percentage reduction in safety stocks achievable through inventory consolidation. Evers and Beier (1993) extended the portfolio effect model to an even more generalized form by incorporating fluctuations in lead time.

Another method of risk-pooling is the use of transshipments between stock-keeping locations. The use of emergency transshipments to satisfy out-of-stock conditions and of planned transshipments as a deliberate operating condition have been found to increase customer service levels. Numerous researchers have developed analytical models to examine emergency transshipments in periodic review inventory control systems, including Gross (1963), Krishnan and Rao (1965), Hoadley and Heyman (1977), Cohen et al. (1986), Tagaras (1989), Robinson (1990), Chang and Lin (1991) and Tagaras and Cohen (1992). Notable among these is Chang and Lin who incorporated cost considerations into the analysis. Lee (1987), Axsater (1990), Dada (1992) and Sherbrooke (1992) examined the impact of emergency transshipments in continuous review systems and found that the use of transshipments could possibly lead to significant backorder reductions and, accordingly, high service levels. Evers (1997) determined that the benefits of emergency transshipments extended beyond demand pooling to include lead time pooling as well. Evers (1996) considered planned transshipments, where each retail activity was supplied by multiple, separate distribution centers. The impact of this conscious policy was that overall demand variance was decreased and system safety stock required to maintain the desired level of inventory support was reduced.

Cantagalli (1987) evaluated the impact of four different emergency transshipment policies using the (s,S) inventory system, where an order is placed to bring the inventory level up to the desired maximum stock level (S) when the inventory on hand is equal to or less than reorder point (s). These transshipment policies can be classified as either complete pooling or partial pooling and form the basis of the transshipment rules examined later in this report. Though (s,S) systems are in use, they tend to be difficult to work with in terms of establishing the control parameters (Silver and Peterson, 1985). As a result, this study considers a reorder quantity, reorder point (Q,R) system instead, where an order of fixed size (Q) is placed whenever the reorder point (R) is reached. The (Q,R) system is a very common and relatively straightforward system whose primary drawback is associated with demands of appreciable magnitude (Silver and Peterson, 1985).

Transshipments increase transportation costs; therefore, the impact of this increase needs to be considered in determining the cost-effectiveness of variance reduction through transshipments. Transportation-inventory tradeoffs were demonstrated early on by Baumol and Vinod (1970), showing that the use of premium transportation could lead to potential savings in inventory holding costs. A substantial number of studies on this topic have since been reported (c.f. Constable and Whybark, 1978; and Blumenfeld et al., 1985)—indeed, a more general transportation–inventory–production tradeoff has also been analyzed (Blumenfeld et al., 1985). Tyworth (1991) reviewed a series of these transportation–inventory models and recommended improvements in their flexibility and cost-effectiveness, especially in their ability to model various probability distributions (see also Tyworth, 1992). While in general these studies found that, by reducing the
variance of any stochastic process in the system (such as lead time and demand) potential savings resulted, they did not consider the use of transshipments.

This research attempts to fill a gap in the literature by incorporating cost into the decision on whether to use emergency transshipments in a \((Q,R)\) system. At the same time, a method for analyzing this decision is also suggested.

3. THE SIMULATION MODEL

3.1. Model structure

In this study, a multi-echelon, multi-location inventory system with two warehouses, each supporting three retail locations, is examined. A lead buyer monitors information flows, directs shipments and centrally controls stock levels for the entire system. This system is similar to that used by some large retailers where the location of the buyer is independent of the stocking and selling locations. It is also comparable to those used by the military for many items, substituting depots, bases and materials managers for warehouses, retailers, and buyers, respectively.

In this system, expedited shipments from warehouses are attempted before transshipments so as not to reduce inventory availability at the other retail locations. Thus, if random customer demand for an item cannot be completely filled from stock at the retail location where it arises, the first option is an expedited (overnight) shipment of enough product to cover the customer’s request from the retailer’s assigned warehouse. If the assigned warehouse cannot fill the entire emergency request, the second option is an expedited (overnight) shipment of the remaining unfilled amount from the other warehouse. The third alternative is to request (overnight) transshipment of any remaining unfilled quantities from other retailer locations within the region. The final option is to request (second day) transshipment of product from retailers in the other region. It is assumed that customers are willing to wait up to 2 days for an item if they know it will be provided on an emergency basis.

A sequential procedure is used to determine from which retail center a transshipment will be requested first. That is, within a region, retail outlets are numbered 1, 2 and 3 and transshipments are requested from the lowest numbered location first. (Therefore, if either location 2 or 3 faces demand but is out of stock, it contacts 1 first—while location 1 would contact 2 first.) This method can be used to assign priority designations to the retail stores. Alternatively, a distance criterion could have been used; however, with three locations, the numerical designation serves the same purpose (note that the facility nearest to the other two facilities could be designated location 1, with the facility closest to 1 designated location 2). If the requested inventory is not available at a retail center in the same region as the initial demand, the search continues into the other region, again in a sequential fashion.

If a single retail store cannot be found that has sufficient inventory on hand to fill the transshipment request, a search begins for two stores that in combination has the inventory. First, the two other retail outlets in the region where the initial demand arose are considered. If the request cannot be completely satisfied, the first store is retained (based on numerical order) and the first store in the other region is considered. The model logic continues to compare retail centers until either the request can be satisfied from the inventory of two locations or all two-way combinations of retail centers have been checked. If the two-way combinations are unable to fill the transshipment request, then three-, four-, and five-way combinations are considered in a similar fashion. If no combination of locations satisfies the demand that remains after expedited shipments from the warehouses take place, no transshipments are made and the unfilled portion of demand is lost. In other words, if the entire order cannot be filled, customers are assumed to prefer going elsewhere for the remainder of the order over receiving a number of transshipments that in combination still does not completely fill their requirements. Therefore, the rule used in the model is to satisfy from transshipments either the full remaining demand or none of it.

Along with a baseline policy of no transshipments, four alternative transshipment policies covering both partial and complete pooling are considered. In the baseline case, expedited shipments from either warehouse are allowed, but emergency transshipments between retailers are not. In the case of partial pooling, one option is where a retailer is allowed to transship any inventory over and above its reorder point (expected demand during lead time plus safety stock). A second type of
partial pooling is where a retailer is allowed to transship any inventory over and above its safety stock. A third type of partial pooling is where a retailer is allowed to transship any inventory over and above its expected demand during the next period (in this study, the next day). In the case of complete pooling, a retailer is allowed to transship all of its inventory.

Once a retailer has reached its reorder point because of either normal demands from customers or emergency transshipment requests from other retailers, it places a standard reorder with its assigned warehouse. Lead times for these regular retailer reorders vary. If the warehouse has sufficient inventory on hand, it ships the desired reorder quantity; if it does not have sufficient inventory, it ships whatever it can (if any) and backorders the rest—filling all backorders on a first-come-first-serve basis. Once a warehouse reaches its reorder point (because of either normal demands from assigned retailers or expedited shipment requests from retailers), it places a reorder with an outside vendor having sufficient capacity to fill all orders in exactly 5 days.

The standard reorder point, reorder quantity \((Q,R)\) approach is used to control inventory levels at both the retail centers and the warehouses. Safety stocks at the retail centers are based on the standard equation:

\[
SS = k\sqrt{L\sigma_D^2 + \sigma_L^2D^2}
\]

where:

- \(SS\) = safety stock,
- \(k\) = safety factor associated with a desired level of inventory availability,
- \(L\) = average lead time,
- \(\sigma_L\) = standard deviation of lead time,
- \(D\) = average demand during one time period, and
- \(\sigma_D\) = standard deviation of demand during one time period.

No safety stocks are held at the warehouses.

The simulation model was written in SIMAN (Pegden et al., 1990), a high-level simulation language, and validated from a military perspective by using three inventory system defense experts to review the simulation logic and output from an experimental replication. They provided guidance in ensuring that the system was properly designed, the distribution operations properly described, the reorder processes and inventory calculations reasonable, and the transshipment rules (i.e. expedite first, transship next) similar to existing military systems. Verification of the model was performed by manually stepping through the program code to ensure the correct paths were taken. Additional verification was accomplished by tracing entities through actual runs of the simulation model.

### 3.2. Model inputs

A number of input parameters associated with retail demands and lead times had to be established. Mean demand and lead time, as well as fluctuations in demand and lead time, were each examined at three different levels. Along with the measures of central tendency and dispersion, probability distributions associated with both retail demands and lead times were also established. For daily demand, the normal distribution truncated at zero was assumed. Admittedly, a drawback of using the normal distribution is that it is less appropriate for low volume items (Silver and Peterson, 1985); however, it was chosen for multiple reasons: because it does not place restrictions on the values of the mean and variance (compared to, say, the Poisson distribution which requires that they be equal); because its properties are well known; and because it is typically the basis for examining continuous demand. For lead time, the gamma distribution was used since it “appears to be well-suited for modeling lead time because it is defined only for non-negative values and can model positively skewed distributions” (Tyworth, 1991, p. 310).

To ensure consistent comparisons among the three different means, common levels of dispersion were required. This was operationalized in the simulation by using the coefficient of variation, which is equal to the average divided by the standard deviation. The use of the coefficient of variation is preferred because, among other things, it “preserves the orthogonality of the independent
variables in the simulation” (Zinn and Marmorstein, 1990, p. 100), thus ensuring that the dispersion of the distribution is comparable regardless of the mean. Table 1 shows the input parameter values used in the study.

Since the system being modeled is similar to the military’s logistics systems, the military’s baseline $k$ factor of 1 was used. If demand during lead time is assumed to be normally distributed, a $k$ factor of 1 equates to an 84% probability of no stockout during a replenishment cycle. Even if the normal distribution does not apply, a conservative level of service can still be determined based on the use of a distribution-free approach.

### 3.3. Model outputs and costs

Fourteen outputs were collected from the simulation. They included initial demands, unfilled demands, expedited shipments, emergency transshipments, orders placed, inventory levels and other relevant system measures. The simulation outputs were then entered into a spreadsheet model to find the total cost of each system. Consequently, cost data associated with these outputs was needed to determine the total cost associated with each transshipment policy. In this study, the total cost was derived from six distinct components: in-storage inventory carrying costs, in-transit inventory carrying costs, ordering costs, routine transportation costs, rush transportation costs (either expedited or transshipment), and penalty costs associated with stockouts. Other potentially significant transshipment-related costs, such as additional handling and loss and damage costs, were not directly considered (though some of them could be factored into the cost of rush transportation on a per unit basis if desired).

Three different levels of item cost (product value) were considered: low ($30), medium ($60) and high ($100). The in-storage carrying charge was assumed to be 30%/year of the item cost since organizations typically use internal rates of return higher than the cost of capital to evaluate investments. The in-transit carrying charge of 25%/year of the item cost was based on the in-storage carrying charge less 5% to account for the value of storage, shrinkage, damage, and obsolescence not incurred when goods are in transit (c.f. Coyle et al., 1996). The cost of placing an order, $20, was based on the amount a branch of the military uses for computing ordering costs. In lieu of computing a fill rate, the cost of a stockout was assumed to be a penalty cost. Since the actual cost of a stockout can differ widely depending upon the circumstances, a penalty cost of 50% of the item cost was used.

Transportation costs were estimated from actual rates. As inputs, the costs of routine shipping via a less-than-truckload (LTL) carrier and of rush shipping via an overnight carrier were collected for a common part: an electronic circuit assembly weighing 30 pounds in a one cubic foot box being shipped from the east coast to the west coast. Based on this description, the cost of a LTL shipment was approximately equal to $15 per unit, while the cost of an overnight shipment was $45 per unit (or three times as much). Although the transportation cost per unit may be different based on the declared value of an item, it is expected that the ratio of LTL rates to overnight rates remains fairly constant.

### 3.4. Experimental design

The number of simulation replications was based on a full factorial model with four input parameters (each taking on three levels) and five transshipment policies. This resulted in 405 ($3^4 \times 5$) replications. Each of the 405 runs was then evaluated using the three item-cost levels, producing a total of 1215 simulation sets.

Law and Kelton noted that “a very common [but incorrect] mode of operation is to make a single simulation run of somewhat arbitrary length and then to treat the resulting simulation...
estimates as the ‘true’ model characteristics’ (Law and Kelton, 1991, p. 522). This mistake is avoided by recognizing that an inventory system is a non-terminating system and then designing the experiment to evaluate the system in a steady-state condition. Therefore, the experiment must provide sufficient independent observations to allow statistical tests to be used and obtain statistical significance. Cantagalli (1987) used a single long simulation run of 8000 periods, while Evers (1997) used 10,000 periods, to produce enough independent observations. Based on this prior research and examination of trial runs associated with the baseline (no pooling) policy, each of the 405 replications in this study was run for 8000 days (periods).

Because individual observations in a non-terminating inventory simulation from one period to the next are seldom uncorrelated, a method to obtain independent observations is required if standard statistical techniques are to be used to evaluate the output. While several techniques are available, the method used in this analysis was batch means.

The batch means technique is based on a single, long run and as such produces a transient period only once. This warm-up period is identified—in this study, the transient period was deemed equal to 100 time days—and truncated from the total number of observations since it is not representative of the steady state. The remaining periods are then evenly divided into batches that provide a sufficient number of independent observations, each based on the mean of the batch. The major problem with using batch means is the possibility that the means of the batches will also be correlated, thus yielding a severely biased estimator. Law and Kelton (1991) recommended examining the correlation of successive batch means and selecting a batch size that exhibits little correlation to ensure that the resulting sample variance is not biased. A batch size of 200 days was used in this study, resulting in 39 batches per replication and providing an ample number of independent observations with which to perform statistical analyses.

In addition, common random numbers were used across replications to reduce unintended variance. This technique dampened purely random fluctuations across the 405 runs.

4. RESULTS

An initial attempt to evaluate the transshipment policies was performed assuming mean demand at each retail center was equal. The result was so few transshipments between retailers that no meaningful results could be discerned. In order to assess the transshipment policies, average demand at each retail store was allowed to take on a different magnitude, with retail centers 1 and 4 having three times as much, and retail center 2 having twice as much, demand as the other three retail centers (3, 5 and 6). To calculate demand at each retail facility, the magnitudes were multiplied by the random numbers generated from the demand distributions and their corresponding parameters. Use of these magnitudes ensured that some level of transshipments between retail centers would be observed.

Paired t-tests with unequal variances for all pairwise comparisons were used to determine which transshipment policy provided the lowest average total cost. All pairwise comparisons were tested to ensure that all possible combinations were considered. As an example, Table 2 shows the results of all pairwise comparisons for the simulation runs when the item cost is high ($100). At a 99% confidence interval, the baseline policy of no transshipments between retailers is significantly lower cost than either the complete pooling policy or any of the partial pooling policies. Similar results not shown here were obtained for the two other item-cost levels (for convenience and conciseness, only results for the high item-cost level will be shown herein since the results are comparable across levels; see Needham, 1997 for more details).

Since the baseline policy of no transshipments provided the lowest cost in all cases, sensitivity analysis was performed on the output data. Various cost combinations were examined in order to determine the sensitivity of the transshipment policy decision to the cost parameters. A metamodel was also developed to provide insight into the relationship between the system’s total cost and the transshipment policies and to assist in selecting the appropriate transshipment policy to implement.

4.1. Sensitivity analysis

The sensitivity analysis concentrated on finding the cost levels that would change the total system cost associated with each transshipment policy to the point where a policy other than the baseline was the lowest cost. In order to examine combinations of the various costs, the
4.1. Stockout costs. Stockout costs were used to assign a penalty when a customer request could not be filled. The original stockout cost was 50% of the item cost and was selected to represent the potential profit lost of not selling an individual unit. However, there are clearly times when the stockout cost penalty may exceed the item cost because of the value of the end item being supported or because of the possibility of losing not only the current sale but also future sales as well.

Analysis of the simulation results showed that the fill rate for the baseline policy was only 61%, for the partial pooling policy of holding back the reorder point 99%, for the partial pooling policy of holding back the safety stock 99%, for the partial pooling policy of holding back expected demand 94%, and for the complete pooling policy 94%. Because of the dramatic differences in fill rates, stockout costs were expected to affect the transshipment decision. To examine the point where a different transshipment policy would cost less than the baseline policy, if one existed, the stockout cost was allowed to vary as a percentage of the item cost, while all other costs were held constant. Part A of Table 3 shows the average total cost of each transshipment policy when penalty costs range from 50 to 800% of the item cost.

The table indicates that when the stockout cost reaches four times the item cost, the lowest cost policy changes from no transshipments to complete pooling. When the stockout cost reaches five times the item cost, the complete pooling policy remains the lowest, and the partial pooling policies (hold back tomorrow's expected demand, hold back the safety stock, and hold back the reorder point) now have lower costs than the baseline policy as well. Based on a $t$-test, when the stockout cost is seven times the item cost, the complete pooling policy is significantly less costly than the baseline pooling policy. Finally, when the stockout cost is eight times the item cost, the partial pooling policy of holding back the reorder point was found to provide the lowest total cost. From this analysis, it can be concluded that the pooling policies become lowest cost as stockout costs increase because of the differences in fill rates.

4.1.2. Holding cost. Since two holding charges were used in the analysis (in-storage and in-transit), they were moved in tandem to ensure the basic relationship between them was retained. The range of possible values varied from 5% in-transit to 100% in-storage. The expectation was that changes in holding charges would have little effect on the selection of a transshipment policy since inventory levels were initially set to maintain an 84% probability of no stockouts.

Table 3, Part B shows the average total cost of the policies when the in-storage holding charge expressed as an annual percentage of the item cost vary from a low of 10% to a high of 100%
The in-transit charge was equal to the in-storage charge less 5%). The table indicates that the baseline policy remained the lowest total cost as the holding costs changed from 10 to 100% of the item cost. From this, it can be claimed that the decision of which transshipment policy is lowest cost does not vary on the basis of the holding cost.

4.1.3. Transportation cost. The sensitivity of the results to transportation cost changes were examined by allowing both routine and rush transportation costs to vary. Routine transportation costs varied from a low of $1 to a high of $100 per item, while rush shipment costs ranged from a low of $5 to a high of $200 per item. These ranges were not related to one another except that the cost of a routine shipment was required to always be less than the cost of a rush shipment (otherwise, it would be cheaper to use emergency transportation than to use regular transportation for all shipments).

As with the holding charges, changing the transportation costs was not expected to significantly impact the total cost relationship of the transshipment policies. The baseline policy did not allow for any transshipments; therefore, a change in the transshipment cost did not have any impact on the total cost of the baseline policy. The pooling policies employed transshipments which, while reducing stockouts, incur increased costs as a result of transshipping. Part C of Table 3 shows the average total cost of each transshipment policy for six combinations of routine and rush transportation costs. Indeed, the baseline policy remained the lowest cost in all cases.

4.1.4. Ordering cost. Ordering costs in the simulation model were accumulated each time an order was placed from either a retail center or a distribution center. A range of ordering costs was considered from a low of $1 to a high of $100 (the cost of the high cost item). All orders placed, both routine and rush, were assumed to incur the same cost. Since the baseline policy did not

### Table 3. Sensitivity analysis of individual costs (high item-cost runs)

<table>
<thead>
<tr>
<th>(A) Stockout cost</th>
<th>No transshipments</th>
<th>Hold back reorder point</th>
<th>Hold back safety stock</th>
<th>Hold back demand</th>
<th>Hold back nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% (original)</td>
<td>6574^a</td>
<td>9186</td>
<td>9243</td>
<td>9219</td>
<td>8414</td>
</tr>
<tr>
<td>100%</td>
<td>6927^a</td>
<td>9190</td>
<td>9296</td>
<td>9274</td>
<td>8477</td>
</tr>
<tr>
<td>400%</td>
<td>9052</td>
<td>9220</td>
<td>9619</td>
<td>9613</td>
<td>8854^a</td>
</tr>
<tr>
<td>500%</td>
<td>9761</td>
<td>9231</td>
<td>9727</td>
<td>9726</td>
<td>8980^a</td>
</tr>
<tr>
<td>700%</td>
<td>11,177</td>
<td>9251</td>
<td>9942</td>
<td>9952</td>
<td>29231^a</td>
</tr>
<tr>
<td>800%</td>
<td>11,885</td>
<td>9261^a</td>
<td>10,049</td>
<td>10,065</td>
<td>9357</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B) Holding cost</th>
<th>No transshipments</th>
<th>Hold back reorder point</th>
<th>Hold back safety stock</th>
<th>Hold back demand</th>
<th>Hold back nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% (original)</td>
<td>2264^a</td>
<td>3160</td>
<td>3221</td>
<td>3245</td>
<td>2902</td>
</tr>
<tr>
<td>30% (original)</td>
<td>6574^a</td>
<td>9186</td>
<td>9243</td>
<td>9219</td>
<td>8414</td>
</tr>
<tr>
<td>50%</td>
<td>10,884^a</td>
<td>15,211</td>
<td>15,265</td>
<td>15,192</td>
<td>13,927</td>
</tr>
<tr>
<td>80%</td>
<td>17,349^a</td>
<td>24,250</td>
<td>24,298</td>
<td>24,152</td>
<td>22,195</td>
</tr>
<tr>
<td>100%</td>
<td>21,659^a</td>
<td>30,275</td>
<td>30,320</td>
<td>30,126</td>
<td>27,707</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(C) Trans. cost (routine/rush)</th>
<th>No transshipments</th>
<th>Hold back reorder point</th>
<th>Hold back safety stock</th>
<th>Hold back demand</th>
<th>Hold back nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/$5</td>
<td>6324^a</td>
<td>8636</td>
<td>8684</td>
<td>8615</td>
<td>7882</td>
</tr>
<tr>
<td>$1/$45</td>
<td>6324^a</td>
<td>8868</td>
<td>8925</td>
<td>8899</td>
<td>8096</td>
</tr>
<tr>
<td>$15/$45 (original)</td>
<td>6574^a</td>
<td>9186</td>
<td>9243</td>
<td>9219</td>
<td>8414</td>
</tr>
<tr>
<td>$15/$100</td>
<td>6574^a</td>
<td>9505</td>
<td>9574</td>
<td>9609</td>
<td>8709</td>
</tr>
<tr>
<td>$100/$105</td>
<td>8809^a</td>
<td>11,465</td>
<td>11,536</td>
<td>11,583</td>
<td>10,668</td>
</tr>
<tr>
<td>$100/$200</td>
<td>8809^a</td>
<td>12,016</td>
<td>12,108</td>
<td>12,257</td>
<td>11,176</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(D) Ordering cost</th>
<th>No transshipments</th>
<th>Hold back reorder point</th>
<th>Hold back safety stock</th>
<th>Hold back demand</th>
<th>Hold back nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1</td>
<td>6518^a</td>
<td>9046</td>
<td>9099</td>
<td>9075</td>
<td>8327</td>
</tr>
<tr>
<td>$10</td>
<td>6545^a</td>
<td>9112</td>
<td>9167</td>
<td>9143</td>
<td>8369</td>
</tr>
<tr>
<td>$20 (original)</td>
<td>6574^a</td>
<td>9186</td>
<td>9243</td>
<td>9219</td>
<td>8414</td>
</tr>
<tr>
<td>$50</td>
<td>6661^a</td>
<td>9406</td>
<td>9470</td>
<td>9446</td>
<td>8552</td>
</tr>
<tr>
<td>$75</td>
<td>6734^a</td>
<td>9590</td>
<td>9660</td>
<td>9635</td>
<td>8667</td>
</tr>
<tr>
<td>$100</td>
<td>6807^a</td>
<td>9774</td>
<td>9849</td>
<td>9824</td>
<td>8781</td>
</tr>
</tbody>
</table>

^aTransshipment policy with lowest overall mean total cost at specified cost level.

(where the in-transit charge was equal to the in-storage charge less 5%). The table indicates that the baseline policy remained the lowest total cost as the holding costs changed from 10 to 100% of the item cost. From this, it can be claimed that the decision of which transshipment policy is lowest cost does not vary on the basis of the holding cost.
attempt to fill any stockouts using transshipments, the results were expected to be insensitive to changes in the cost of placing an order. Part D of Table 3 shows that the baseline case indeed remains the lowest total cost regardless of order cost level.

4.2. Metamodel

Use of a metamodel is proposed here to assist managers in deciding when to employ transshipments. A metamodel represents an algebraic expression of a simulation model and is commonly generated using least square regression (c.f. Law and Kelton, 1991). It is used to estimate the response surface based on the input parameters without having to repeatedly run the simulation. Since the point of this simulation was to determine which transshipment policy to adopt based on total cost, the dependent variable is the total cost of the system and the level of observation is the individual simulation sets.

The selection of independent variables focused on the controllable input factors: mean demand, mean lead time, the coefficient of variation of demand, the coefficient of variation of lead time, and the five transshipment policies. In order to use transshipment policy as an independent variable, dummy variables were employed. Based on the results of the sensitivity analysis, an additional set of independent variables was used to incorporate the effects of stockout costs. Since stockout costs were found to affect the preferred transshipment policy, interaction terms between the stockout cost level and the transshipment policy were employed as well. For each transshipment policy, stockout costs were examined at five different levels; as a result, a total of 2025 observations ($3^5 \times 5 \times 5$) were used in the regression for each item-cost level. The regression results for the high item-cost level are reported in Table 4.

The overall adjusted $R^2$ of nearly 0.51 indicates that the regression is not extremely precise in terms of forecasting the total cost of the system. However, the regression does provide valuable insight into the behavior of the independent variables. The variables are all of the expected sign and, with the exception of most of the interaction terms, are statistically significant. Each of the four input parameters explains a significant amount of total cost and demonstrates that as either the average or the variability of either demand or lead time increases, total costs increase. Indeed, the two variation measures had great influence on total cost. The dummy variables indicate that, when stockout costs are negligible, the four transshipment policies are inherently more costly than the no pooling policy. This corresponds with the findings of the sensitivity analysis where the baseline policy produced the lowest total cost when the penalty cost was at low levels.

### Table 4. Regression results (high item-cost runs)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>13</td>
<td>31898987809.992</td>
<td>2453768293.076</td>
<td>162.054</td>
</tr>
<tr>
<td>Residual</td>
<td>2011</td>
<td>30449807472.459</td>
<td>15141624.799</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2024</td>
<td>62348795282.442</td>
<td>30449807472.459</td>
<td></td>
</tr>
<tr>
<td>$R^2 = 0.5116$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates:

<table>
<thead>
<tr>
<th>variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>$t$-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5868.247</td>
<td>447.547</td>
<td>-13.112</td>
<td>0.001</td>
</tr>
<tr>
<td>Coef. of var. of lead time</td>
<td>2363.671</td>
<td>234.863</td>
<td>10.064</td>
<td>0.001</td>
</tr>
<tr>
<td>Ave. lead time</td>
<td>1884.973</td>
<td>52.953</td>
<td>35.597</td>
<td>0.001</td>
</tr>
<tr>
<td>Coef. of var. of demand</td>
<td>4571.344</td>
<td>293.391</td>
<td>15.581</td>
<td>0.001</td>
</tr>
<tr>
<td>Average demand</td>
<td>42.194</td>
<td>2.135</td>
<td>19.762</td>
<td>0.001</td>
</tr>
<tr>
<td>Hold back reorder point (dummy)</td>
<td>2961.182</td>
<td>476.902</td>
<td>6.209</td>
<td>0.001</td>
</tr>
<tr>
<td>Hold back safety stock (dummy)</td>
<td>2969.494</td>
<td>476.902</td>
<td>6.227</td>
<td>0.001</td>
</tr>
<tr>
<td>Hold back demand (dummy)</td>
<td>2942.319</td>
<td>476.902</td>
<td>6.170</td>
<td>0.001</td>
</tr>
<tr>
<td>Hold back nothing (dummy)</td>
<td>2131.842</td>
<td>476.902</td>
<td>4.470</td>
<td>0.001</td>
</tr>
<tr>
<td>Stockout cost * no transshipment</td>
<td>7.082</td>
<td>0.789</td>
<td>8.972</td>
<td>0.001</td>
</tr>
<tr>
<td>Stockout cost * reorder point</td>
<td>0.100</td>
<td>0.789</td>
<td>0.126</td>
<td>0.900</td>
</tr>
<tr>
<td>Stockout cost * safety stock</td>
<td>1.075</td>
<td>0.789</td>
<td>1.361</td>
<td>0.174</td>
</tr>
<tr>
<td>Stockout cost * demand</td>
<td>1.129</td>
<td>0.789</td>
<td>1.430</td>
<td>0.153</td>
</tr>
<tr>
<td>Stockout cost * nothing</td>
<td>1.257</td>
<td>0.789</td>
<td>1.592</td>
<td>0.112</td>
</tr>
</tbody>
</table>

Dependent variable = total system cost of simulation set.
Combined with their respective coefficients on the dummy variables, the interaction term coefficients are useful in determining at what stockout cost level, if any, each policy provides the lowest total cost. For example, the coefficient of the baseline policy interaction term (7.08) is almost six times the coefficient of the complete pooling policy interaction term (1.26); however, the complete pooling policy dummy variable has a coefficient of 2131.84. Therefore, the stockout cost level at which the complete pooling policy produces the same total cost as the no pooling policy is 366.30% \((2131.84/[7.08–1.26])\). This indifference point confirms the sensitivity analysis which found that, when the stockout cost level increased to 400%, the complete pooling policy provided the lowest total cost. As will be discussed later, generation of a metamodel producing this type of information can be quite valuable. Management could use it to quickly determine when transshipments should be used and, if so, which type to use.

5. CONCLUSIONS AND MANAGERIAL IMPLICATIONS

This study expands the research in transshipments by actually incorporating cost considerations into the analysis. Transshipment cost tradeoffs have been previously noted in the literature, and this research attempts to specifically account for these concerns. In addition to finding that stockout costs tend to drive the transshipment decision, this research presents a straightforward method for determining the indifference point between using and not using transshipments. The method of analysis presented here can be directly extended to other situations by running a set of simulation experiments with appropriate cost factors and assumptions and then developing a metamodel through linear regression, the results of which can be used to find the indifference point.

5.1. Conclusions

Based on this research, a number of conclusions can be drawn regarding transshipments. First, as might be expected, the primary determinant affecting the decision to use transshipments is the stockout cost level. As penalty costs increase in relation to the cost of the item, the use of transshipments becomes more cost effective. Indeed, the stockout cost was the only cost found to significantly influence the decision, even though transportation, holding, and ordering costs were also considered.

In this research, a penalty cost of roughly 366% was determined to be the point at which the total cost associated with the use of transshipments (more precisely, complete pooling) was equal to the total cost associated with not using transshipments. Above this indifference point, using transshipments resulted in a lower total cost than not using transshipments did (and vice versa below this point). Of course, the actual indifference point will vary depending upon the cost levels considered and assumptions made. Nevertheless, even though the validity of some of the simulation assumptions (such as customer willingness to wait up to 2 days for an emergency delivery) and inputs (such as magnitudes of demand at various locations) used here may be somewhat tenuous, the underlying finding that stockout costs drive transshipment usage should continue to hold.

Another conclusion of the research is confirmation that transshipments dramatically improve inventory availability. The fill rates associated with not using transshipments were much lower than those with using transshipments (61% vs well over 90% based on the input parameters considered in this study). Moreover, the different transshipment policies all produced relatively similar fill rates: the more-sharing policies [hold back nothing (complete pooling) and hold back expected demand (one type of partial pooling)] each had fill rates of 94% while the less-sharing policies [hold back safety stock and hold back reorder point (the other two types of partial pooling considered)] had fill rates of 99%. Therefore, in the case that individual facility managers express concern about making their inventories available to others, some inventory can be held back while still benefitting from transshipments (though total system cost may increase somewhat depending especially upon the cost of a stockout).

Perhaps the most valuable contribution of this research is the development of a metamodel for use as a managerial decision aid. Since stockout costs can be extremely difficult to quantify, a metamodel based on simulated runs of the system in question can be used to identify the point at which transshipments become economically feasible. Management could then compare their
assessment of stockout costs for individual products with the cost of a stockout at the indifference point to determine which items, if any, would be amenable to transshipments.

Additional research is needed to expand the range of cases considered in this study. More generalizable statements regarding the indifference point along with a more encompassing metamodel could be generated from a wider range of item-cost levels and from other methods of inventory control besides the standard reorder point, reorder quantity approach. Relaxation of some of the more onerous simulation assumptions could also be made. For example, using different probability distributions and allowing transshipment times to randomly fluctuate are two possible directions that could be investigated.

5.2. Managerial implications

The most critical element of management’s decision with respect to transshipments is the determination of the cost of a stockout. As both the sensitivity analysis and metamodel indicate, the stockout cost level significantly influences the implementation decision. Indeed, certain managerial implications arise from this research beyond the proposed metamodeling technique.

Where customers require low prices and are willing to incur an occasional stockout or substitute one item for another, the use of transshipments is expected to be minimal. Mass merchants, discount retailers, and grocery stores represent examples of this type. As an illustration, consider a grocery store customer who finds all but a couple of items on the shelf. Typically, the customer will either buy a different size or substitute a competing product. The cost of a stockout in this case would be very minor, since the customer is generally satisfied with the purchase transaction and the store still earns some margin from the substituted items. As a result, the use of transshipments would be rare in these cases.

On the other hand, where customers are willing to pay higher prices for better service, the use of transshipments may be more commonplace. For example, high-end department stores and boutique shops may find that customers who experience a stockout on just one item are inclined to begin shifting their loyalties elsewhere—not just for that item, but for all of their purchases. Here, the stockout cost may be quite high, perhaps multiples of the actual value of the product; thus transshipments may represent a realistic solution.

Similarly, transshipments may be common when an item is used to support another item. Parts for machinery, vehicles, and other equipment are examples of this. The cost of downtime associated with an airplane or bulldozer may be such that a relatively inexpensive part still entails a huge cost if a stockout arises. Therefore, the use of transshipments may again be warranted.

In general then, emergency transshipments represent a potentially useful and cost-effective approach to satisfying unexpected customer demands. Management should consider the use of transshipments, and this paper shows how to employ a metamodel to assist in making the decision.

Acknowledgements—This report stems from the first author’s doctoral dissertation. The authors would like to thank Professors Thomas M. Corsi, Martin E. Dresner, Joseph G. Mattingly and Paul M. Schonfeld, all of the University of Maryland, for their assistance during the dissertation stage of this research and two anonymous reviewers for their comments on an earlier version of this paper.

REFERENCES


