10-2015

A two-item two-warehouse periodic review inventory model with transshipment

Koushik S. RAMAKRISHNA
Singapore Management University, koushik.sr.2007@mom.smu.edu.sg

Sharafali MOOSA
Singapore Management University, sharafalim@smu.edu.sg

Yun Fong LIM
Singapore Management University, yflim@smu.edu.sg

Published in Annals of Operations Research, October 2015, Volume 233, Issue 1, pp 365-381.
http://dx.doi.org/10.1007/s10479-013-1483-4

A TWO-ITEM TWO-WAREHOUSE PERIODIC REVIEW INVENTORY MODEL WITH TRANSSHIPMENT

Koushik S. Ramakrishna · Moosa Sharafali ·
Yun Fong Lim

Received: date / Accepted: date

Abstract We study a two-item two-warehouse periodic review inventory model that allows transshipment between warehouses and emergency orders. The model helps to determine two decisions: (i) the order up to levels for the two items at the two warehouses, and (ii) whether to accept a transshipment request from the other warehouse. The acceptance of a transshipment request in a warehouse depends on the time until the next review epoch and the inventory level of the warehouse. We propose a search procedure, which is a combination of greedy heuristics and a Lagrangian relaxation method, to minimize the total operating cost of the system. Numerical experiments suggest that optimality can be achieved using the proposed procedure. The time threshold to accept a transshipment request in a warehouse increases with the emergency shipment cost. In addition, the more inventory available in the warehouse the more likely for a transshipment request to be accepted.

Keywords Multi-item inventory · Transshipment · Pooling · Markov decision models · Lagrangian relaxation

1 Introduction

We consider a problem faced by an online retailer of pet food described in Miller et al. (2006). This online retailer has multiple warehouses at strategic locations in order to serve her customers. When a

Koushik S. Ramakrishna
Lee Kong Chian School of Business, Singapore Management University, SINGAPORE 178899
E-mail: koushiksrinivasr@gmail.com

Moosa Sharafali
Lee Kong Chian School of Business, Singapore Management University, SINGAPORE 178899
Tel.: +65-68280899
Fax: +65-68280777
E-mail: sharfalim@smu.edu.sg

Yun Fong Lim
Lee Kong Chian School of Business, Singapore Management University, SINGAPORE 178899
Tel.: +65-68280774
Fax: +65-68280777
E-mail: yflim@smu.edu.sg

customer's order cannot be fulfilled completely from a single warehouse due to shortage, the company usually tries to deliver the order from multiple warehouses directly to the customer. Miller et al. (2006) and Xu et al. (2009) call this split deliveries. According to Miller et al. (2006), customer satisfaction is low when they receive their orders in split deliveries. The online retailer in their study was concerned about this issue. So, in that paper, the authors develop heuristics to determine appropriate bundles of products that each warehouse should carry so that the total number of split orders is minimized. In their model the authors assume a bulk factor that depends on the order quantity and safety stock for an item.

In spite of the flexibility that a multi-item multi-warehouse system offers, timely deliveries still depend on stock availability at the warehouses. One way to avoid split deliveries is to assemble a customer's order in a warehouse near the customer by making lateral shipments of items from other warehouses, and then perform a single delivery to the customer. As highlighted by Miller et al. (2006), customer satisfaction is high in spite of the higher cost that is usually passed on to the customers.

Motivated by the above real-life observation, we aim to find an optimal inventory policy given the bundles of items carried by the warehouses. We assume that the company has an option to use transshipment between warehouses to avoid split deliveries. It is appealing to use transshipment in practice as it reduces the number of split orders, which simplifies the handling of final assembly and product returns. In addition, it is also well known that transshipment helps companies to pool inventory, and allows them to carry a larger number of items in their catalogs.

There exists a lot of work in the literature on inventory models with transshipment. It is also common in many industries for companies to use a periodic review policy to manage their inventory. However, according to Paterson et al. (2011), the study of transshipment in periodic review inventory models is limited to mainly single-item models. In their concluding remarks, the authors mention that research in multi-item inventory models is lacking and point out that this is an area worth pursuing. Specifically, joint optimization on multiple items over multiple warehouses with transshipment has not received much attention.

One paper by Archibald et al. (1997) derives formulas for a single-item two-warehouse periodic review inventory model, which permits transshipment and emergency shipments. The authors also propose a heuristic to find solutions for a two-item two-warehouse model. However, we do not find any expression of the objective function for the two-item two-warehouse model in that paper. The authors also admit that the proposed heuristic is not successful for certain problem instances.

Our objective in this paper is to propose a heuristic that provides solutions for the two-item two-warehouse periodic review inventory model. Specifically, we consider a system with two items and two capacitated warehouses where transshipment between warehouses is permitted. Similar to Miller et al. (2006), we assume items could be bundles of products. Each warehouse carries both bundles with the objective to minimize the total operating cost, which comprises the variable ordering cost, the holding cost at the warehouses, the cost of transshipment between warehouses, and the cost of emergency orders if transshipment is not possible. We formulate the problem as an infinite horizon dynamic program.

We propose a search procedure, which is a combination of greedy heuristics and a Lagrangian relaxation method, to solve the problem. Numerical experiments suggest that the proposed procedure can reach optimality. Furthermore, we also identify conditions where an emergency order is preferred over transshipment. We emphasize that our solution approach is different from that by Archibald et al.

(1997). Our approach can also be extended to the multi-item two-warehouse case in a straightforward manner.

This paper is structured as follows. Section 2 reviews the literature and highlights the lack of work on periodic review inventory models with transshipment. In Section 3 we present the formulation of the two-item two-warehouse periodic review inventory model with transshipment. Section 4 presents a search procedure for solving the model. The procedure consists of three heuristics used in succession. Section 5 demonstrates the applicability of our approach through numerical experiments. The last section concludes the paper with a summary of findings.

2 Literature review

A comprehensive review of the literature on inventory models with lateral shipments is provided by Paterson et al. (2011). To avoid repetition, we confine here to a very brief review of research related to our work in particular. Tables 3 and 4 of Paterson et al. (2011) reveal that only four papers analyze a multi-item inventory model with transshipment. We provide an extract of this information in Table 1 wherein we have also included our work in the last row. Of the five papers listed, only two discuss periodic review models and both of them use an order up to policy in a multi-item two-warehouse system.

Author(s)	**Continuous or periodic review?**	**Policy considered**	**# of warehouses**	**# of items**	**Transshipment Type**	**Transshipment Pooling**
Archibald et al. (1997)	Periodic	Order up to	2	m	Reactive	Partial
Kranenburg et al. (2009)	Continuous	$(S-1, S)$	n	m	Reactive	Partial
Wong et al. (2005)	Continuous	$(S-1, S)$	n	m	Reactive	Complete
Wong et al. (2006)	Continuous	$(S-1, S)$	2	m	Reactive	Complete
Our paper	Periodic	Order up to	2	m	Reactive	Partial

Table 1 Inventory models with transshipment.

Transshipment facilitates pooling of inventory. In the literature we find two types of pooling strategies. In *complete pooling* of inventory a transshipment location shares all of its inventory in case of a transshipment request. In *partial pooling* a transshipment location shares only a fraction of its inventory, keeping the rest to meet its own demand. A further classification of the models is on the type of transshipment used. In a *proactive* model transshipment can be effected only at fixed points in time, while transshipment can occur at any time in a *reactive* model. In the literature we find predominantly reactive transshipment policies (see Paterson et al. (2011)).

Wong et al. (2006) classify the multi-item multi-warehouse models according to the level at which the solutions are found and analyzed. The analysis could be at the *item* level or at the *system* level. In the item-level approach, optimization is performed item wise (by considering the problem as a set of independent single-item cases). The system-level approach as proposed by Sherbrooke (2004) resorts to system optimization by considering all items jointly. Obviously, system optimization is more cost effective. However, the bulk of the literature involving transshipment uses only the item-level approach. Papers that use the system-level approach include Archibald et al. (1997) and Wong et al. (2006).

Among the papers on inventory models with transshipment, the most relevant to our work is that of Archibald et al. (1997). The authors mainly analyze an order up to inventory policy with transshipment, and develop expressions of order quantities for a two-item one-warehouse problem. They also propose a heuristic to solve the two-item two-warehouse problem. However, the proposed heuristic is based on some restrictive assumptions, and it also fails to find solutions in certain circumstances. Wong et al. (2006) propose a Lagrangian relaxation based heuristic to solve a multi-item continuous review inventory model with waiting time constraints.

As indicated above, little work has been done on using the system-level approach to solve periodic review inventory models with transshipment. In addition, capacitated warehouses are also not widely considered. In this paper, we propose a heuristic to solve a periodic review capacitated inventory model for the two-item two-warehouse system. The approach can easily be extended to the multi-item two-warehouse case.

3 Model and analysis

We consider an online retailer with two capacitated warehouses as depicted in Figure 1. Each of these warehouses serves a geographical region. For simplicity, we assume that the warehouses carry only two items. The retailer uses a periodic review inventory policy. Without loss of generality, we assume that the length of each period is 1. Demand for item j in warehouse k occurs according to a Poisson process with rate λ_{kj}, $j, k = 1, 2$. The Poisson processes are independent. Each warehouse makes joint replenishment orders for both items to a central warehouse. If a demand occurs during a stockout at a warehouse, that demand can be satisfied by either a transshipment from the other warehouse or by an emergency order placed to the central warehouse. Obviously, this results in extra costs.

The transshipment policy employed by the system is reactive with partial pooling of inventory. That is, in the event of a transshipment request, the decision to accept it by the other warehouse is dependent on the time still remaining until the next period (or the next order) and the stock availability at the point in time. For example, consider just one of the two items (say item j) which is held in both warehouses. The inventory behaviour is depicted in Figure 2. At time $1 - t$ a demand for item j occurs at warehouse 1 during a stockout while at the warehouse 2 we find i_{2j} (> 0) units available. So, a transshipment request from warehouse 1 to warehouse 2 is sent instantly. But, warehouse 2 honours this request only if the remaining time to the next order, i.e., t time units is less than a threshold time of τ^2_{2j} (a decision variable) which is dependent on i_{2j}. Otherwise, the transshipment request is rejected in which case warehouse 1 makes an emergency order. This is very natural as the other warehouse would not want to risk a stock out occurring in its own warehouse before the next period.

We also assume as in Archibald et al. (1997) that the unsold inventory at the end of the period can be returned to the central warehouse for a full refund and then a new order is placed and received at the warehouses at the start of the next period. Without loss of generality we assume that the skus (stock keeping units) are in units of pallets and hence their bulk factors are the same. Consequently, the storage space required by any sku in the warehouse is the same.

We assume instantaneous replenishment of orders from the central warehouse. The ordering of stocks and its instantaneous delivery occurs at the start of the review period for all the items. The review period is also assumed to be the same for both the items (and for both warehouses) and hence all orders and their receipts occur simultaneously and jointly at the same time.

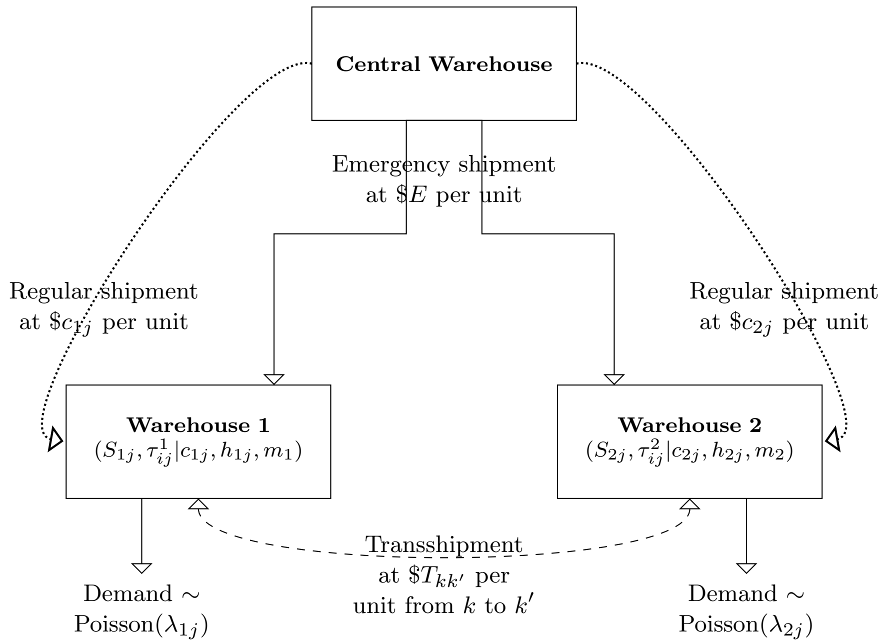

Fig. 1 A two-item two-warehouse model with transshipment where $j = 1, 2$.

We note that there are two decision problems in this model. The first is the decision to choose between an emergency order or a transshipment at an instant within a review period when a demand occurs at a warehouse for a stocked out item. The second problem is the reordering decision at a review epoch. Archibald et al. (1997) have performed an excellent analysis of these by using a finite horizon continuous time Markov decision process for the first problem and an infinite horizon discounted Markov decision process for the second. We adapt their results for our model. The underlying stochastic process is the inventory level process at the warehouses given by $\{(\mathbf{S}_1(t), \mathbf{S}_2(t)), t \geq 0\}$ with the state space given by $E = \{(\mathbf{s}_1, \mathbf{s}_2)\}$ where the vectors $\mathbf{s}_k = (s_{k1}, s_{k2})$ with $0 \leq s_{k1} + s_{k2} \leq m_k$ and m_k is the capacity of the warehouse k, $k = 1, 2$. The objective is to minimize the system-wide total cost which comprises of the variable ordering cost, holding and transshipment costs. We assume that there is no fixed ordering cost. Specifically, we use the notation as given in Table 2.

3.1 The main optimization problem

In order to formulate our main optimization problem, we first derive our objective function which is the expected discounted cost for the infinite horizon. We recall our assumption that we return the unsold items just prior to the start of the next period at no cost and with full refund for the returned items. We then make the order for the next period. Since, there is no fixed order cost in the model and the variable order cost is linear, the ordering decision at the start of any period is independent of the stock level just before the review. Hence, it is enough to consider only one state of the system at a review epoch.

For convenience, as in Archibald et al. (1997), we take for the system, this state to be $(0, 0)$ for each item $j = 1, 2$. Consequently, at any ordering instant, the system will not have any items in the warehouse.

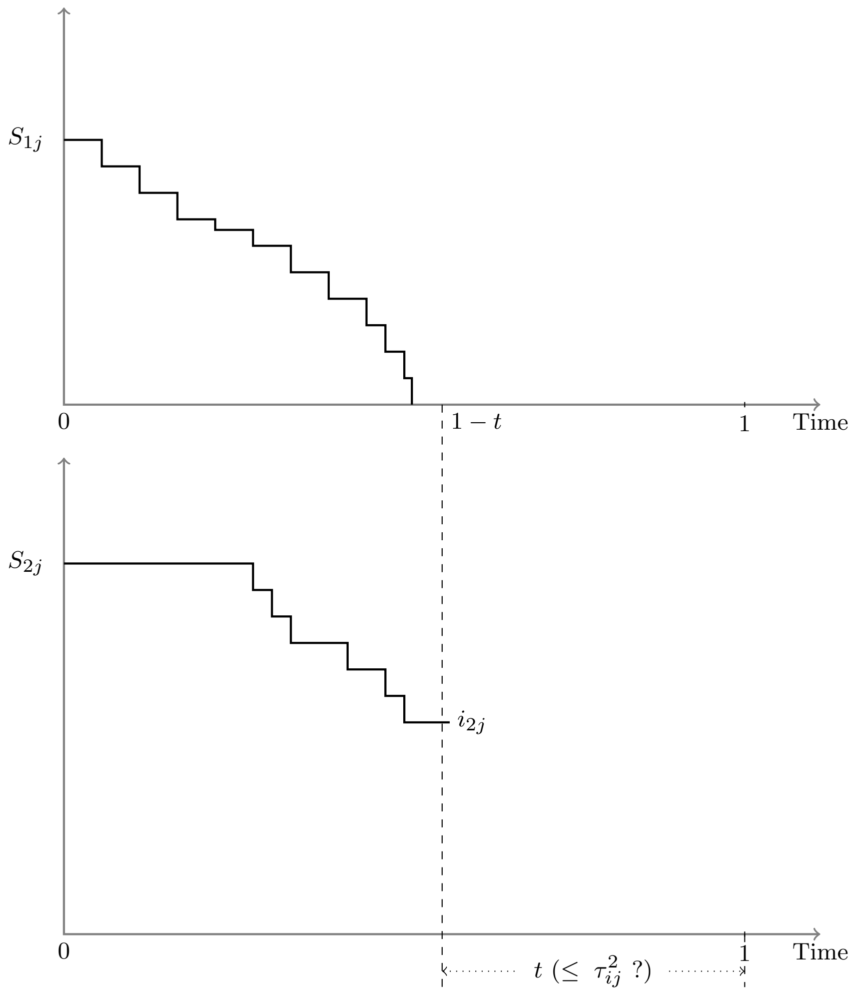

Fig. 2 Decision time within a period in a one-item two-warehouse model when a stockout occurs at $1-t$ in warehouse 1 and the stock in warehouse 2 is $i_{2j}(>0)$ units.

So, the decision problem is to know how many units to have at the start of any period. Let us assume that at the start of every period after returning the unsold items, we will order for s_{kj} units of item j for warehouse k, with $k, j = 1, 2$. Let $W_1(\mathbf{s}_1, \mathbf{s}_2)$ be the one-period cost for maintaining the inventory and for satisfying the demand from stock or from transshipment or by emergency order. Then, it is very easy to see that, if $V(\mathbf{0}, \mathbf{0})$ is the infinite horizon minimum expected total discounted cost for this system of two warehouses, then the optimal value function satisfies the following optimality equation:

$$V(\mathbf{0},\mathbf{0}) = \min_{s_{k1}+s_{k2}\leq m_k, k=1,2} \left\{ \sum_{j=1}^{2} \left\{ c_{1j}s_{1j} + c_{2j}s_{2j} \right\} + \beta(W_1(\mathbf{s}_1,\mathbf{s}_2) + V(\mathbf{0},\mathbf{0}) \right\}, \qquad (1)$$

β	Discount factor
c_{kj}	Cost of regular order per unit for item j in warehouse k
E	Emergency order cost per unit
$f(\lambda, n, t)$	$= e^{-\lambda t} \frac{(\lambda t)^n}{n!}$
$F(\lambda, n, t)$	$= \sum_{i=0}^{n} e^{-\lambda t} \frac{(\lambda t)^i}{i!}$
λ_k	Demand rate at warehouse k
m_k	Capacity of warehouse k
$p_k, 1 - p_k$	Probability of a demand for items 1 and 2, respectively, at warehouse k
s_{kj}	Stock level of item j in warehouse k
S_{kj}	Order up to level of item j in warehouse k
$T_{kk'}$	Transshipment cost from warehouse k to warehouse k'
τ_{ij}^{k}	The threshold time until the next period for accepting a transshipment request at warehouse k holding inventory of i units of item j at the instant of the transshipment request (*a decision variable*)
$V_j(0, 0)$	Infinite horizon discounted total cost due to item j
$V(\mathbf{0}, \mathbf{0})$	Infinite horizon discounted total cost of the system
$w_t^E(s_{1j}, s_{2j})$	The minimum expected total cost until the next review epoch, given that the time to the next review epoch is t and there is an unmet demand for item j with stock (s_{1j}, s_{2j}) that is satisfied with an emergency order
$w_t^T(s_{1j}, s_{2j})$	The minimum expected total cost until the next review epoch, given that the time to the next review epoch is t and there is an unmet demand for item j with stock (s_{1j}, s_{2j}) that is satisfied with a transshipment
$W_1(s_{1j}, s_{2j})$	The minimum expected total cost per period to satisfy demand for item j by transshipment or emergency orders
$W_1(\boldsymbol{s}_1, \boldsymbol{s}_2)$	The minimum expected total cost per period to satisfy demand by transshipment or emergency orders

Table 2 Notation used in this paper.

where β is the discounting factor. It is now clear that as this a single state Markov decision process, the optimal policy is the order up to level policy (see Archibald et al. (1997)). Note that $V(\mathbf{0}, \mathbf{0})$ is a function of $(\mathbf{s}_1, \mathbf{s}_2)$.

A moment of reflection on the value function for the finite horizon Markov Decision process (MDP) will reveal the complexity involved in considering the events that occur in the evolution of the inventory level stochastic process. Hence, as a first level of approximation, we propose to consider the two-item two-warehouse model to be two separate one-item two-warehouse models. The two separate models are then inter-related through the capacity constraints.

So, if we consider each item independently and assume that $V_j(0, 0)$ to be the infinite horizon value function corresponding to only item j, $j = 1, 2$ in both warehouses, then we can just borrow single-item two-warehouse results of Archibald et al. (1997). For easy reference, these results are surveyed in the Appendix at the end of this paper. Then, we have

$$V_j(0, 0) = \min_{s_{k1}+s_{k2} \le m_k, k=1,2} \left\{ c_{1j}s_{1j} + c_{2j}s_{2j} + \beta(W_1(s_{1j}, s_{2j}) + V_j(0, 0) \right\}. \tag{2}$$

Now, $W_1(s_{1j}, s_{2j})$ remains to be found. By conditioning on the instant of the occurrence of the first unmet demand, $W_1(s_{1j}, s_{2j})$ can easily be derived as in Archibald et al. (1997) which is given in (16). Thus, for two-item two-warehouse capacitated model we have the following optimization problem.

P0:

$$\min V(\mathbf{0},\mathbf{0}) = \sum_{j=1}^{2} V_j(0,0) \tag{3}$$

subject to

$$s_{11} + s_{12} \leq m_1;$$

$$s_{21} + s_{22} \leq m_2;$$

where m_1, m_2 are the capacities of warehouses 1 and 2 respectively and $V_j(0,0)$ are given by (2).

The solution of the problem is the vector of order up to levels, $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2)$ with $\mathbf{S}_1 = (S_{11}, S_{12})$ and $\mathbf{S}_2 = (S_{21}, S_{22})$ which minimizes the total cost of the problem for the given capacity constraints. Let the associate total cost be $C(\mathbf{S})$. Let C^* be the optimal cost corresponding to optimal solution $\mathbf{S}^*$ for problem **P0**.

It is clear that the problem is a non-linear integer programming problem with linear constraints. The model is in a suitable format for the application of *Lagrangian relaxation* (see Fisher (1985)).

3.2 The Lagrangian relaxation problem

Let the vector $\underline{\mu} = (\mu_1, \mu_2)^T \epsilon \Re^2$ with $\mu_j \geq 0$, $j = 1, 2$ be the Lagrange multipliers. Now, by relaxing the capacity constraints in problem **P0** we obtain the following relaxed problem.

P1:

$$\min \sum_{j=1}^{2} V_j(0,0) + \mu_1(\sum_{j=1}^{2} s_{1j} - m_1) + \mu_2(\sum_{j=1}^{2} s_{2j} - m_2)$$

$$= \sum_{j=1}^{2} (V_j(0,0) + \mu_1 s_{1j} + \mu_2 s_{2j}) + \mu_1 m_1 + \mu_2 m_2. \tag{4}$$

As the constant term $\mu_1 m_1 + \mu_2 m_2$ can be ignored during the optimization process, it is clear that the Lagrangian relaxation problem **P1** is separable in item and so the optimal solution to **P1** can be obtained by solving the following problem for each $j = 1, 2$.

P1$_j$**:**

$$\min V_j(0,0) + \mu_1 s_{1j} + \mu_2 s_{2j}. \tag{5}$$

Let $C^1_{\underline{\mu}}(\mathbf{S})$ be the optimal cost of problem **P1**, for given set of μ's. Then the capacity used at the warehouse k for this ordering policy $\mathbf{S}$ is, $M_k = S_{k1} + S_{k2}$. We now have the following result (see Wong et al. (2006)).

Property 1 From the formulation in **P1** we have the following properties:

(i) $C^* \geq C^1_{\underline{\mu}}(\mathbf{S})$ for all $(\mu_1, \mu_2) \geq (0, 0)$.
(ii) $C^* \geq \max_{\underline{\mu}} C^1_{\underline{\mu}}(\mathbf{S})$.
(iii) If for some $(\mu_1, \mu_2) \geq (0, 0)$ the optimal solution for problem **P1** is $\mathbf{S}^{1*}$ and $M_k \leq m_k$, $k = 1, 2$ then $\mathbf{S}^{1*}$ is feasible for problem **P0** and $C(\mathbf{S}^{1*}) - C^* \leq \mu_1(m_1 - M_1) + \mu_2(m_2 - M_2)$.

(iv) If for some $(\mu_1, \mu_2) \geq (0, 0)$ the optimal solution for problem **P1** is $\mathbf{S}^{1*}$ and for $k = 1, 2$, $M_k = m_k$, if $\mu_k > 0$; $M_k \leq m_k$ if $\mu_k = 0$, then $\mathbf{S}^{1*}$ is the optimal ordering policy for problem **P0.**

Proof

(i) The result easily follows from the observation that any optimal solution to problem **P0** is a feasible solution for problem **P1** for any given $(\mu_1, \mu_2) \geq (0, 0)$. In turn, any feasible solution to problem **P1** should yield an objective value that is more than or equal to its optimal objective value.
(ii) This can easily be inferred from property (i) above.
(iii) The first part is the consequence of definition of problem **P0** while the second part follows from (i).
(iv) This follows from (iii). One can refer to Everett (1963) from which also the result follows. □

The properties above are useful in our search for the optimal solution to problem **P0**. First, we note that Property (i) above provides us with a lower bound for the optimal objective function value of problem **P0**. The next property (ii) indeed provides us with the best such lower bound. Further, from (iii) we have an upper bound for the gap between the objective function value for any feasible solution to **P0** and its optimal objective value. The final result indicates that the relaxed solution can be optimal to problem **P0** and if so the capacity of a warehouse will be fully utilized when the corresponding multiplier is positive and not fully utilized when the multiplier is zero.

Thus, to develop the algorithm, we first need the optimal solution to the relaxed problem for given set of Lagrange multipliers which would give us a lower bound. Then, we need to get the tightest lower bound for which we should find the best Lagrange multipliers. The approaches for finding these are presented below.

It is known (see Archibald et al. (1997)) that the functions in involved in (5) are convex with respect to the inventory levels and sub-modular (see Topkis (1978)) in time and inventory level variables. So, these separated problems can be solved independently using the algorithm of Archibald et al. (1997) (please refer to the Appendix). But, we note that to apply this procedure we need the values of the multipliers μ_1 and μ_2. From the property above, for a given set of (μ_1, μ_2), the solution to the separated problems yields a lower bound for the optimal objective value of problem **P0**. The best possible lower bound can be obtained by optimizing over μ_1 and μ_2. As the problem is a non-linear integer programming problem, it is not differentiable and so we cannot apply methods like steepest ascent. For such situations, the approach usually employed is the subgradient optimization method (Bazaraa et al. (1993)) which could be considered similar to steepest ascent method with the subgradient based direction replacing the gradient direction. We refer the readers to Fisher (1985) for an excellent exposition of this procedure. We employ this procedure to our problem to find the optimal values of (μ_1, μ_2).

Thus, our task is now to obtain the tightest lower bound and the associated best Lagrange multipliers (μ_1^*, μ_2^*) for $C^1_{\underline{\mu}}(\mathbf{S})$. The method is to first solve Problem $\mathbf{P1}_j$ given in (5) by taking some initial values for (μ_1, μ_2) and then iteratively updating (μ_1, μ_2) using the subgradient optimization until the best lower bound is obtained. We now describe these two procedures below.

Initial values for the multipliers

To apply the subgradient procedure, we need to initialize (μ_1, μ_2). Usually, $(0, 0)$ is chosen as the initial values for the μs but a more efficient procedure to choose the initial values for (μ_1, μ_2) is proposed by

Wong et al. (2006). The algorithm first finds three possible initial values for (μ_1, μ_2) and then chooses the best among the three.

Algorithm 1 Finding Initial Multipliers

1: First set: Let $\mu_1 = 0$. Find the smallest value of μ_2 for which the capacity constraints of problem $P0$ are satisfied.
2: Second set: Let $\mu_2 = 0$. Find the smallest value of μ_1 for which the capacity constraints of problem $P0$ are satisfied.
3: Third set: Let now $\mu_1 = \mu_2$. Find the smallest value of both μ_1 and μ_2 for which the capacity constraints of problem $P0$ are satisfied.
4: Of the three above, choose the one with the largest objective function value for Problem $P1$ as the initial set of μs for the subgradient procedure.
5: END

We also note that the optimal solution corresponding to the selected set of multipliers is a feasible solution for problem **P0.** This now provides an initial value for the upper bound (call it $\widehat{C}$) that is required in the subgradient procedure. We now describe this procedure below.

Subgradient optimization

The procedure involves updating (μ_1, μ_2) at each iteration using the subgradient direction calculated in that iteration. If at the n-th iteration, $\underline{\mu}^n = (\mu_1^n, \mu_2^n)$ are the Lagrangian multipliers and if M_k^n be the total capacity used in warehouse k, the subgradient direction at iteration n is given by

$$\gamma_k^n = M_k^n - m_k. \tag{6}$$

The Lagrange multipliers are updated as follows:

$$\mu_k^{n+1} = \max(0, \mu_k^n - t^n \gamma_k^n). \tag{7}$$

In the above, t^n is the step size which also needs updating at every iteration. The most commonly used updating procedure is the one proposed and justified by Held et al. (1974). The updating formula is

$$t^n = s^n \frac{C^1_{\underline{\mu}^n} - \widehat{C}}{(\gamma_1^n)^2 + (\gamma_2^n)^2}, \tag{8}$$

where $\widehat{C}$ is the best known upper bound for problem **P0** and s^n is a scalar between 0 and 2. If after a specified number of iterations, there is no improvement in the value of the objective function, the step size is updated by halving the value of s^n.

Having prepared the ground for solving our problem to optimality, we now propose in the next section the procedure to obtain the optimal solution. For more details on our experience which led to this solution procedure, we refer the readers to Koushik (2010).

4 Solution procedure

As described in the previous section, our optimization problem which is a non-linear integer programming problem can only be solved numerically. To this end, we described in the last section the Lagrangian relaxation based procedure supported by subgradient optimization.. We ran the procedure for some example problems, verifying the solution through the use of brute force. Our experience revealed that the convergence of the procedure and also the quality of the final solution depended very much on the initial value chosen for $\mathbf{S}$, the order levels. The usual choice of $\mathbf{S} = (\mathbf{0}, \mathbf{0})$ appeared to be a very poor for the procedure. Thus, choosing a good initial feasible $\mathbf{S}$ became important. For this purpose, we propose a greedy heuristic. This heuristic starts at an infeasible solution $\mathbf{S} = (m_1\boldsymbol{e},\, m_2\boldsymbol{e})$, where $\boldsymbol{e} = (1, 1)$. That is, we ignore the capacity constraints and proceed iteratively to find a good initial feasible solution. One other noteworthy observation is that in many of the example problems we solved, the solution from the greedy heuristic itself turned out to be optimal as confirmed during the application of the Lagrangian relaxation approach.

The next step is to use the Lagrangian relaxation approach supported by the subgradient optimization procedure to update $(\mu_1,\, \mu_2)$. As we explained in the previous section, we also needed a good initial $(\mu_1,\, \mu_2)$ for which we resorted to the procedure proposed by Wong et al. (2006). It is known that the Lagrangian relaxation approach does not always lead to the optimal solution to the original problem. Thus, we propose a local neighbourhood procedure to improve the solution provided by Lagrangian relaxation approach. Algorithm 2 summarizes the procedure to solve the problem.

Algorithm 2 Solution Procedure

1: Use the greedy heuristic to find an initial feasible solution.
2: Use the Lagrangian relaxation method supported by the subgradient procedure to test whether the initial solution is optimal. If not, find iteratively a better solution. If optimality cannot be reached then go to the next step.
3: Use the improvement heuristic to find a better solution to the one found in Step 2.
4: End

We note that in the above procedure, to solve the item-wise separated optimization problems, we use the procedure of Archibald et al. (1997). To evaluate the performance of our procedure, we indeed compared the solution from the above procedure to the optimal solution obtained using brute force. It is heartening to note that our procedure indeed performed better. We now describe each of the steps in our main procedure given above.

4.1 Initial greedy heuristic

As explained above, our initial greedy heuristic starts with an infeasible solution, ignoring the capacity constraints. To move towards feasibility, the technique then removes one unit of either product from the current solution and calculates the difference in the value of cost before and after the removal. The item that produces the least of the differences is removed from the current solution. This removal procedure is repeated until a feasible solution is reached. It should be noted that the i-th unit of product j is a candidate for removal only if the $(i-1)$-th unit of product j had already been removed. That is, at any iteration only one unit is removed. This heuristic is based on the observation that the function $V_j(0,0)$ is convex with respect to (s_{1j}, s_{2j}) and so $\triangle V_j(0,0)$ increases with increase in either s_{1j} or s_{2j}. Hence removal of one unit is certainly cheaper than removal of the subsequent unit. Algorithm 3 describes the initial greedy heuristic.

Algorithm 3 Initial Greedy Heuristic

$(s_{11}, s_{12}, s_{21}, s_{22}) \leftarrow (m_1, m_1, m_2, m_2)$
$S_1 \leftarrow s_{11} + s_{12}$ and $S_2 \leftarrow s_{21} + s_{22}$
repeat
  if $s_{11} = 0$ **then**
    $s_{12} \leftarrow s_{12} - 1$
  else if $s_{12} = 0$ **then**
    $s_{11} \leftarrow s_{11} - 1.$
  else
    $d_1 \leftarrow V_1(s_{11} - 1, s_{21}) - V_1(s_{11}, s_{21})$, $d_2 \leftarrow V_2(s_{12} - 1, s_{22}) - V_2(s_{12}, s_{22})$
    if $d_1 < d_2$ **then**
      $s_{11} \leftarrow s_{11} - 1$
    else
      $s_{12} \leftarrow s_{12} - 1$
      $S_1 \leftarrow s_{11} + s_{12}$
    end if
  end if
until $S_1 \leq m_1$
repeat
  if $s_{21} = 0$ **then**
    $s_{22} \leftarrow s_{22} - 1$
  else if $s_{22} = 0$ **then**
    $s_{21} \leftarrow s_{21} - 1$
  else
    $d_3 \leftarrow V_1(s_{11}, s_{21} - 2) - V_1(s_{11}, s_{21})$, $d_4 \leftarrow V_2(s_{12}, s_{22} - 1) - V_2(s_{12}, s_{22})$
    if $d_3 < d_4$ **then**
      $s_{21} \leftarrow s_{21} - 1$
    else
      $s_{22} \leftarrow s_{22} - 1$
      $S_2 \leftarrow s_{21} + s_{22}$
    end if
  end if
until $S_2 \leq m_2$
END

It is very clear that if the problem is unconstrained, this heuristic itself will converge to the optimal solution. In the capacity constrained case, the method would follow the steepest descent till the inventories satisfy the capacity constraints. For discrete convex optimization such as this problem, this greedy procedure does not guarantee convergence to global optimality. Hence, we limit the use of the greedy heuristic only to obtain an initial feasible solution. However, the Lagrangian relaxation procedure can test its optimality in all cases. We now describe the Lagrangian relaxation method.

4.2 Lagrangian relaxation method

We fist recall the Lagrangian relaxation problem (**P1**) formulated in (5). Note that each separate item-wise problem in it is a single-item two-warehouse problem which can be solved using the procedure of Archibald et al. (1997). Algorithm 4 describes the Lagrangian heuristic.

Algorithm 4 Lagrangian Heuristic

1: Choose a value for the scalar $s\varepsilon(0,2)$.
2: Get the initial solution **S** and its associated cost C from Algorithm 3
3: Get the initial values for μ_1 and μ_2 from Algorithm 1 in section 3.2.
4: Solve the problem **P1** with μ_1 and μ_2 obtained in the last step. Let the corresponding solution be $\mathbf{S}_{\underline{\mu}}$ and the total cost be $C_{\underline{\mu}}$
5: **if** $\mathbf{S}_{\underline{\mu}}$ is feasible and $C_{\underline{\mu}} \leq C$ **then**
6:     $\mathbf{S} \leftarrow \mathbf{S}_{\underline{\mu}}$ and $C \leftarrow C_{\underline{\mu}}$
7: **end if**
8: **if** $\mid C - C_{\underline{\mu}} \mid \sim 0$ **then**
9:     go to step 14
10: **else if S** is the same as any solution obtained in previous iterations **then**
11:     scalar $s \leftarrow s/2$.
12: **end if**
13: Update the values of μ_1, μ_2 as follows and go to step 4

- $\gamma_1 \leftarrow s_{11} + s_{12} - m_1$
- $\gamma_2 \leftarrow s_{21} + s_{22} - m_2$
- $t \leftarrow \frac{s(C_{\underline{\mu}} - C)}{(\gamma_1)^2 + (\gamma_2)^2}$
- $\mu_1 \leftarrow \max(0, \mu_1 - t\gamma_1)$
- $\mu_2 \leftarrow \max(0, \mu_2 - t\gamma_2)$

14: End

4.3 Improvement heuristic

In spite of using good initial solution and good Lagrangian multipliers, we found instances when the procedure did not converge to the optimal solution. To improve this and to also explore the feasible region for a better solution, we propose a neighborhood search heuristic. We describe this improvement heuristic as follows.

From the value of $V_j(0,0)$ obtained as the current upper bound in the Lagrangian relaxation procedure we identify the feasible neighbors of $\mathbf{s}_j = (s_{1j}, s_{2j})$ for product $j = 1, 2$. Combinations of neighbors from the two sets are checked for satisfying the capacity constraints, discarding those combinations that violate these constraints. The feasible neighbour yielding the minimum cost replaces the solution giving the current upper bound. The process is repeated till the current upper bound solution did not change. Algorithm 5 describes the improvement heuristic. Here, $\boldsymbol{e}_i$ is the unit row vector with 1 in the i-th place and $\boldsymbol{e} = (1, 1)$.

Algorithm 5 Improvement Heuristic

1: Get the solution $\mathbf{S}_\ell$ and its associated cost C_ℓ from Algorithm 4
2: $\mathbf{S}^0 \leftarrow \mathbf{S}_\ell$ and $C^0 \leftarrow C_\ell$
3: Get the feasible neighborhood set $N(\mathbf{S}^0) = \{(\mathbf{s}_1^0 \pm \mathbf{e}_i, \mathbf{s}_2^0 \pm \mathbf{e}_j) : (\mathbf{s}_1^0 \pm \mathbf{e}_i)\mathbf{e}^T \leq m_1, (\mathbf{s}_2^0 \pm \mathbf{e}_j)\mathbf{e}^T \leq m_2; i, j = 1, 2\}$
4: Find $\mathbf{S}' = \arg\min_{\mathbf{S} \epsilon N(\mathbf{S}^0)} V(\mathbf{0}, \mathbf{0})$ and the corresponding $V_{\mathbf{S}'}$, the value of the value function at $\mathbf{S}'$
5: **if** $V_{\mathbf{S}'} \neq C^0$ **then**
6: $\quad \mathbf{S}^0 \leftarrow \mathbf{S}'$
7: $\quad C^0 \leftarrow C'$
8: GO TO 3
9: **else**
10: STOP
11: **end if**
12: End

The Improvement procedure is used only in cases when the Lagrangian relaxation fails to reach optimality. The Lagrangian heuristic is useful because the solution provided by it is based on the inputs such as good initial solution and Lagrangian multipliers. So, if the Lagrangian heuristic fails then with the solution from it, the improvement heuristic would be able to find a better local minimum. Now, to illustrate our procedure and to analyse its effectiveness we ran some numerical experiments. In the next section, we describe this aspect of our work.

5 Numerical experiments

This section performs numerical experiments based on realistic settings to study the proposed solution procedure. We focus on convergence to optimality of the heuristics and sensitivity analysis with respect to changes in parameters.

5.1 Experiment settings

We have selected a range of values for the parameters that are realistic for the e-fulfillment industry. Further we have pegged the arrival rates to the capacities of the warehouses so as to ensure that an unreasonable arrival rate is not generated for a specific capacity. We have considered the book e-tailing industry as a source for the parameters used in the experiments. A set of 288 experiments for the two-item two-warehouse problem was run with various combinations of parameters as listed in Table 3.

We recall our basic assumption that the demands at the warehouses occur according to Poisson processes. We use Poisson splitting of demands at each warehouses, i.e. if λ_k is the overall demand rate at warehouse k, then the demand rate for item 1 is $\lambda_k p_k$ and for item 2 is $\lambda_k(1-p_k)$. In real life situations, the inventory allocated to the warehouse and the capacity of the warehouse are usually inter-related. A serious mismatch of the capacity of the warehouse and inventory intended for the warehouse provides us with a capacity planning problem rather than an inventory allocation one. Hence the arrival rates are derived from the capacities of the warehouses through solving the following formulas. We assume that

Parameters	Units	# of Instances	Values
m_1	sku	3	20, 40, 60
m_2	sku	1	60
z-value from $N(0,1)$	–	2	2.57583, 1.95996
p_1	%	2	25%, 50%
p_2	%	2	25%, 50%
h	\$/sku-yr	3	0.01, 0.03, 0.05
c_{1j}	\$/sku	1	20
c_{2j}	\$/sku	1	20
E	\$/sku	2	50, 60
$T_{kk'}$	\$/sku	2	30, 40
Total number of settings		288	

Table 3 Values for the input parameters used in the numerical experiments.

the warehouses employ a simple service level based safety stock policy given by the well-known formula $\lambda + z_\alpha\sqrt{\lambda}$ where λ is the mean demand and z_α is the value of the standard normal variate corresponding to a service level of $1-\alpha$. Thus, we have,

$$\lambda_1 + z\sqrt{\lambda_1} - m_1 = 0 \tag{9}$$

$$\lambda_2 + z\sqrt{\lambda_2} - m_2 = 0 \tag{10}$$

Solving these equations (9 and 10) for $\lambda_1(> 0)$ and $\lambda_2(> 0)$, we get the total arrival rates for each warehouses. Then, using Poisson splitting, we get the demand rates for each of the products at each warehouse as follows:

$$\lambda_{11} = p_1 r_1 \tag{11}$$

$$\lambda_{12} = (1-p_1) r_1 \tag{12}$$

$$\lambda_{21} = p_2 r_2 \tag{13}$$

$$\lambda_{22} = (1-p_2) r_2 \tag{14}$$

Three specific holding cost rates 0.01, 0.03, 0.05 were used. The insights from the 288 experiments are described below.

5.2 Sensitivity analysis

At the outset, we would like to highlight that for many of the instances the initial greedy heuristic itself converged to optimality. Of course, we could verify this only during the application of the Lagrangian heuristic. We also encountered some instances outside of this experiment for which the initial greedy heuristic failed to converge but the Lagrangian heuristic was able to converge to the optimal solution.

Sensitivity analysis with respect to capacities
We note that at the optimal solution, either one or all of the μ's may be zero. If any of the μ's is zero, then the corresponding capacity constraint becomes tight and so irrelevant. In all our experiments,

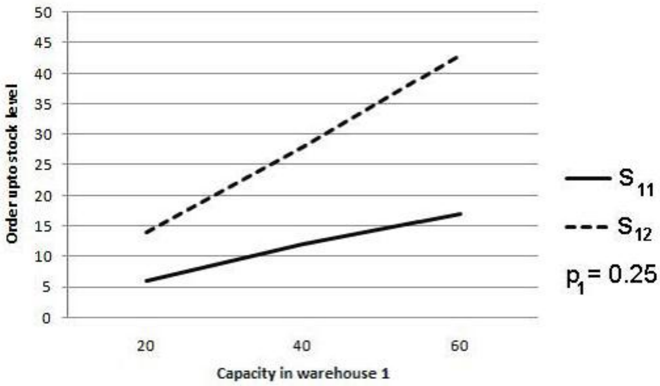

Fig. 3 Order up to level and capacity of warehouse 1.

the optimal solutions filled the warehouses to their capacities. This is also an observation and a crucial assumption for Archibald et al. (1997)'s algorithm for the two-warehouse case to work. We point out that in our case this assumption is not necessary. In Figure 3, as capacity of warehouse changes from 20 to 60, the optimal order level for the respective item in the warehouse increases proportionally. This is possibly due to the way we have pegged capacity to demand rates.

Sensitivity analysis with respect to costs

We have considered for the experiment two values each of emergency and transshipment costs and 3 values of holding costs. The solution was not sensitive to the changes in these costs. Since the solution fills the inventories to the capacity, the extra cost does not change the solution. However, 'τ_{ij}^{k}' the time instant which determines the choice between transshipment and emergency in case of transshipment requests due to stock outs, changes. For increase in holding cost, the values of the time limit τ_{ij}^{k} are unaffected and remain the same. But, τ_{ij}^{k} appears to be sensitive to emergency cost as shown in Figure 4. The figure clearly reveals the transshipment decision policy for warehouse 2, given an emergency cost E. The curves demarcate the regions over which the decision on transshipment request changes from acceptance to rejection. The decision depends on two variables - the time till the next review epoch and the inventory present in the other warehouse. It is interesting to note that for $E = 60$, the demarcation curve is a straight line while for $E = 50$, it is a curve. Clearly, as the emergency cost increases, the region in which emergency shipments are preferred (i.e. rejection of transshipment requests) shrinks. Further, the diagram reveals some conservatism that if at a transshipment request the review is further away then the algorithm recommends emergency shipment, especially so in the case of lower E.

6 Conclusions

In this paper we consider a two-item two-warehouse periodic review inventory model that permits transshipment and emergency orders in the event of a stockout at a warehouse. We develop a procedure to solve the problem with greedy and Lagrangian relaxation based heuristics. The procedure helps to determine two decisions: (i) the order up to levels for the two items at the two warehouses, and (ii)

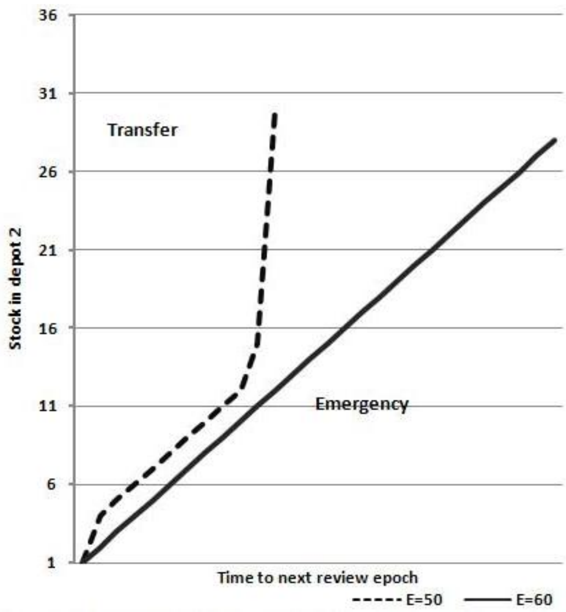

Fig. 4 Transshipment acceptance and rejection regions for warehouse 2.

whether to accept a transshipment request from the other warehouse. When a warehouse is requested for transshipment, the decision depends on two factors – the time until the next review epoch and the inventory level of the warehouse.

Our numerical experiments provide the following insights:

1. The initial greedy heuristic is itself quite good in terms of the number of times it converges to an optimal solution.
2. The initial greedy heuristic provides a feasible solution for the Lagrangian relaxation method. The latter converges to an optimal solution or provides bounds on the optimal solution.
3. The time threshold to accept a transshipment request in a warehouse increases with the emergency shipment cost. In addition, the more inventory available in the warehouse the more likely for a transshipment request to be accepted.
4. The decision to use an emergency shipment depends on the inventory level of the other warehouse.

We highlight that the Lagrangian relaxation method results in a separable optimization problem, which can be exploited to solve the multi-item two-warehouse problem. Interested readers may refer to Koushik (2010) for analysis of such a system. A natural extension to the problem discussed in this paper is to study the multi-item multi-warehouse problem.

Acknowledgements The first author thanks the Ministry of Education of Singapore and Singapore Management University for the generous scholarship. The third author is supported by the Lee Kong Chian School of Business, Singapore Management University and the Neptune Orient Lines (NOL) Fellowship, Singapore.

7 Appendix

In this appendix, we survey the results of Archibald et al. (1997) pertaining to the single-item two-warehouse system that is used in our Lagrangian relaxation method. For the sake of clarity, we retain the notation of that paper.

Lemma 1 (Archibald et al. (1997)) *Let $w_t^E(s_{1j}, s_{2j})$ and $w_t^T(s_{1j}, s_{2j})$, $j = 1, 2$, denote the minimum expected total costs until the next review epoch given that the time to the next review epoch is t when the system is in state (s_{1j}, s_{2j}) and there is an unmet demand at one of the warehouses for item j, which is satisfied by an emergency order and a transshipment respectively. Then*

$$\begin{aligned} w_t^E(s_{1j}, 0) = E + \int_0^t \lambda_{2j} f(\lambda_{2j}, s_{2j}, u) \Bigg\{ & \sum_{i_{1j}}^{s_{1j}} f(\lambda_{1j}, i_{1j}, u) w_{t-u}(s_{1j} - i_{1j}, 0) \\ & + \sum_{i_{1j}=s_{1j}+1}^{\infty} f(\lambda_{1j}, i_{1j}, u) \big((i_{1j} - s_{1j})E + w_{t-u}(0,0) \big) \Bigg\} du \\ & + f(\lambda_{2j}, 0, t) \Bigg\{ \sum_{i_{1j}}^{s_{1j}} f(\lambda_{1j}, i_{1j}, u) W_0(s_{1j} - i_{1j}, 0) \\ & \qquad + \sum_{i_{1j}=s_{1j}+1}^{\infty} f(\lambda_{1j}, i_{1j}, u) \big((i_{1j} - s_{1j})E + W_0(0,0) \big) \Bigg\} du \end{aligned}$$

and

$$w_t^T(s_{1j}, 0) = T_{1,2} - E + w_t^E(s_{1j} - 1, 0) \tag{15}$$

for $0 < s_{1j} \leq m_1$. In the above equations, the functions $w_t(s_{1j}, 0)$ and $w_t(s_{1j}, 0)$ are given as

$$w_t(s_{1j}, 0) = \min\{w_t^E(s_{1j}, 0),\ w_t^T(s_{1j}, 0)\} \text{ for } 0 < s_{1j} \leq m_1$$

and

$$w_t(0, s_{2j}) = \min\{w_t^E(0, s_{2j}), w_t^T(0, s_{2j})\} \text{ for } 0 < s_{2j} \leq m_2.$$

Further,

$$W_0(s_{1j}, s_{2j}) = h_{1j}s_{1j} + h_{2j}s_{2j} - c_{1j}s_{1j} - c_{2j}s_{2j}.$$

Proof Conditioning on the time at which the next demand at warehouse 2 occurs yields the above result. □

Theorem 1 (Archibald et al. (1997)) *For the finite horizon Markov decision process, the optimal value function satisfies the following: For $j = 1, 2$, the minimum expected total cost per period to satisfy* *the demands is*

$$\begin{aligned} W_1(s_{1j}, s_{2j}) = & \int_0^1 \lambda_{1j} f(\lambda_{1j}, s_{1j}, t) \sum_{i_{2j}=0}^{s_{2j}} f(\lambda_{2j}, s_{2j}, t) w_{1-t}(0, s_{2j} - i_{2j}) dt \\ & + \int_0^1 \lambda_{2j} f(\lambda_{2j}, s_{2j}, t) \sum_{i_{1j}=0}^{s_{1j}} f(\lambda_{1j}, s_{1j}, t) w_{1-t}(s_{1j} - i_{1j}, 0) dt \\ & + \sum_{i_{1j}=0}^{s_{1j}} \sum_{i_{2j}=0}^{s_{2j}} f(\lambda_{1j}, s_{1j}, t) f(\lambda_{2j}, s_{2j}, t) W_0(s_{1j} - i_{1j}, s_{2j} - i_{2j}) \end{aligned} \tag{16}$$

and

$$w_t(0,0) = E[1 + \lambda_{1j} + \lambda_{2j})t] + W_0(0,0) \tag{17}$$

Proof Usual probabilistic conditioning arguments on the instant of the occurrence of the first unmet demand, the above theorem can easily be proved. See also Archibald et al. (1997). □

The expected discounted cost for the infinite horizon problem can now be derived. We recall the assumption that the unsold items will be returned just prior to the start of the next period at no cost and with full refund. We then make the order for the next period. Since, there is no fixed order cost in the model and the variable order cost is linear, the ordering decision at the start of any period is independent of the stock level just before the review. Hence, it is enough to consider only one state of the system at a review epoch. For convenience, that state is taken to be $(0,0)$ for each item $j = 1, 2$. Consequently, at any ordering instant, the system will not have any items in the warehouse. So, the decision problem is to know how many units to have at the start of any period. As per the definition, if we decide to have s_{kj} units of item j in warehouse k, with $k, j = 1, 2$, then the minimum expected total cost of satisfying the demands during the next period is $W_1(s_{1j}, s_{2j})$ given by (16). We now have the following theorem:

Theorem 2 (Archibald et al. (1997)) *If $V_j(0,0)$ is the infinite horizon minimum expected total discounted cost for item j in the two warehouses, then the optimal value function satisfies the following optimality equation.*

$$V_j(0,0) = \min \; \left\{ cs_{1j} + cs_{2j} + \beta(W_1(s_{1j}, s_{2j}) + V_j(0,0) \right\} \tag{18}$$

where β is the discounting factor for future costs.

Proof The proof is omitted as it is straightforward. □

As explained by Archibald et al. (1997), the finite horizon MDP is a single state MDP and hence there is only one decision to make, viz the order up to levels. This decision is clearly the order up to levels (S_{1j}, S_{2j}) that minimize the RHS of (18). It is also known (Archibald et al. (1997)) that the functions involved are convex with respect to the inventory levels and sub-modular (see Topkis (1978)) in time and inventory level variables. Thus, the following theorem characterizes the optimal policy:

Theorem 3 *For $j = 1, 2$, there exist non-negative integers S_{1j} and S_{2j} such that the optimal reorder policy is to order up to level S_{1j} at warehouse 1 and to order up to level S_{2j} at warehouse 2.*

Using value iteration, Archibald et al. also prove the following structural results.

Theorem 4 (Archibald et al. (1997)) *For $j = 1, 2$, there exist real values $\tau^1_{1j} \leq \tau^1_{2j} \leq \cdots \leq \tau^1_{m_1 j}$ such that the minimising action in state $(i, 0)$ when there is an unmet demand at warehouse 2 for item j and t time units to go until the next review epoch is to transfer an item from warehouse 1 to warehouse 2 if $t < \tau^1_{ij}$ and to place an emergency order otherwise. Similar result exists for the other warehouse.*

References

Archibald, T.W., Sassen, S.A.E., Thomas, L.C. An Optimal Policy for a Two Depot Inventory Problem with Stock Transfer. *Management Science*, Vol. 43, No. 2, 173–183 (1997).

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., *Nonlinear Programming: Theory and Algorithms.* John Wiley, New York (1993).

Everett III, H. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, Vol. 11, No. 3, 399–417 (1963).

Fisher, M.L. An applications oriented guide to Lagrangian relaxation. *Interfaces*, Vol. 15, No. 2, 10–21 (1985).

Held, M., Wolfe, P., Crowder, H. Validation of sub-gradient optimization. *Mathematical Programming*, Vol. 6, 62–88 (1974).

Koushik, S.R. *A Two-Item Two-Warehouse Periodic Review Inventory Model with Transshipment.* Master's thesis, Lee Kong Chian School of Business, Singapore Management University (2010).

Kranenburg, A.A., van Houtum, G.J. A new partial pooling structure for spare parts networks. *European Journal of Operational Research* Vol. 199, No. 3, 908–921 (2009).

Miller, R., Xu, S.H., Co, H. Clustering of skus to reduce split delivery cost and improve on-time delivery in online merchandising. *Working Paper*, College of Business, California State Polytechnic University, Pomona (2006).

Paterson, C., Kiesmüller, G.P., Teunter, R.H., Glazebrook, K.D. Inventory models with lateral transshipments: A review. *European Journal of Operational Research*, Vol. 210, No. 2, 125–136 (2011).

Sherbrooke, C.C. *Optimal Inventory Modeling of Systems.* Kluwer, Boston (2004).

Topkis, D.M. Minimizing a sub-modular function on a lattice. *Operations Research*, Vol. 26, No. 2, 305–321 (1978).

Wong, H., van Houtum, G.J., Cattrysse, D., van Oudheusden, D. Simple efficient heuristics in multi-item multi-location spare parts systems with lateral transshipments and waiting time constraints. *Journal of the Operational Research Society*, Vol. 56, No. 12, 1419–1431 (2005).

Wong, H., van Houtum, G.J., Cattrysse, D., van Oudheusden, D. Multi-item spare parts systems with lateral transshipments and waiting time constraints. *European Journal of Operational Research*, Vol. 171, No.3, 1071–1093 (2006).

Xu, P.J., Allgor, R., Graves, S.C. Benefits of reevaluating real-time order fulfillment decisions. *Manufacturing and Service Operations Management*, Vol. 11, No. 2, 340–355 (2009).