

THE IMPACT OF DIFFERENT DEMAND ALLOCATION RULES ON TOTAL STOCK LEVELS

Peter Wanke*

Centre for Logistics Studies
The COPPEAD Graduate School of Business
Federal University of Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ
peter@coppead.ufrj.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 02/2008; aceito em 06/2009
Received February 2008; accepted June 2009

Abstract

This manuscript compares two different rules for allocating demands to serving facilities and their impact on total stock levels: the one proposed by Tyagi & Das (1998) and the cross filling practice. The EOQ and the ROP methods are used for setting cycle and safety stocks. It is demonstrated that the minimization of the consolidated inventory levels within these rules leads to different allocation policies, frequently adopted by companies: one single facility sharing, dedicated facilities and full decentralization. Sensitivity analyses are conducted to identify the most relevant variables accountable for the differences in total stock levels among these three policies. Results suggest different benefit opportunities that may favor one policy to the detriment of the others. A framework synthesizing the findings is presented, so as to help in decision making. Potential impacts in terms of service levels and distribution costs are also evaluated qualitatively.

Keywords: allocation rules; stock levels; inventory portfolios.

Resumo

Esse artigo compara duas diferentes regras para alocação da demanda às instalações de serviço e seus impactos nos níveis totais de estoque: a regra proposta por Tyagi & Das (1998) e a prática de transferência entre instalações. Os métodos do lote econômico de compras e do ponto de pedido são usados para determinação dos níveis de estoque de ciclo e de segurança. É demonstrado que a minimização dos níveis agregados de estoque nessas regras leva a diferentes políticas de alocação freqüentemente adotadas pelas empresas: compartilhamento de uma única instalação de serviço, instalações de serviço dedicadas e descentralização total do atendimento. Análises de sensibilidade são conduzidas para identificar as variáveis mais relevantes, responsáveis pela diferença nos níveis totais de estoque entre essas três políticas. Os resultados sugerem diferentes oportunidades que podem favorecer uma dada política em detrimento das outras. Um quadro conceitual sintetizando os achados é apresentado, de modo a auxiliar na tomada de decisão. Os impactos potenciais em termos de níveis de serviço e dos custos de distribuição também são avaliados em termos qualitativos.

Palavras-chave: regras de alocação; níveis de estoque; portfólio de estoques.

1. Introduction

Research into virtual inventory management, variously referred to as demand allocation, inventory pooling, inventory consolidation, portfolio effect, and consolidation effect, has been around in literature in several forms for quite some time (Landers *et al.*, 2000; Ballou & Burnetas, 2003). The basic idea is that inventory increases as the standard deviation of either demand or lead time increases, and, as a result, companies may attempt to reduce inherent variation by pooling it. The statistical economies that arise from uncertainty pooling can be achieved in numerous ways, including not only inventory centralization, but also cross filling (Evers, 1999). Frequently, these economies are determined based upon an independent distribution system, in which each one of the markets is exclusively served by dedicated facilities.

Inventory centralization physically consolidates stock at a limited number of locations (often a single facility) from which all demand is satisfied. Since the mid 70's, a substantial amount of research on consolidation effects and inventory portfolios has greatly enhanced the understanding of the effects of facility consolidation on inventory levels, total costs, and the major variables to account for. Most of this research originated from the seminal works of Maister (1976) and Zinn *et al.* (1989) and their basic idea of using the 'square-root law' to measure safety stock savings derived from consolidation. Generally speaking, inventory centralization results in one single facility sharing; stocks should be consolidated at the facility with the smallest standard deviation of lead time.

Cross filling occurs when a facility satisfies whatever demand coming from another territory. On a regular basis, cross filling implies that a given proportion of demand is supplied from facilities located in different markets, regardless of whether there is sufficient inventory in the original serving facility. If cross filling results because the original serving facility is out of stock, then it will be known as an emergency transshipment (Evers, 1997).

Much research on cross filling has focused on mathematical modelling of particular systems. Krishnan & Rao (1965) modelled the costs for emergency transshipments from n locations. Tagaras (1989) generalized Krishnan & Rao's cost structure. Lee (1987), Axsäter (1990), and Kukreja *et al.* (2001) used queuing theory methods to estimate the cross filling effect in continuous review multi-location inventory systems with stochastic demand. In general, all these studies confirm that the cross filling practice reduce stock outs by improving inventory availability when compared to centralized systems. This occurs, however, to the detriment of higher distribution costs associated to longer distances for a given market (in case of regular transshipments) and to expedited shipping (in case of emergency transshipments).

Evers' (1996; 1997) and Ballou & Burnetas' (2003) studies specifically link the cross filling practice to portfolio and consolidation effects. Evers (1996) showed how the portfolio effect model could be employed to evaluate the use of regular transshipments so as to reduce safety stocks. The author also assessed under what conditions it might be effective to do so. The analysis neither considered the effect of transshipments on cycle stocks nor the impact on distribution costs. The author found that the use of cross filling should result in significant inventory reductions in cases where markets exhibit high degrees of demand variability. In Evers (1997), emergency transshipments were included into the portfolio effect analysis indicating that they result in the pooling of demand and lead time uncertainties. Finally, Ballou & Burnetas (2003) incorporated cycle stock consolidating effects to the portfolio effects on safety stock, when studying the cross filling between locations. Several restrictive assumptions were made regarding no lead time uncertainty, uncorrelated demands, and equal

inventory-related costs. The authors illustrated the countervailing inventory forces that affect the cross filling decision.

This paper aims to explore the impacts of inventory centralization and cross filling on total stock levels, employing the consolidation effect model in terms of their underlying demand allocation rules. More precisely, should a given centralized facility supply the same fraction of demand to each decentralized location, like in Tyagi & Das (1998), or should a fraction of the demand be supplied by a primary source and the remainder by secondary sources, like in Ballou & Burnetas (2003)? The impact of both allocation rules on the consolidation effect and their implications in terms of inventory levels are addressed in order to show how the consolidation effect model can be used as a cornerstone tool to evaluate under what conditions it may be preferable to adopt one demand allocation rule to the detriment of the other. These results are also compared to those of an independent distribution system, in which there are dedicated facilities to serve each market.

Specifically considering Tyagi & Das' (1998) and Ballou & Burnetas' (2003) results, this paper differs from them, not only for considering less restrictive assumptions in terms of lead time uncertainties, demand correlations, and inventory-related costs, but also for developing a framework for deciding the more adequate demand allocation rule in light of different lead time and demand characteristics.

This paper starts off by discussing the foundations of inventory consolidation planning; then, the two demand allocation rules are reviewed and linked to their respective inventory-pooling models in the ambit of the consolidation effect. Finally, several analytical expressions are derived and sensitivity analyses are performed, so as to address their adequacy in terms of product, demand, and operation characteristics for a minimal total cost.

2. A Brief Review on Inventory Consolidation Planning

Despite their shortcomings, the basic economic order quantity (EOQ) and the reorder point (ROP) models are widely used in the inventory consolidation research and practice. Their shortcomings are mostly related to the following assumptions: constant and perpetual demand, normally distributed demand and lead times, and demand and lead time independence (Nahmias, 2001). However, these models are the cornerstones of several software packages for inventory control, and their formulae are used to analyze important trade-offs related to inventory consolidation decisions (Lee & Nahmias, 1993).

With regard to the safety stock component embedded in ROP, several authors adopted variants of the safety-factor approach to set safety stocks both before and after consolidation when analyzing the portfolio effect. A non-exhaustive list of authors includes, for example, Maister (1976), Zinn *et al.* (1989), Caron & Marchet (1996), Evers & Beier (1998), Tyagi & Das (1998), and Das & Tyagi (1999). Their basic idea is that the safety stock savings, due to inventory consolidation, can be approximated by the statistical effect of the 'square-root law'. For instance, according to Maister (1976), the reduction in safety stocks is contingent to the square-root of the ratio between the number of centralized and decentralized locations. This effect, of course, depends upon the restrictive assumptions adopted, which were gradually relaxed over the years. For example, different standard deviations of demand were considered by Zinn *et al.* (1989); lead time uncertainty and unequal average lead times by Tallon (1993) and Evers & Beier (1998).

Cycle stocks were also considered in broader consolidation effect models. For instance, Evers (1995), Ballou & Burnetas (2003), and Ballou (2005) reexamined the traditional square-root formulation of the portfolio effects in light of this augmentation. Their basic idea was to account for the countervailing inventory force that arises when the consolidated cycle stock is considered. The cycle stocks savings are based on the fact that the square-root of the aggregate demand is always smaller than the sum of the square roots of individual demands.

The basic models to determine the consolidated levels of cycle and safety stocks, which were adopted as starting points in several of the previously mentioned researches, were firstly found in Evers & Beier (1993) and in Evers (1995), and they are presented in the next sections. However, before proceeding, readers should note that, if another inventory model is used where stock levels are directly pegged to demand levels, the effects on cycle and safety stocks, which are further detailed, will not necessarily occur (Maister, 1976; Ronen, 1990). This is what will probably happen, if, for instance, order-up-to-level models are used instead of EOQ/ROP models. The reason is the absence of square root terms in the order-up-to-level models.

2.1 Consolidated Cycle Stock

The total cycle stock as a result of the inventory consolidation from multiple n locations (or stocking points that serves demand) into a fewer number of m locations (where $n \geq m \geq 1$) is given by (Evers, 1995):

$$CSc = \frac{1}{2} \cdot \sum_{f=1}^m EOQ_f, \quad (1)$$

$$EOQ_f = \sqrt{2 \cdot \frac{P_f}{h_f} \cdot \sum_{i=1}^n (W_{i,f} \cdot D_i)}, \quad (2)$$

where: CSc = total cycle stock at m centralized locations,

EOQ_f = economic order quantity at centralized location f ,

D_i = mean demand during one time period at decentralized location i ,

P_f = fixed order processing cost at centralized location f (\$ per order),

h_f = unitary inventory holding cost per time period at centralized location f (\$ per time period), and

$W_{i,f}$ = proportion of mean demand during one time period transferred from decentralized location i to centralized location f (where $0 \leq W_{i,f} \leq 1$

for all i and f and $\sum_{f=1}^m W_{i,f} = 1$ for all i).

2.2 Consolidated Safety Stock

In its turn, the total safety stock at m centralized locations, as a result of the inventory consolidation from n decentralized locations is (Evers & Beier 1993):

$$SSc = k \cdot \sum_{f=1}^m \sqrt{\left(\sum_{i=1}^n W_{i,f} \cdot D_i \right)^2 \cdot s_{LT,f}^2 + \left[\left(\sum_{i=1}^n W_{i,f}^2 \cdot s_{D,i}^2 \right) + 2 \cdot \left(\sum_{i=1}^n \sum_{j=1}^{i-1} W_{i,f} \cdot W_{j,f} \cdot \rho_{i,j} \cdot s_{D,i} \cdot s_{D,j} \right) \right]} \cdot LT_f, \quad (3)$$

where: SSc = total safety stock at m centralized locations,
 LT_f = mean lead-time at centralized location f ,
 $s_{LT,f}$ = standard deviation of lead-time at centralized location f ,
 k = safety-stock factor,
 $\rho_{i,j}$ = correlation of demand between decentralized locations i and j , and
 $s_{D,i}$ = demand standard deviation during one period at decentralized location i .

It should be noted that equations (2) and (3) rest on the following assumptions:

1. All facilities use the safety-factor approach ($k \cdot s$, where s is the standard deviation of demand during lead-time) to setting safety stocks both before and after consolidation.
2. All facilities adopt the same safety-stock factor for desired service level.
3. All facilities adopt the EOQ approach for lot sizing and use the ROP for inventory control.
4. Average total system demand remains the same after consolidation.
5. Lead-times and demands are independent random variables.
6. Demand, lead-time and demand during lead-time are normally distributed variables.

Another important aspect to be mentioned about the inventory consolidation research is the allocation rule underlying these models. Should a given centralized facility supply the same fraction of demand to each decentralized location, like in Tyagi & Das (1998), or should the cross filling practice be adopted, like in Ballou & Burnetas (2003)? Under the cross filling, a fraction of the demand is supplied by a primary source and the remainder by secondary sources. The impacts of both allocation rules on total inventory levels and their implications for the consolidation decision are explored next. The results derived in the following sections consider the particular case where $n = m = 2$ and can be extended by induction to any value of m .

3. Tyagi and Das' Allocation Rule

Under Tyagi and Das' allocation rule (1998), a maximal consolidation effect on both cycle and safety stocks is obtained when inventories are centralized in one single facility. In this particular case, the condition of a given centralized facility supplying the same fraction of demand to each decentralized location is satisfied. In other words, their rule is given by:

$$W_{1,f} = W_{2,f} = \dots = W_{n,f} = W_f \text{ for any } 0 \leq W_f \leq 1 \text{ and } \sum_{f=1}^m W_f = 1 \text{ for all } i. \quad (4)$$

Considering the safety stocks at the centralized locations, replacing $W_{i,f}$ by W_f in equation (3), and observing the particular case where $n = m = 2$ (and by extension $W_2 = 1 - W_1$), SSc is given as follows:

$$SSc = k \cdot \left(\sqrt{W_1^2 \cdot D_C^2 \cdot s_{LT,1}^2 + W_1^2 \cdot s_{D,C}^2 \cdot LT_1} + \sqrt{(1-W_1)^2 \cdot D_C^2 \cdot s_{LT,2}^2 + (1-W_1)^2 \cdot s_{D,C}^2 \cdot LT_2} \right), \quad (5)$$

where: $D_C = D_1 + D_2$ is the total average demand, and

$$s_{D,C}^2 = s_{D,1}^2 + s_{D,2}^2 + 2 \cdot \rho_{1,2} \cdot s_{D,1} \cdot s_{D,2} \text{ is the total demand variance.}$$

Considering now the cycle stocks and replacing $W_{i,f}$ by W_f in equations (2) and (3), assuming that $h_f = h$ (unit holding costs) for all f , it comes for the $n = m = 2$ particular case that:

$$CSc = \sqrt{\frac{D_C}{2 \cdot h}} \cdot \left(\sqrt{P_1 \cdot W_1} + \sqrt{P_2 \cdot (1 - W_1)} \right). \quad (6)$$

4. Comparing with the Cross Filling Allocation Rule

Cross filling is a practice where demand for a stocking location results from the primary demand area as well as from the demand areas of other stocking locations (Ballou & Burnetas, 2003). Despite its negative impacts on cycle stocks, cross filling can be used either to reduce safety stock levels while maintaining very high fill rate levels. The cross filling effect on cycle and safety stocks can be generalized and compared to Tyagi and Das' allocation rule, considering the particular case where $n = m = 2$. Suppose that, under the cross filling, each centralized facility f supplies the same fraction of demand to its primary decentralized location i , that is:

$$W_{1,1} = W_{2,2}, W_{1,2} = W_{2,1} \text{ and } W_{1,2} = 1 - W_{2,2}. \quad (7)$$

This implies an allocation rule different from Tyagi & Das (1998), where $W_{1,1} = W_{2,1}$ and $W_{1,2} = W_{2,2}$. For instance, under the cross filling, 90% of demand of decentralized location 1 will be served from centralized facility 1 (primary flow) and 10% of it from centralized facility 2 (secondary flow). The opposite is true for decentralized location 2, where 90% is filled by centralized facility 2 (primary flow) and 10% by centralized facility 1 (secondary flow). Under Tyagi and Das' allocation rule, centralized (primary) facility 1 will serve 90% of demand of decentralized locations 1 and 2, and centralized (secondary) facility 2 will serve the other 10% at both decentralized locations. These differences are illustrated in Figure 1.

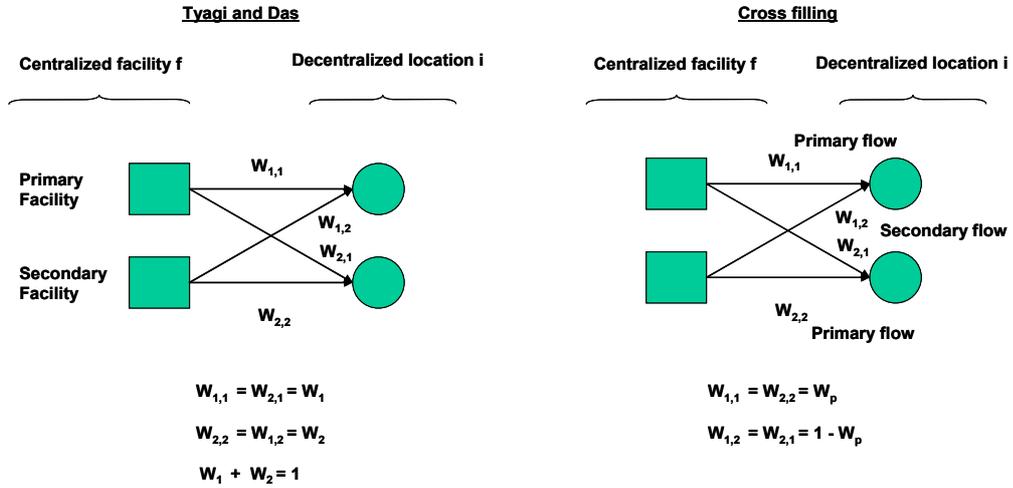


Figure 1 – Tyagi and Das and cross filling allocation rules.

4.1 Maximum Cycle Stock Savings

Applying the cross filling allocation rule in equations (2) and (3), and replacing $W_{1,1}$ and $W_{2,2}$ by W_p , $W_{1,2}$ and $W_{2,1}$ by $1 - W_p$ and h_f by h , it comes ($n = m = 2$):

$$CSc = \frac{1}{2} \cdot \sqrt{\frac{2}{h}} \cdot \left[\sqrt{[W_p \cdot D_1 + (1 - W_p) \cdot D_2] \cdot P_1} + \sqrt{[(1 - W_p) \cdot D_1 + W_p \cdot D_2] \cdot P_2} \right]. \quad (8)$$

Differentiating equation (8) with respect to W_p and equaling it to zero, one gets the W_p value that maximizes the consolidated cycle stock: $(P_2 \cdot D_2 - P_1 \cdot D_1) / [(P_1 + P_2) \cdot (D_2 - D_1)]$. If $P_2 = P_1$, W_p is equal to 0.5 like in Tyagi and Das' allocation rule. In Figure 2 the consolidated cycle stock behavior for Tyagi and Das and cross filling allocation rules is illustrated. One realizes that consolidation in one single facility minimizes the cycle stocks under Tyagi and Das' rule, whatever the conditions of the demand and of the fixed order processing costs are. The cycle stocks levels are considerably lower, when compared to the cross filling practice, for extreme values of W_1 .

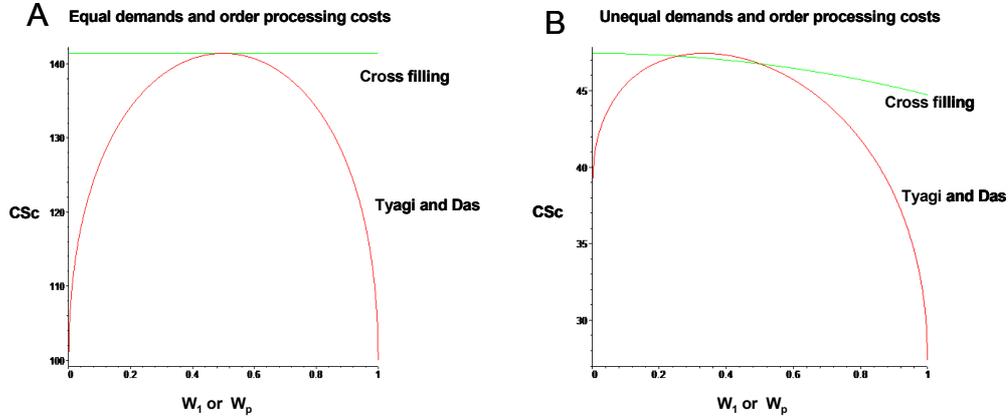


Figure 2 – Impact of the allocation rule on the consolidated cycle stock levels.

This can be proved for the particular case where demands ($D_1 = D_2 = D$) and order processing costs are equal and can be extended by induction for different conditions. When demands and order processing costs are equal, the consolidated cycle stock under the cross filling allocation rule is a flat line given by:

$$CSc = \sqrt{\frac{2 \cdot D \cdot P}{h}}, \quad (9)$$

but, under Tyagi and Das' allocation rule, it depends on W_1 :

$$CSc = \sqrt{\frac{D \cdot P}{h}} \cdot (\sqrt{W_1} + \sqrt{1 - W_1}). \quad (10)$$

One can realize that, when W_1 is 1 or 0, the consolidated cycle stocks under the cross filling are 1.41 times higher than under Tyagi and Das' allocation rule.

4.2 Maximum Safety Stock Savings

As regards the consolidated safety stocks, applying the cross filling allocation rule in equation (3) and considering equal lead-time ($LT_1 = LT_2 = LT$) and demand means and standard deviations ($s_{D,1} = s_{D,2} = s_D$ and $s_{LT,1} = s_{LT,2} = s_{LT}$), its expression is given next for the $n = m = 2$ case:

$$SSc = 2 \cdot k \cdot \sqrt{D^2 \cdot s_{LT}^2 + s_D^2 \cdot [W_p^2 + (1 - W_p) \cdot (1 + W_p \cdot (2 \cdot \rho_{1,2} - 1))]} \cdot LT. \quad (11)$$

Differentiating equation (11) with respect to W_p , it follows that the W_p value that minimizes the consolidated safety stock under the cross filling rule is 0.5. Under the same lead-time and demand conditions, Tyagi and Das allocation rule leads to a consolidated safety stock given as follows:

$$SSc = k \cdot \sqrt{4 \cdot D^2 \cdot s_{LT}^2 + 2 \cdot s_D^2 \cdot (1 + \rho_{1,2}) \cdot LT}. \quad (12)$$

In Figure 3 the consolidated cycle stock behavior for both allocation rules is illustrated. When lead-times and demands have equal means and standard deviations, the correlation coefficient of equilibrium that equals equations (11) and (12) is 1. Particularly, when W_p is equal to 1 or 0 and $\rho_{1,2} = 1$, equations (11) and (12) are reduced to $2 \cdot k \cdot \sqrt{D^2 \cdot s_{LT}^2 + s_D^2 \cdot LT}$ and the difference between these two allocation rules is zero. The same occurs when W_p is equal to 0.5, whatever the value of $\rho_{1,2}$ are. Outside these limiting conditions, the consolidated safety stocks are lower under Tyagi and Das allocation rule, regardless the correlation coefficient and the proportion of the decentralized demand to be allocated to a given centralized facility.

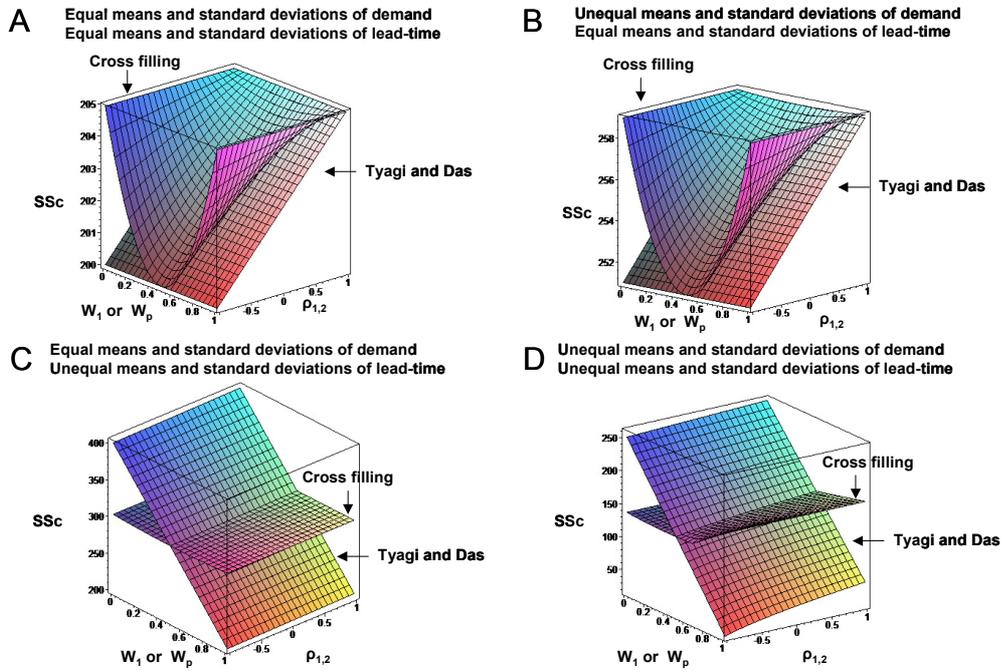


Figure 3 – Impact of the allocation rule on the consolidated safety stock levels.

4.3 Maximum Total Stock Savings

If, in one hand, minimal total stock occurs for extreme values of W_1 (0 or 1) under Tyagi and Das' rule, thus implying consolidation in one single facility, on the other hand, total stock may be minimal for intermediate values of W_p , under the cross filling practice. One can realize this fact via equation (8), when equal demands are considered: the cycle stock component does not longer depend on W_p and the safety stock is determinant to achieve the

minimal total stock (see Figures 2A and 3A). However, under different conditions of lead-time and demand, a quick inspection on Figures 2B and 3D indicates that total stock may be also minimal for extreme values of W_p . The next section addresses this issue with further details.

5. Sensitivity Analysis

Since the optimum value of W_p cannot be analytically evaluated by solving the total stock derivative (with respect to W_p) for different conditions of demand and lead-time, nonlinear programming and optimization was performed at AIMMS 3.8 on 10,000 different simulated scenarios of demand, lead-time, unitary inventory holding costs and fixed order processing costs. More precisely, the LGO routine (*Lipschitz-continuous Global Optimizer*), developed by *Pinter Consulting Services*, was used to find the optimal solutions within these scenarios.

The objectives of this sensitivity analysis are to evaluate:

- (a) the behavior of the optimal value of W_p under the cross filling allocation rule;
- (b) how this behavior relates to different allocation patterns or policies under the cross filling allocation rule;
- (c) the impacts of Tyagi and Das and cross filling allocation rules on total stock levels; and finally;
- (d) the most relevant variables accountable for the percent difference in total stock levels between these two allocation rules.

The nonlinear model implemented at AIMMS® is given next.

Minimize the total stock level under the cross filling practice:

$$\begin{aligned}
 & k \cdot \sqrt{\left(W_p \cdot D_1 + (1 - W_p) \cdot D_2 \right)^2 \cdot s_{LT,1}^2 + \left(W_p^2 \cdot s_{D,1}^2 + (1 - W_p)^2 \cdot s_{D,2}^2 + 2 \cdot (1 - W_p) \cdot W_p \cdot \rho_{1,2} \cdot s_{D,1} \cdot s_{D,2} \right) \cdot LT_1} + \\
 & k \cdot \sqrt{\left((1 - W_p) \cdot D_1 + W_p \cdot D_2 \right)^2 \cdot s_{LT,2}^2 + \left((1 - W_p)^2 \cdot s_{D,1}^2 + W_p^2 \cdot s_{D,2}^2 + 2 \cdot (1 - W_p) \cdot W_p \cdot \rho_{1,2} \cdot s_{D,1} \cdot s_{D,2} \right) \cdot LT_2} + \\
 & \frac{1}{2} \cdot \sqrt{2} \cdot \sqrt{\frac{\left(W_p \cdot D_1 + (1 - W_p) \cdot D_2 \right) \cdot P_1}{h}} + \frac{1}{2} \cdot \sqrt{2} \cdot \sqrt{\frac{\left((1 - W_p) \cdot D_1 + W_p \cdot D_2 \right) \cdot P_2}{h}},
 \end{aligned} \tag{13}$$

subject to:

$$0 \leq W_p \leq 1. \tag{14}$$

The parameters used to build these scenarios were chosen based on their test-values found in literature review papers (see the Appendix for listing of these parameters and papers). For each one of the 10,000 different scenarios, uniformly distributed random parameters were generated for demands, lead times, fixed order processing costs and unitary inventory holding costs, as it is displayed in Table 1.

Table 1 – Random generation of parameters.

Variables	Uniformly Distributed Random Parameters	
	Minimum	Maximum
D_1 and D_2 Per day	80	120
$s_{D,1}$ and $s_{D,2}$ Per day	3	30
LT_1 and LT_2 In days	1	5
$s_{LT,1}$ and $s_{LT,2}$ In days	0.50	2.00
$\rho_{1,2}$	-1	1
k	1	3
P_1 \$/order	17	67
P_2 \$/order	20	140
h \$/unit/day	0.35	0.68

The findings of the sensitivity analysis are addressed in the following sections.

5.1 Magnitude and Behavior of the Optimal Values of W_p

The distribution of frequencies for the optimal values of W_p is displayed in Figure 4. As expected, the optimal values of W_p are symmetrically distributed and concentrated around 0.50. In 7,199 of the 10,000 scenarios, the optimal solution for the demand allocation problem under the cross filling practice ranged within values greater than 0 and smaller than 1. For the remainder scenarios, the optimal solution was 0 and 1; totaling 1,397 and 1,404 cases, respectively.

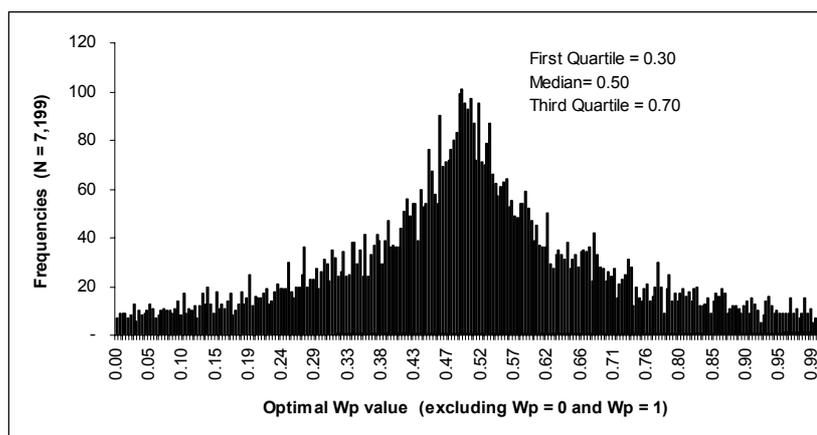


Figure 4 – Frequency distribution for optimal W_p values.

5.2 Allocation Patterns or Policies under the Cross Filling

Not only the optimal solutions under the cross filling behave differently from those of Tyagi and Das' allocation rule (0 or 1), but they also imply different allocation patterns or policies. Whereas both decentralized locations are exclusively served from a dedicated facility when the cross filling optimal solution is 0 or 1, under Tyagi and Das' allocation rule both decentralized locations share one single serving facility. In other words, differently from Tyagi and Das' allocation rule, an optimal W_p value equal to 0 or 1 does not imply demand consolidation into a single facility, neither the sharing of this single facility by both markets.

However, when the optimal values of W_p range from values greater than 0 and smaller than 1, a fully decentralized allocation pattern takes place under the cross filling. This means that both decentralized locations are served from all centralized facilities. Intermediate values of W_1 will also produce the same pattern under Tyagi and Das' allocation rule, but readers should recall that these are non-optimal solutions.

Therefore, three different possible allocation patterns or policies, which are commonly adopted within companies, emerge as a consequence of the optimization of the respective allocation rule (see Figure 5). In summary, they are: dedicated facilities (when W_p is equal to 0 or 1 under the cross filling practice), one single facility sharing or full consolidation (under Tyagi and Das' allocation rule), and full decentralization (when W_p ranges within values greater than 0 and smaller than 1 under the cross filling practice).

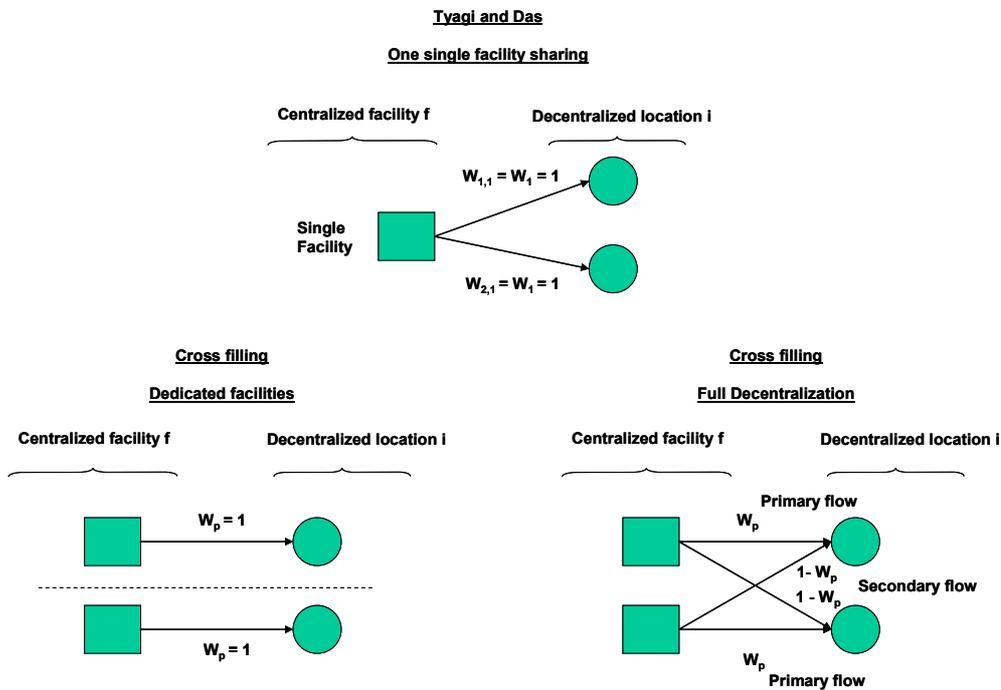


Figure 5 – Allocation patterns or policies.

Non-parametric tests were conducted to explore the differences in medians between the two cross filling patterns or policies: dedicated facilities and full decentralization. Not only the variables presented in Table 1 were tested, but also some relevant ratios calculated directly from these variables. They are given as follows:

$\alpha = LT_2 / LT_1$ is the ratio between lead-time means at locations 2 and 1,

$\beta = s_{LT,2} / s_{LT,1}$ is the ratio between lead-time standard deviations at locations 2 and 1,

$\gamma = P_2 / P_1$ is the ratio between fixed order processing costs at locations 2 and 1,

$\delta = D_2 / D_1$ is the ratio between demand means at locations 2 and 1, and

$\varepsilon = s_{D,2} / s_{D,1}$ is the ratio between demand standard deviations at locations 2 and 1.

The results for the Kruskal-Wallis Test and the Median Test are shown in Table 2 and significant differences ($p < 0.05$) between these two policies are flagged.

Table 2 – Non-parametric tests.

Variable	Impact of allocation policies on several variables					
	Median values		Kruskal-Wallis Test		Median Test	
	Dedicated Facilities (N = 2,801)	Full Decentralization (N = 7,199)	Chi-Square	Sig.	Chi-Square	Sig.
$\rho_{1,2}$	0.460	-0.160	1,105.710	0.000	717.736	0.000
% difference ^a	36.700	24.900	648.829	0.000	497.987	0.000
$s_{D,1}$	12.680	18.240	493.919	0.000	400.292	0.000
$s_{D,2}$	12.480	18.120	490.813	0.000	394.972	0.000
$s_{LT,2}$	1.380	1.210	123.559	0.000	99.460	0.000
LT_1	2.680	3.100	115.428	0.000	94.276	0.000
LT_2	2.680	3.070	95.994	0.000	69.494	0.000
β	1.160	0.970	50.402	0.000	80.750	0.000
k	1.960	2.020	12.341	0.000	8.958	0.003
h	0.330	0.350	3.346	0.067	3.325	0.068
D_1	100.650	99.760	3.234	0.072	4.379	0.036
$s_{LT,1}$	1.260	1.240	2.735	0.098	2.357	0.125
P_2	80.990	79.600	1.637	0.201	0.752	0.386
γ	1.940	1.870	1.347	0.246	3.364	0.067
α	1.020	1.000	0.639	0.424	1.862	0.172
P_1	42.450	42.020	0.481	0.488	0.575	0.448
δ	1.010	1.000	0.447	0.504	2.238	0.135
D_2	100.130	100.130	0.230	0.632	0.000	0.998
ε	0.990	0.990	0.003	0.955	0.024	0.876

^a Relative to Tyagi and Das allocation rule (full centralization or single facility sharing).

Both policies present a significant increase in total stock levels when compared to the single facility sharing policy, the latter as the result of the optimization of the Tyagi and Das' allocation rule. However, the median percent increase verified in the dedicated facilities policy is about 1.5 times higher than in the full decentralization policy. This effect occurs basically because the demands fluctuations during the lead-time cannot be compensated when facilities are exclusively dedicated to a decentralized location. Therefore, considering

the optimal solutions, correlation coefficients tend to be positive when centralized facilities are dedicated and negative whenever the centralized facilities serve all decentralized locations.

As a direct consequence of the compensation of the demands fluctuations during the lead-time, the full decentralization policy is more adequate to cases where lead-time means and demand standard deviations are higher when compared to the dedicated facilities policy. The former is also more adequate when both centralized facilities present equivalent standard deviation of lead-times. When the ratio between the lead-time standard deviations at different locations is high, the adoption of a dedicated facilities policy may be pertinent to avoid undesirable lead-time variability impacts at the decentralized location primarily served by the best performance facility.

5.3 Percent Differences in Total Stock Levels and Major Variables

The frequency distribution of the percent differences in total stock levels between Tyagi and Das and cross filling allocation rules is shown in Figure 6. One can realize that this distribution is moderately concentrated on the left side of the graph, thus indicating that the percent difference in the majority of cases ranges from 5 to almost 30%.

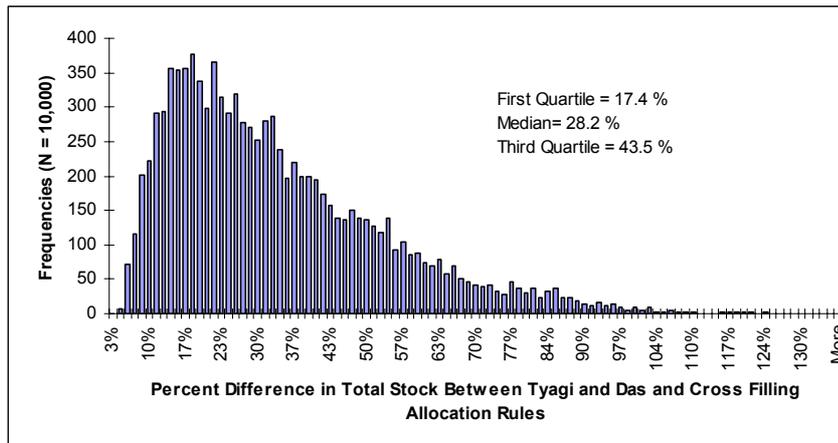


Figure 6 – Frequency distribution for the percent difference in total stock levels.

In order to identify the most relevant variables accountable for the percent difference in total stock levels, a correlation analysis between these differences and the variables presented in Table 3 was conducted. Readers should recall that whereas Tyagi and Das' allocation rule is solely related to the one single facility sharing policy, the cross filling rule is related to two different policies: dedicated facilities and full decentralization. Therefore, a dummy variable named policy (*Pol*) was created and included in the analysis. The value 1 was attributed to the full decentralization policy. The results of the correlation analysis are presented in Table 3.

Table 3 – Correlation coefficients.

Variable	Pearson Correlation	Sig. (2-tailed) N = 10,000	Variable	Pearson Correlation	Sig. (2-tailed) N = 10,000
$\beta = s_{LT,2} / s_{LT,1}$	0.533	0.000	$\delta = D_2 / D_1$	-0.020	0.050
$s_{LT,1}$	-0.375	0.000	$s_{D,1}$	-0.016	0.118
Pol	-0.232	0.000	D_2	-0.013	0.183
γ	0.102	0.000	LT_1	-0.013	0.190
h	-0.086	0.000	D_1	0.010	0.302
k	-0.072	0.000	$s_{D,2}$	-0.008	0.406
P_1	0.070	0.000	LT_2	-0.008	0.443
P_2	-0.063	0.000	$\alpha = LT_2 / LT_1$	0.005	0.602
$s_{LT,2}$	0.036	0.000	$\varepsilon = s_{D,2} / s_{D,1}$	0.000	0.996
$\rho_{1,2}$	-0.033	0.001			

One can realize that the two most significant variables related to the percent differences between Tyagi and Das and cross filling allocation rules are: the ratio between lead-times at locations 2 and 1 (β) and the optimal policy (Pol) adopted under the cross filling rule.

The greater the value of β , the higher the difference in total stock levels in favor to Tyagi and Das' allocation rule (one single facility sharing or full consolidation policy). When demands are allocated to centralized facilities with discrepant levels of lead-time variability, the safety stock savings opportunity from consolidating total demand at the best performance facility is being missed.

As mentioned before, the full decentralization policy is related to a lower percent difference in total stock levels when compared to the dedicated facilities policy. Besides that, Pol is also correlated with β , k , and $\rho_{1,2}$, among other variables. Therefore, the reasons for this behavior and, implicitly, the adequacy of one policy to detriment of the other will be explained in terms of these variables.

Returning to β , the major decision is: which is the best allocation policy, full decentralization or dedicated facilities, when the consolidation opportunity at the best performance facility, in terms of safety stocks savings, is missed? One must decide between serving positively correlated demands from dedicated facilities, which may also prevent from undesirable lead-time variability impacts on service levels, and serving negatively correlated demands from all centralized facilities, in order to benefit from the compensation of the demands fluctuations, despite the cross-effects on lead-time variability to each decentralized location.

In summary, when adopting the one single facility sharing policy, one is benefiting both from the consolidation at the best performance facility in terms of lead-time variability and from the compensation of demand fluctuations at decentralized locations. However, when adopting the full decentralization policy, one just benefits from the compensation of demand fluctuations at decentralized locations. And finally, when adopting the dedicated facilities policy, none of these benefits is achieved except for the fact that undesirable cross-effects in terms of lead-time variability and, thus, on service levels, are avoided.

Specifically considering k , the greater the safety-stock factor, the smaller the percent difference in total stocks between Tyagi and Das and cross filling allocation rule. Although this effect is purely mathematical and occurs due to increases in the safety stock terms at the numerator

and the denominator, with cycle stocks terms remaining unchanged, it still reflects the common sense that high service level networks may often imply decentralized allocation policies.

Now, regarding $\rho_{1,2}$, the smaller the correlation, the greater the percent difference in total stocks between these two allocation rules. This is because the benefits of a full inventory consolidation, in terms of safety stocks, increase while the correlation decreases. Under the cross filling, the possibility of exploiting this benefit will lead to the full decentralization policy, principally if lead-time means and standard deviations of demand are higher. Whenever the correlation is high, the dedicated facilities policy will be preferable for lower lead-time means and standard deviations of demand.

At last, two other relevant variables for explaining differences in total stock levels are the unitary inventory holding cost and the ratio between fixed order processing costs at decentralized locations 2 and 1.

As regards the latter variable, its interpretation is similar to β : the greater the value of γ , the higher the difference in total stock levels in favor to Tyagi and Das' allocation rule. When demands are allocated to centralized facilities with discrepant levels of fixed order processing costs, the cycle stock savings opportunity from consolidating total demand at the lowest cost facility is missed. In its turn, the greater the inventory holding cost, the smaller the percent difference in total stock levels, basically because cycle stocks tend to be lower in both allocation rules, thus reducing the benefits of a full consolidation of inventories.

Two linear models for the percent difference in total stock levels between Tyagi and Das and cross filling allocation rule are presented in Table 4: one for $\beta \geq 1$ and the other for $\beta < 1$.

A stepwise regression method was conducted with SPSS. Only the most significant variables were included in the models, while simultaneously controlling for multicollinearity. It is important to mention that:

- both models present a good explanatory power for the percent differences;
- β is the most significant variable in both models, and its coefficient is greater than the summation of the coefficients of the other variables;
- in both models, the coefficient signs indicate that Tyagi and Das' allocation rule is favored when the correlation is negative, the order processing cost is high in relation to the inventory holding cost and the safety-stock factor is low;
- in the model for $\beta < 1$, exchanged signs in the remainder variables suggest that it is better to consolidate stocks in location 2.

Table 4 – Linear models for the percent difference in total stock levels.

Linear Model for $\beta \geq 1$					Linear Model for $\beta < 1$				
Variables	B	Std. Error	t	Sig.	Variables	B	Std. Error	t	Sig.
Constant	-0.1490	0.0050	-32.657	0.000	Constant	0.8290	0.0070	122.031	0.000
β	0.3310	0.0020	217.450	0.000	β	-0.6960	0.0060	-109.176	0.000
γ	0.0474	0.0010	73.078	0.000	γ	-0.0217	0.0010	-23.999	0.000
h	-0.1390	0.0050	-30.412	0.000	$\rho_{1,2}$	-0.0132	0.0020	-6.247	0.000
k	-0.0344	0.0020	-22.556	0.000	h	-0.0368	0.0060	-5.789	0.000
$\rho_{1,2}$	-0.0160	-0.0160	-10.483	0.000	k	-0.0104	0.0020	-4.825	0.000
F (N = 5,064)	Sig.	R²	Adj. R²	Std. Error	F (N = 4,936)	Sig.	R²	Adj. R²	Std. Error
10,988	0.000	0.915	0.915	0.06329	2,531	0.000	0.719	0.718	0.08689

6. Managerial implications and Conclusions

Companies frequently link demands at decentralized locations to serving facilities based on allocation policies. This manuscript explored three of them (one single facility sharing, full decentralization and dedicated facilities) in terms of their adequacy to different variables and their impact on total consolidated stock levels.

While providing managers with a framework to decide among these policies, this research tracked their origins with respect to two major allocation rules discussed in several papers on portfolio/consolidation effect: Tyagi and Das' allocation rule and the cross-filling practice. This research also evaluated, via non-linear programming and optimization, the typical optimal allocation fractions under the cross filling and compared the respective optimal stock levels to those obtained with Tyagi and Das' rule, where inventories are centralized in one single facility. A summary of the research major results is provided in Table 5.

Table 5 – A framework to choose among different allocation policies.

Allocation Rules	Tyagi & Das (1998)		
Allocation Policies	One Single Facility Sharing	Cross filling	
Allocation Policies	One Single Facility Sharing	Dedicated Facilities	Full Decentralization
Optimal W value	0 or 1	0 or 1	Ranges within 0 and 1
Adequacy	<ul style="list-style-type: none"> • High negative correlation; • High lead-time variability discrepancies within locations. 	<ul style="list-style-type: none"> • High positive correlation; • Low standard deviations of demand; • Low lead-time means. 	<ul style="list-style-type: none"> • Moderate negative correlation or weak positive correlation; • High standard deviations of demand; • High lead-time means.
Rationale of Benefit	One benefits both from the consolidation at the best performance facility in terms of lead-time variability and from the compensation of demand fluctuation.	Except for the fact that undesirable cross-effects in terms of lead-time variability are avoided, none of the already mentioned benefits is achieved.	One just benefits from the compensation of demand fluctuation.

Despite higher total stock levels, the cross filling allocation rule may exact its price when service levels and transportation costs are taken into consideration. The service level effect occurs basically because the average cycle stock level allocated to a given facility is proportionally higher than the average lead-time demand variability allocated to the same give facility under the cross filling. This is particularly true when a full decentralization policy is adopted.

Considering the well-known result for determining fill rate (FR), in which cycle and safety stocks are traded-off, $FR = 1 - f(k) \cdot s / Q$ (where $f(k)$ is a function of the normal loss curve, s is the standard deviation of the lead-time demand allocated to a given facility and Q is the replenishment order quantity), it is easy to note that, under the cross filling practice, the safety- factor k tends to be smaller when compared to Tyagi and Das' allocation rule for a given facility and a desired fill rate level to be achieved. This effect is illustrated in Table 6 and equations (9), (10), (11) and (12) were used to build up this table.

Table 6 – Service level effect.

	Tyagi and Das	Cross filling ($W_p = 0.5$)			Cross filling ($W_p = 0$ or 1)		
	Single Facility	Centralized Location 1	Centralized Location 2	Difference	Centralized Location 1	Centralized Location 2	Difference
Cycle stock	31.62	22.36	22.36	13.10	22.36	22.36	13.10
Variability of lead-time demand	20.00	10.00	10.00		14.14	14.14	
k (Fill Rate = 98%)	1.47	1.31	1.31		1.47	1.47	
Safety stock	29.40	13.10	13.10	(3.20)	20.79	20.79	12.18
Total stock at the facility	61.02	35.46	35.46	9.90	43.15	43.15	25.28
Total consolidated stock	61.02	70.92			86.30		
Percent difference		16%			41%		

Test-values for the $n = m = 2$ particular case: $D_1 = D_2 = 100$; $P_1 = P_2 = 10$; $h = 1$; $s_{D,1} = s_{D,2} = 10$; $LT_1 = LT_2 = 2$; uncorrelated demands and no lead-time variability.

In its turn, with respect to transportation, centralization may certainly increase distribution costs, especially if stocking items present low unitary holding costs and the order processing costs are low. However, the cross-filling practice allows a higher degree of decentralization and, therefore, greater proximity to customers and lower distribution costs, with a moderate increase in total stock levels, particularly when lead-time standard deviations at centralized locations present equivalent magnitude. The choice of the appropriate demand allocation policy requires an in-depth understanding of the total consolidated stock behavior and its response to different allocation rules. Managers should be aware of the opportunities for improving customer service via deciding for dedicated facilities or a full decentralization policy, with minimal increases in aggregate inventory, while keeping distribution costs low.

For instance, product inventories which result in larger savings from consolidation in one single facility will provide clear opportunities for total cost reduction and the one single facility sharing decision should be taken into consideration. Nevertheless, inventories which result in lower savings from consolidation may represent pitfalls in terms of increasing total costs, especially if these inventories present both low per-unit holding costs and fixed order processing costs. At the same time, these inventories represent opportunities for improving customer service via decentralization, with minimal increases in aggregate inventory and, eventually, with decreases in total costs. This is particularly true when the lead-time variability discrepancy between locations is low and the correlation coefficient is positive.

References

- (1) Axsäter, S. (1990). Modelling emergency lateral transshipments in inventory systems. *Management Science*, **36**(11), 1329-1338.
- (2) Ballou, R.H. & Burnetas, A. (2003). Planning multiple location inventories. *Journal of Business Logistics*, **24**(2), 65-89.
- (3) Ballou, R.H. (2005). Expressing inventory control policy in the turnover curve. *Journal of Business Logistics*, **26**(2), 143-164.
- (4) Caron, F. & Marchet, G. (1996). The impact of inventory centralization/decentralization on safety stock for two-echelon systems. *Journal of Business Logistics*, **17**(1), 233-257.

- (5) Das, C. & Tyagi, R. (1999). Effect of correlated demands on safety stock centralization: patterns of correlation versus degree of centralization. *Journal of Business Logistics*, **20**(1), 205-213.
- (6) Evers, P.T. & Beier, F.J. (1993). The portfolio effect and multiple consolidation points: a critical assessment of the square root law. *Journal of Business Logistics*, **14**(2), 109-125.
- (7) Evers, P.T. (1995). Expanding the square root law: an analysis of both safety and cycle stocks. *The Logistics and Transportation Review*, **31**(1), 1-20.
- (8) Evers, P.T. (1996). The impact of transshipments on safety stock requirements. *Journal of Business Logistics*, **17**(1), 109-133.
- (9) Evers, P.T. (1997). Hidden benefits of emergency transshipments. *Journal of Business Logistics*, **18**(2), 55-76.
- (10) Evers, P.T. & Beier, F.J. (1998). Operational aspects of inventory consolidation decision making. *Journal of Business Logistics*, **19**(1), 173-189.
- (11) Evers, P.T. (1999). Filling customer orders from multiple locations: a comparison of pooling methods. *Journal of Business Logistics*, **20**(1), 121-139.
- (12) Krishnan, K.S. & Rao, V.K. (1965). Inventory control in n warehouses. *Journal of Industrial Engineering*, **16**(3), 212-215.
- (13) Kukreja, A.; Schmidt, C.P. & Miller, D.M. (2001). Stocking decisions for low-usage items in a multilocation inventory system. *Management Science*, **47**(10), 1371-1383.
- (14) Landers, T.L.; Cole, M.H.; Walker, B. & Kirk, R.W. (2000). The virtual warehousing concept. *Transportation Research Part E*, **36**(2), 115-125.
- (15) Lee, H.L. (1987). A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, **33**(10), 1302-1316.
- (16) Lee, H.L. & Nahmias, S. (1993). Single-product, single-location models. **In:** *Logistics of Production and Inventory* [edited by S.C. Graves, A.H.G. Kahn and P.H. Zipkin], North-Holland, Amsterdam, 3-55.
- (17) Maister, D.H. (1976). Centralisation of inventories and the 'square root law'. *International Journal of Physical Distribution & Materials Management*, **6**(3), 124-134.
- (18) Nahmias, S. (2001). *Production and Operations Analysis*. 4th ed. McGraw-Hill, New York.
- (19) Ronen, D. (1990). Inventory centralization/decentralization – the 'square root law' revisited again. *Journal of Business Logistics*, **11**(2), 129-138.
- (20) Tagaras, G. (1989). Effects of pooling on the optimization and service levels of two-location inventory systems. *IIE Transactions*, **21**(3), 250-257.
- (21) Tallon, W.J. (1993). The impact of inventory centralization on aggregate safety stock: the variable supply lead time case. *Journal of Business Logistics*, **14**(1), 185-203.
- (22) Tyagi, R. & Das, C. (1998). Extension of the square root law for safety stock to demands with unequal variances. *Journal of Business Logistics*, **19**(2), 197-203.
- (23) Zinn, W.; Levy, M. & Bowersox, D.J. (1989). Measuring the effect of inventory centralization/decentralization on aggregate safety stock: the 'square root law' revisited. *Journal of Business Logistics*, **10**(2), 1-14.

Appendix – Parameters collected from selected papers.

Primary Parameters												Secondary Parameters						
D	LT	CI/d	CI/t	s _D	s _{LT}	k	ρ _{1,2}	P	C	i	h = C * i	P/h	CI/d = s _n / D	CI/t = s _{LT} / LT	α	β	γ	
Tallon (1993) ^a	500; 650; 900; 1000; 1200	1; 2; 3; 5		20; 25; 100; 200; 250	0.25; 0.50	1.65	-1; +1						0.038; 0.04; 0.11; 0.17; 0.25	0.05; 0.083; 0.10; 0.125; 0.167; 0.25; 0.50	0.2; 0.33; 0.4; 0.5; 0.6; 0.67; 1.5; 1.67; 2; 2.5; 3; 5			
Caron and Marchet (1996) ^b		5			1; 2		0.2; 0.8						0.2; 0.4			0.5; 2		
Evers (1997) ^c	100; 300; 500	5; 10; 15	0.1; 0.3; 0.5												0.33; 0.5; 0.67; 1.5; 2; 3			
Tyagi and Das (1998) ^d				1; 3; 10														
Evers (1999) ^e	50; 100	5; 10	0.25; 0.5		0; 0.5										0.5; 2			
Das and Tyagi (1999) ^f							-1; +1											
Ballou and Burnetas (2003) ^g	77; 350; 750	6		25; 150; 300		1.65		10	200	2.5%	50	0.2						
Ballou (2005) ^h	100	5		10				50	3; 10	2.5%	0.75; 2.5	20; 66.7	0.10					
Ballou (2005) and Ballou and Burnetas (2003)											0.75; 2.5	20; 66.7	0.10					0.2; 5

^a Average demand per period; average lead-time in time periods; standard deviation of demand per period; standard deviation of lead-time in time periods (p. 187 and p. 194);

^b Average lead-time in days; standard deviation of lead-time in days (p. 244);

^c Average demand per period; coefficient of variation of demand per period; average lead-time in time periods (p. 64);

^d Standard deviation of demand per period (p. 201);

^e Average demand per day; coefficient of variation of demand per day; average lead-time in days (p. 126);

^f p. 209;

^g C = item cost (\$/unit); i = carrying cost (\$/unit); P = order cost (\$/order); average lead-time in weeks; average demand per week; standard deviation of demand per week (pp. 75-76);

^h C = item cost (\$/unit); i = carrying cost (\$/unit); P = order cost (\$/order); average demand per day; standard deviation of demand per day; average lead-time in days (p. 148).