Managing single echelon inventories through demand aggregation and the feasibility of a correlation matrix

Kefeng Xu\textsuperscript{a}, Philip T. Evers\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a}College of Business, University of Texas, San Antonio, 6900 North Loop 1604 West, San Antonio, TX 78249, USA
\textsuperscript{b}Robert H. Smith School of Business, University of Maryland, 3417 Van Munching Hall, College Park, MD 20742, USA

Received 1 September 2000; received in revised form 1 June 2001

Abstract

This paper examines the assertion that partial pooling of customers is sometimes favored over complete pooling on the sole basis of demand correlation. First, the management of inventory within a supply chain is discussed, with specific attention paid to risk-pooling. Then, the claim that partial pooling can dominate is theoretically discussed. The conditions under which previous research found that partial aggregation could, at times, be preferable are investigated next. From this, methods are proposed for checking correlation matrices to ensure their validity. It is concluded that partial pooling can never do better and that examples supporting partial aggregation are based on inconsistent correlation matrices.

Scope and purpose

Arising in part from uncertainty, inventory is a key component of the supply chain. This paper reviews alternative methods for managing demand uncertainty across various levels of the supply chain as well as within a single echelon of the system. Particular attention is paid to the fulfillment of customer demands from multiple stock-keeping facilities versus from a single location, with the superiority of the latter presented in terms of requisite safety stock levels. In doing so, methods for ensuring the feasibility of correlation matrices for simulation and numerical analyses are also highlighted. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Single echelon inventories; Decentralization of facilities; Customer grouping; Correlation matrices

1. Introduction

The mission of supply chain management is to coordinate the activities of all relevant entities for the purpose of providing goods and services to the end customer in an efficient and effective
manner. One function that receives considerable attention in this alignment is the management of inventory. A typical goal is to increase supply chain velocity by minimizing total inventories within the supply chain, while maintaining an appropriate level of customer service. By doing so, products flow faster from raw materials sources to final consumers, leading to lower holding costs, less obsolescence, and reduced storage-space requirements, among other things.

Even in a fully functional supply chain, however, there is justification for holding inventory. No matter how good the demand forecasting system being used is, a firm usually cannot predict end customer orders with complete accuracy. Assuming that the customer is unwilling to wait for the product to be manufactured, inventories must be maintained to meet demand as it occurs. In other situations, it simply may not be economically feasible to produce on a just-in-time basis. Even in the case of Internet-based companies like e-tailers, the expected response time to requests may be such that items must be produced prior to actual customer requirements. As a consequence, methods are needed for efficiently maintaining inventories in general and, of particular interest here, inventories (commonly known as safety stocks) used to meet unexpected customer requirements.

In Tyagi and Das [1], a model and solution procedure for grouping customers in order to minimize total resource requirements such as safety stocks was proposed. Their work stemmed from the notion that demand correlations can be optimally combined in such a way that partial pooling of customers into multiple groups could, in some cases, result in fewer resource requirements than complete pooling of customers into one group. This paper, while not contesting their solution technique, objects to their fundamental contention that partial aggregation of customers can be better than complete aggregation based on demand correlations alone and argues that complete aggregation of demand is always as good as, if not better than, partial aggregation on the basis of minimized total resources. Further, this paper will establish that the results from Tyagi and Das [1] claiming to show partial aggregation to be better were based on mathematically inconsistent correlation matrices. To do so, this paper will demonstrate the existence of feasible and infeasible correlation structures and describe methods that allow researchers to check the validity of a correlation matrix for both numerical and simulation purposes. But before addressing these issues, the notion of pooling and its relationship to inventory management will be explained first.

2. Statistical economies of scale

Within supply chains, the principle of postponement relies heavily on statistical economies of scale, a term originally coined by Eppen and Schrage [2] and defined as “advantages that result from the pooling of uncertainty [3, p. 51].” Postponement recognizes that when changes in the form, identity, and/or location of a product are delayed, savings accrue because demand is easier to predict [4]. Not only does pushing inventory back up the supply chain lead to lower product valuations since fewer costs have been incurred up to the point at which the item is being held in inventory and, consequently, to reductions in carrying costs, but it also results in statistical economies as forecast uncertainty is reduced. The opposite of postponement is speculation, where changes in a product’s form, identity, and/or location are performed early on in order to take advantage of economics of scale arising from long production runs, large orders, and large shipments [5]. Bucklin explained that the opposing forces of postponement and speculation “form a basis for determining whether speculative inventories … will appear in distribution channels [5, p. 27].” Moreover, he noted that
postponement–speculation could be measured in terms of delivery time; as delivery times increase (decrease), the level of postponement (speculation) increases. Zinn and Bowersox [6] found product value to be the most important factor behind the decision to postpone form and identity changes and found demand uncertainty to be the most important factor behind the decision to postpone location changes. In the latter case, as demand uncertainty increases, the benefits to holding inventory upstream grow. Further discussion of the postponement–speculation tradeoff can be found in van Hoek [7].

Related to postponement–speculation is multi-echelon research focusing on the effects of carrying inventory at different levels of the supply chain. Eppen and Schrage [2] showed that a system in which a central depot places orders for all demand locations at once, but carries no inventory, realizes statistical economies. They divided statistical economies in multi-echelon environments into two distinct effects: the joint-ordering effect and the depot effect. The joint-ordering effect arises because the depot can take advantage of quantity discounts offered by suppliers and of risk-pooling during lead times from supplier to depot (i.e., the final allocation of the order is made when the shipment from the supplier is received at the depot, not when the order from the depot is placed with the supplier). The depot effect arises when the central depot is allowed to hold inventory and allocate it at later dates as future inventory needs become clearer (see Jackson and Muckstadt [8] for more on the depot effect). Substantial research has stemmed from the Eppen and Schrage analysis of multi-echelon inventory; Schwarz [9] provides a concise review of a number of these early works. A succinct summary of managerial implications from multi-echelon inventory research can be found in Waller and Rosenbaum [10]. In general, savings from holding inventory upstream increase as any of the following occur: demand variability increases, the number of downstream stocking facilities increases, the service level increases, or replenishment lead times increase.

The use of multiple echelons mirrors postponement in that inventory, especially safety stock, is carried upstream to achieve cost savings. Like postponement, a significant drawback of holding inventory at multiple echelons is the response time to customer demand. This point has been raised by Schwarz [9] who questioned the multi-echelon approach to risk-pooling. He noted that, although there may be an incentive to using a depot in terms of reducing overall variance, the additional lead time needed to move product through the depot may be prohibitive. Thus, there is good reason to also consider the management of speculative inventories within a single echelon at the downstream end of the supply chain. These finished goods inventories, maintained to ensure quick response to customer demands, represent the stocks from which customers can directly draw from, but tend to be considerably more costly to hold. Nevertheless, statistical economies can also arise in the case of single echelon inventories because it is quite likely that some stock-keeping facilities will experience demand that is higher than expected when, at the same time, others are facing demand that is lower than expected [11]. Therefore, just as forecasting demand for a family of products is usually more accurate than forecasting demand for single items in the case of postponement, aggregated customer demands should lead to lower inventory requirements than disaggregated demand in the case of speculation.

There are two primary forms of demand aggregation: physical and virtual. Physical aggregation, or inventory centralization, occurs when inventories of stock-keeping locations are actually consolidated into a smaller number of locations. Virtual aggregation, or information centralization, occurs when the stock-keeping locations remain decentralized, but management of the inventory is centralized and lateral transshipments between facilities are allowed. In accordance with the analysis of Tyagi
and Das [1], only physical aggregation will be examined here (for more on virtual aggregation, see Tagaras [12]).

3. Partial pooling is never better

To understand why complete pooling can never do worse than partial pooling, consider the following as it relates to physical aggregation. Individual customers are initially classified into \( n \) distinct markets on some basis relevant to the firm (say, geographic area). As a starting point, each of these markets is assigned to a different stock-keeping location. Upon pooling, some or all of these markets are coupled together. Thus, pooling can range from one market per location (representing the starting point and called complete disaggregation) to all \( n \) markets served from one location (termed complete aggregation). Partial aggregation lies between these extremes: for example, some markets may be combined while others are not. Now, recall that Tyagi and Das [1] focus only on the sum of the standard deviations of demand (i.e., Eq. (4) of [1]); therefore, the level of safety stock maintained at each location is equal to the standard deviation of demand at that location multiplied by some exogenously determined safety factor (for convenience, assume that the same safety factor is used at all locations and is equal to, say, 1). Using notation similar to theirs

\[
S_k = \sqrt{\sum_{i=1}^{n} Z_{ik}^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} Z_{ik} Z_{jk} r_{ij} \sigma_i \sigma_j},
\]

where \( S_k \) is the standard deviation of annual demand at location \( k \), \( \sigma_i \) the standard deviation of annual demand for market \( i \), \( r_{ij} \) the correlation coefficient between annual demands at markets \( i \) and \( j \), \( Z_{ik} = 1 \) if demand from market \( i \) is assigned to location \( k \), 0 otherwise, and \( n \) is the total number of markets.

In terms of inventory centralization, Eppen [13] has shown that complete disaggregation is equally preferable to complete aggregation when \( r_{ij} = +1 \); however, when \( r_{ij} < +1 \), complete aggregation is always favored. Recognizing that partial aggregation is simply a form of disaggregation with fewer stock-keeping locations (but more than one), complete aggregation is never suboptimal on the basis of demand correlations alone.

A simple example illustrates this principle, which underlies the notion of statistical economies and is sometimes referred to as the portfolio effect (cf., Schwarz [14]). Without loss of generality, consider the case of three locations, where \( \sigma_1 \leq \sigma_2 \leq \sigma_3 \). If perfect, positive correlation exists among all three locations \( r_{12} = r_{13} = r_{23} = +1 \), then complete aggregation leads to

\[
S_{123} = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_1 \sigma_2 + 2\sigma_1 \sigma_3 + 2\sigma_2 \sigma_3}
\]

or

\[
S_{123} = \sigma_1 + \sigma_2 + \sigma_3,
\]

which requires the same level of safety stock as complete disaggregation \( (S_1 + S_2 + S_3 = S_{123}) \). Any form of partial aggregation also leads to the same result (for example, \( S_{12} + S_3 = S_{123} \)).
At the other extreme, when perfect, negative correlation exists within the locations ($r_{12} = r_{13} = -1$, $r_{23} = +1$; as will be explained in more detail later, it is not possible for all three $r_{ij}$ to be $-1$), then complete aggregation leads to

$$S_{123} = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 - 2\sigma_1\sigma_2 - 2\sigma_1\sigma_3 + 2\sigma_2\sigma_3}$$

or

$$S_{123} = \sigma_2 + \sigma_3 - \sigma_1,$$

which requires fewer safety stocks than complete disaggregation ($S_1 + S_2 + S_3 > S_{123}$). While some forms of partial aggregation are as beneficial as complete aggregation ($S_{12} + S_3 = S_{13} + S_2 = S_{123}$), no form of partial aggregation does better. Switching the negative correlations among the locations (for example, $r_{12} = +1$, $r_{13} = r_{23} = -1$) leads to the same conclusions. Indeed, no feasible combination of $r_{ij}$ can be found that results in partial pooling dominating complete pooling as it is always the case that $S_{123} \leq S_1 + S_2 + S_3$ (cf., Schwarz [14]).

4. Results based on inconsistent correlation matrices

Having shown that partial pooling cannot do better than complete pooling, a logical question arises. How were Tyagi and Das [1] able to claim that partial aggregation was preferred in some cases? In particular, of the four different correlation matrices examined ($R_1$, $R_2$, $R_3$, and $R_4$), they found that partial aggregation dominated in three cases ($R_1$, $R_2$, and $R_3$) while complete aggregation dominated in only one ($R_4$). The source of these contrary findings stems not from their proposed solution procedure, but from their randomly generated input data. As the following discussion will indicate, only correlation matrix $R_4$ was valid—the others were all infeasible.

An elementary, yet very important, measure of the relationship among random variables is the Pearson correlation coefficient. Frequently in numerical and simulation studies for example, a correlation structure for input variables is specified. Besides the well-known fact that correlation coefficients can only take on values from $-1$ to $+1$, little else about the structure of a correlation matrix may be evident to a good number of non-statisticians like us. However, researchers need to be aware that the correlation structure cannot simply be arbitrarily determined, but instead certain relationships among correlation coefficients must be taken into account based on matrix algebra. As Lurie and Goldberg [15] succinctly put it, “a correlation matrix is any symmetric, positive semi-definite matrix having unit diagonal elements [p.205]” (see also Marsaglia and Olkin [16]). (For the interested reader, one version of the proof of positive semi-definiteness is presented in Appendix A.)

In other words, there are limits to the values of some correlation coefficients when other correlation coefficients have already been established. For instance, as alluded to earlier, if the correlation between random variables $X_1$ and $X_2$ is $-1$ and the correlation between random variables $X_1$ and $X_3$ is $-1$, the correlation coefficient between $X_2$ and $X_3$ cannot take any random value between $-1$ and $+1$. In this case, it is clear that perfect, positive correlation exists between $X_2$ and $X_3$ since both variables are perfectly, negatively correlated with $X_1$. But what about a less obvious case such as a correlation between $X_1$ and $X_2$ of, say, $-0.9$ and a correlation between $X_1$ and $X_3$ of, say, $-0.9$?
In the case of three random variables, the following rule as originally developed by Hubert [17] is offered to assist researchers in determining both the consistency of a correlation matrix with three variables as well as the bounds of the coefficients. Without loss of generality and for clarity and brevity, all variables discussed in the proof are assumed to be standardized since the correlation matrix of non-standardized variables is the same as that of their standardized counterparts.

4.1. The correlation structure for three random variables

Given the correlation coefficients between $X_1$ and $X_2$ and between $X_1$ and $X_3$, the correlation coefficient between $X_2$ and $X_3$ is bounded as follows:

$$r_{12}r_{13} - \sqrt{(1 - r_{12}^2)(1 - r_{13}^2)} \leq r_{23} \leq r_{12}r_{13} + \sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}.$$  \hfill (2)

**Proof.** Suppose that $X_1$, $X_2$, and $X_3$ are all standardized variables with $E(X_1)=E(X_2)=E(X_3)=0$ and $\text{Var}(X_1)=\text{Var}(X_2)=\text{Var}(X_3)=1$, then

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \text{Cov}(X_1, X_2) = r_{12} \quad \text{(where } |r_{12}| \leq 1\text{)},$$

$$\text{Corr}(X_1, X_3) = \frac{\text{Cov}(X_1, X_3)}{\sqrt{\text{Var}(X_1)\text{Var}(X_3)}} = \text{Cov}(X_1, X_3) = r_{13} \quad \text{(where } |r_{13}| \leq 1\text{)},$$

$$\text{Corr}(X_2, X_3) = \frac{\text{Cov}(X_2, X_3)}{\sqrt{\text{Var}(X_2)\text{Var}(X_3)}} = \text{Cov}(X_2, X_3) = r_{23} \quad \text{(where } |r_{23}| \leq 1\text{)}.$$

First assume that $|r_{12}| \neq 1$. Now consider the covariance and correlation between $X_1$ and $uX_2 + X_3$, an arbitrary linear combination of $X_2$ and $X_3$ with any given value of $u$:

$$\text{Cov}(X_1, uX_2 + X_3) = \text{Cov}(X_1, uX_2) + \text{Cov}(X_1, X_3) = ur_{12} + r_{13} \quad \text{and}$$

$$\text{Var}(uX_2 + X_3) = \text{Var}(uX_2) + \text{Var}(X_3) + 2\text{Cov}(uX_2, X_3) = 1 + u^2 + 2ur_{23}.$$

Since

$$\text{Corr}(X_1, uX_2 + X_3) = \frac{\text{Cov}(X_1, uX_2 + X_3)}{\sqrt{\text{Var}(X_1)\text{Var}(uX_2 + X_3)}}$$

$$= \frac{ur_{12} + r_{13}}{\sqrt{1 + u^2 + 2ur_{23}}}$$

and

$$|\text{Corr}(X_1, uX_2 + X_3)| \leq 1,$$

then

$$(ur_{12} + r_{13})^2 \leq 1 \ast (1 + u^2 + 2ur_{23}),$$

which can be restated as

$$u^2(1 - r_{12}^2) + 2u(r_{23} - r_{12}r_{13}) + (1 - r_{13}^2) \geq 0.$$  \hfill (3)
Rearranging the terms in (3) leads to

\[
(1 - r_{12}^2) \left[ u + \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2} \right]^2 - \frac{(r_{23} - r_{12}r_{13})^2}{1 - r_{12}^2} + (1 - r_{13}^2) \geq 0.
\]

Since \( u \) can be any real value, setting \( u = -(r_{23} - r_{12}r_{13})/(1 - r_{12}^2) \) results in

\[
-\frac{(r_{23} - r_{12}r_{13})^2}{1 - r_{12}^2} + (1 - r_{13}^2) \geq 0
\]
or

\[
(r_{23} - r_{12}r_{13})^2 \leq (1 - r_{12}^2)(1 - r_{13}^2),
\]

which is equivalent to (2).

Now assume that \( |r_{12}| = 1 \) and \( |r_{13}| \neq 1 \). An analysis of the correlation between \( X_1 \) and \( X_2 + uX_3 \) leads to a condition similar to (3):

\[
u^2(1 - r_{13}^2) + 2u(r_{23} - r_{12}r_{13}) + (1 - r_{12}^2) \geq 0.
\] (4)

Following the same logic as above, this can also be transformed into \( (r_{23} - r_{12}r_{13})^2 \leq (1 - r_{12}^2)(1 - r_{13}^2) \), and is again equivalent to (2).

Finally, assume that \( |r_{12}| = 1 \) and \( |r_{13}| = 1 \). Selecting \( u > 0 \) in (3) and \( u < 0 \) in (4) clearly forces \( r_{23} - r_{12}r_{13} = 0 \), which also satisfies (2).

Therefore, condition (2) is satisfied regardless of the values of \( r_{12} \) and \( r_{13} \).  

The implications of this rule are clear. For any three variables \( X_1, X_2, \) and \( X_3 \) (whether standardized or not), once the correlation coefficients between any two pairs are given, the feasible range of values for the third coefficient is then constrained by the condition expressed in (2). The following numerical examples illustrate this principle.

**Example 1:** Suppose \( r_{12} = 1 \). Then condition (2) requires that \( r_{23} = r_{13} \).

**Example 2:** Suppose \( r_{12} = -1 \). Then condition (2) requires that \( r_{23} = -r_{13} \).

**Example 3:** Suppose \( r_{12} = -0.9 \) and \( r_{13} = -0.9 \). Then condition (2) results in \((-0.9)^2 - (1 - (-0.9)^2) \leq r_{23} \leq (-0.9)^2 + (1 - (-0.9)^2) \) or \( 0.62 \leq r_{23} \leq 1 \).

**Example 4:** Suppose \( r_{12} = -0.8 \) and \( r_{13} = -0.9 \). Then condition (2) results in \((-0.8) * (-0.9) - [(1 - (-0.9)^2)(1 - (-0.8)^2)]^{1/2} \leq r_{23} \leq (-0.8) * (-0.9) + [(1 - (-0.9)^2)(1 - (-0.8)^2)]^{1/2} \) or \( 0.4585 \leq r_{23} \leq 0.9815 \).

**Example 5:** Consider the following correlation coefficients from matrix \( R_1 \) of Tyagi and Das [1, p. 1055]: \( r_{12} = 0.1, r_{13} = -1, \) and \( r_{23} = 0.4 \). As shown in Example 2 above, this structure clearly violates condition (2). Consequently, matrix \( R_1 \) is infeasible.

**Example 6:** Consider the following correlation coefficients from matrix \( R_2 \) of Tyagi and Das [1, p. 1055]: \( r_{36} = 0.8, r_{37} = 0.9, \) and \( r_{67} = -0.4 \). This structure also violates condition (2) as \( 0.4585 \leq r_{67} \leq 0.9815 \), indicating that matrix \( R_2 \) is also infeasible.
Example 7: Consider the following correlation matrix excerpted from matrix $R$3 of Tyagi and Das [1, p. 1055]:

$$\begin{pmatrix} X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} = \begin{bmatrix} 1 & -0.5 & 0.5 & -0.5 \\ 1 & -0.5 & 0.5 & -0.5 \\ 1 & 0.5 \end{bmatrix}.$$ 

Any combination of three variables satisfies condition (2). For instance, consider $X_4$, $X_5$, and $X_6$ with $r_{45} = -0.5$, $r_{46} = 0.5$, and $r_{56} = 0.5$. Condition (2) yields: $(-0.5) \cdot (0.5) - [(1 - (-0.5)^2)(1 - (0.5)^2)]^{1/2} \leq r_{56} \leq (-0.5) \cdot (0.5) + [(1 - (-0.5)^2)(1 - (0.5)^2)]^{1/2}$ or $-0.5 \leq r_{56} \leq 0.5$. In other words, the correlations do not violate the rule for three variables.

4.2. The presence of more than three random variables

In the last example more than three variables were present. It is likely to be the case that additional variables pose additional constraints. Thus, a more general test of the feasibility of a correlation matrix is needed to account for higher-level relationships. As mentioned earlier, a correlation matrix must be positive semi-definite. Unfortunately, the calculations necessary for determining the feasibility of a correlation matrix with more than three variables can be quite cumbersome. However, computer code can be written that assists in determining a correlation matrix’s feasibility.

The authors have developed a straightforward C code that calculates the eigenvalues of the correlation matrix. If any of these values are less than zero, the correlation matrix is infeasible. Using the program code (available upon request), more complicated correlation structures can be readily examined for their validity, as the following examples show.

Example 8: Consider the following obviously flawed correlation matrix:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 \end{bmatrix}.$$ 

The matrix eigenvalues are $\{2, 2, 2, -2\}$, indicating that it is not positive semi-definite. Since all eigenvalues are not greater than or equal to zero, the correlation matrix is invalid.

Example 9 (Example 7 revisited): Now, reconsider the correlation matrix excerpted from matrix $R$3 of Tyagi and Das [1, p. 1055]:

$$\begin{pmatrix} X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} = \begin{bmatrix} 1 & -0.5 & 0.5 & -0.5 \\ 1 & -0.5 & 0.5 \\ 1 & 0.5 \end{bmatrix}.$$
With eigenvalues of \{2.118, 1.5, 0.5, -0.118\}, the matrix is not positive semi-definite. As a result, even though the correlation coefficients for three variables lie within the bounds established by condition (2), it can be concluded that the correlation matrix for the four variables examined is not feasible and, therefore, that the matrix \(R_3\) as a whole is invalid.

Example 10: Lastly, consider correlation matrix \(R_4\) of Tyagi and Das [1, p. 1055]:

\[
\text{Corr} = \begin{pmatrix}
X_1 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_2 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_3 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_4 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_5 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_6 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_7 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
X_8 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
\end{pmatrix}
\]

This matrix’s eigenvalues are \{4.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5\}, indicating that it is positive semi-definite. Accordingly, this correlation matrix is valid.

Note that, while the consistency of a correlation matrix is checked, the determination of boundaries necessary for constructing feasible matrices is not performed here. Inclusion of this latter issue can be considerably more involved; for example, Rousseeuw and Molenberghs [18] have identified a number of approaches (including shrinking, eigenvalue, and scaling methods) for building mathematically consistent correlation matrices. Consequently, a code for building feasible matrices is left for future work.

Returning to the question of why partial aggregation was preferred in some cases, the above analysis indicates that matrices \(R_1\), \(R_2\), and \(R_3\) are infeasible (for the full matrices, the eigenvalues are \{3.247, 2.883, 1.671, 0.805, 0.692, 0.258, -0.544, -1.012\} for \(R_1\); \{4.298, 1.595, 1.280, 0.857, 0.577, 0.217, -0.254, -0.570\} for \(R_2\); and \{3.784, 2.285, 1.182, 0.5, 0.5, 0.5, -0.251, -0.5\} for \(R_3\)). As a consequence, these three matrices produce invalid results—results corresponding to a preference for partial aggregation. On the other hand, matrix \(R_4\) is a feasible correlation structure and produces valid results—results that, not surprisingly, correspond to a preference for complete aggregation.

### 5. Conclusion

On the topic of pooling, it was shown that complete aggregation never performs worse than partial aggregation on the sole basis of demand correlation. There may, in fact, be cases where partial aggregation of inventory is preferred, but that entails the consideration of other factors not accounted for here, such as lead times or transportation costs. For example, inventory centralization at the wrong location could necessitate higher safety stocks to cover possible increases in average lead times and lead time variability (see Evers and Beier [19]). Similarly, as the number of stock-keeping facilities is reduced, the cost of transportation must increase in order to maintain the same level of service in terms of response times and may, as a result, outweigh the savings from reductions in
inventory levels (see Das and Tyagi [20]). Developing procedures that combine demand correlation considerations with these additional concerns would represent a useful extension of the single echelon stream of research and the solution techniques of Tyagi and Das [1].

Moreover, since the relationship between correlation coefficients is not apparent to most, the guidance provided by this paper, though not necessarily novel, should be of great benefit to researchers in making sure that infeasible correlation structures are avoided. For instance, if a researcher constructs an exemplar correlation matrix for simulation or numerical purposes, the methods described in this paper can be employed to validate its structure. By doing so, the researcher can be confident that any results obtained have not arisen from faulty inputs related to the specified correlation coefficients. These methods can also be useful in checking the correctness of correlation matrices estimated from empirical data and ensuring that calculation errors were not made.

Acknowledgements

Sections of this paper reviewing the risk-pooling literature were presented at the Workshop on Supply Chain Management Practice and Research: Status and Future Directions, in Rockville, MD, April 2001.

Appendix A. The structure of a valid correlation matrix

Let \( \mathbf{X}_{0:p} \) be a \((p + 1) \times 1\) standardized random vector in which all variables have means of 0 and standard deviations of 1, let \( \Sigma^* \) be its \((p + 1) \times (p + 1)\) variance–covariance matrix which is the same as its correlation matrix, and let \( \mathbf{u} \) be a \(p \times 1\) vector of arbitrary real numbers, with the following partition:

\[
\mathbf{X}_{0:p} = \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_p \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}, \quad \text{var}(\mathbf{X}_{0:p}) = \Sigma^* = \begin{pmatrix} s_{00} & s_{01}^t \\ s_{01} & \Sigma \end{pmatrix}.
\]

Note that \( X_0, u_i, \) and \( s_{00} = 1 \) are scalars, while \( s_{01} \) is a \(p \times 1\) vector.

Given the correlation structure \( \Sigma \), then the correlation coefficient vector \( s_{01} \) will be bounded by the following condition:

\[
\Sigma - s_{01} s_{01}^t \text{ must be positive semi-definite,}
\]

where \( \Sigma \) is the correlation structure among variables \( X_1, \ldots, X_p \), and \( s_{01} \) is the correlation coefficient vector between \( X_0 \) and \( X_{1:p} \).

**Proof.** Suppose that \( X_0, X_1, \ldots, X_p \) are standardized variables with mean of 0 and standard deviation of 1. Now consider the covariance and correlation between \( X_0 \) and \( \mathbf{u}' \mathbf{X}_{1:p} \), where \( \mathbf{u} \) is a \(p \times 1\) vector of arbitrary real numbers:

\[
\text{Cov}(X_0, \mathbf{u}' \mathbf{X}_{1:p}) = \mathbf{u}' \text{Cov}(X_0, \mathbf{X}_{1:p}) = \mathbf{u}' s_{01} \quad \text{and} \quad \text{Var}(\mathbf{u}' \mathbf{X}_{1:p}) = \mathbf{u}' \text{Var}(\mathbf{X}_{1:p}) \mathbf{u} = \mathbf{u}' \Sigma \mathbf{u}.
\]
Since $\Sigma$ and $\Sigma^*$ must be positive semi-definite and
\[
|\text{Corr}(X_0, u'X_{1-p})| = \left| \frac{\text{Cov}(X_0, u'X_{1-p})}{\sqrt{\text{Var}(X_0)\text{Var}(u'X_{1-p})}} \right| \leq 1,
\]
then
\[
\text{Var}(u'X_{1-p}) \geq [\text{Cov}(X_0, u'X_{1-p})]^2 = u's_{01}s_{01}'u.
\]
That is,
\[
u'(\Sigma - s_{01}s_{01}')u \geq 0.
\]
In other words, no matter what values $u$ will take, the left-hand side will be non-negative. (For a non-zero vector $u$, the equality only arises if $X_0$ can be represented by a linear combination of $X_1, \ldots, X_p$.) This means that $\Sigma - s_{01}s_{01}'$ must be a positive semi-definite $p \times p$ matrix (cf., Graybill [21], Johnson and Wichern [22]).

References


Kefeng Xu is an Assistant Professor of Operations Management at the University of Texas at San Antonio. He received his Ph.D. in operations management and logistics/transportation from the University of Maryland. His publications have appeared in journals such as Transportation Research, Journal of Business Logistics, Journal of Transportation Economics and Policy, and International Journal of Physical Distribution and Logistics Management. His research interests are in inventory management; logistics, service and manufacturing strategy; supply chain modeling; transportation policy, service, cost and demand studies.

Philip T. Evers is an Associate Professor of Logistics Management at the University of Maryland. He received his Ph.D. in logistics management from the University of Minnesota. His research interests include inventory management, transportation operations management, and intermodal transportation issues. He has published numerous articles in journals such as Transportation Research, Journal of Business Logistics, Transportation Journal, and Production and Inventory Management Journal.