Modeling the Effects of Demand Correlation in Location Problems that account for Inventory Pooling

Michael J. Bucci*, Donald P. Warsing†, Michael G. Kay*, Reha Uzsoy*, James R. Wilson†
*Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA
†Department of Business Management, North Carolina State University, Raleigh, NC 27695, USA

Abstract
We investigate large-scale facility location models that consider the risk pooling effects of centralizing safety stocks when customer demands are correlated across customer locations. This problem has not been well studied, perhaps because the calculation of safety stock levels with correlated demands requires significant computational effort. The experimental evaluation of any proposed procedure is also difficult, since industrial data is generally unavailable and it is difficult to create a statistically valid correlation matrix of customer demands required to generate random test instances of realistic size. We show that even very low levels of correlation have significant impact on solution cost and structure. This is of practical significance since distribution networks increasingly rely on contracts of relatively short duration, implying low or no fixed facility costs, such that the cost of holding safety stock can represent a large portion of the controllable distribution network costs. We present a method for generating a statistically valid correlation matrix when the correlations between customer demands are inversely proportional to the distance between customer locations. Using a metaheuristic solution procedure, we study a set of test instances that incorporate both safety stock holding costs and outbound transportation costs, illustrating the importance of explicitly considering demand correlation in location models.

Keywords
Facility location, inventory pooling, safety stock, demand correlation

1. Introduction
We investigate large-scale facility location models that consider the risk pooling effects of centralizing safety stocks in the presence of customer demands that may be correlated across customer locations. This problem is challenging to study for two reasons. First, it is surprisingly difficult to create a statistically valid (i.e., positive semi-definite) correlation matrix for the customer demands in all but trivially sized networks. This renders the generation of meaningful random test instances difficult, but such instances are necessary for research progress since industrial data sets with correlation between demand sites are extremely hard to obtain. Second, the computation of the covariance-related term in the safety stock calculation is computationally intensive even for problems with only a few customer locations. These factors make it difficult to evaluate the possible safety stock costs a priori, and limit the number of safety stock cost calculations that can be performed if the problem is to be solved in reasonable time. The thesis
of this paper is that addressing these obstacles is important, as even low levels of correlation can significantly impact the solution cost and structure.

The model studied represents a single-source, single-tier distribution network located in the continental United States that may experience correlated demand between customer locations. The objective of the problem is to locate an unknown number of distribution centers to minimize the sum of transportation and safety stock holding costs. As distribution networks have increasingly come to rely on contracts of relatively short duration in which warehouse space is rented from third parties [Armstrong and Associates 2009], the customary fixed costs due to the size of a warehouse are not incurred or are significantly reduced. In these environments, where the cost of safety stock can represent a large portion of the distribution network costs, the incorporation of risk pooling effects may significantly influence the structure of the network-design solution. Moreover, since customer demands are not likely to be independent in any realistic distribution network, it is of interest to understand the impact of demand correlation across customer locations on the safety stock costs, and ultimately on the optimal facility location decisions.

Our analysis begins with the study of problems that assume a constant correlation coefficient for all customer pairs. Although this scenario is unlikely in practice, it can nevertheless provide valuable insights. We first solve these problems assuming independent customer demands; then, while keeping this solution fixed, we recalculate the inventory cost that would be incurred under a common correlation coefficient for all customer pairs. Figure 1 shows the results obtained on a typical 400-customer instance studied in this paper. The solution has four facilities of roughly the same size, and the figure clearly depicts the significant increase in the safety stock holding cost as a common inter-customer correlation coefficient is introduced.

Having established the importance of considering inter-customer demand correlation, we then consider the challenge of solving realistic instances of such problems. One of the difficulties in studying location models that reflect demand correlation is the lack of realistic test data sets that allow researchers to study the effects of different correlation levels and solution techniques. Industrial data is hard to obtain for confidentiality reasons, often leaving the generation of random test instances as the only alternative. However, a major difficulty in this latter approach is the generation of valid correlation matrices where the correlations between all pairs of demands can be controlled by the experimenter [Xu and Evers 2003]. This paper presents a method to generate valid correlation matrices for large datasets under the premise
that correlation of demand between customer demands are inversely proportional to the distance between customer locations (i.e. the correlation decreases as the distance between customer locations increases). We demonstrate the use of this technique to generate random test instances, and show that the safety stock costs and solution structure are significantly affected by the presence of even modest levels of demand correlation. The second challenge is solving a location model with this type of complexity. Solving location problems exactly is widely known to be a challenging combinatorial optimization problem [Daskin 1995]. Thus, we use a heuristic procedure, the discrete alternate location–allocation (ALA) neighborhood search heuristic, which has been shown to find near optimal solutions with low computational effort [Cooper 1963, Bucci et al. 2009]. As the comparison between solutions is assessed using a common solution (i.e., with no change in facility locations or customer allocations), these near optimal solutions are sufficient for our study.

The remainder of the paper is organized as follows. In Section 2 we review the literature related to this problem, and we present the general formulation of the problem and the metaheuristic solution procedure in Section 3. Section 4 presents the experimental design and analysis of results with a constant inter-customer correlation coefficient across all customer pairs. Section 5 demonstrates the use of the correlation matrix generation technique, and provides analysis of solutions with variable inter-customer correlation coefficient values. Section 6 offers conclusions and directions for future research.

Figure 1: Percentage increase in safety stock, as inter-customer demand correlation is introduced (all points represent a common solution to a 400-customer network served by four facilities)
2. Literature Review

We can group the related literature into three areas: studies of pooling effects in inventory systems, studies of location problems that include inventory pooling costs, and literature related to the generation of correlated demand data.

The impacts of inventory pooling have been widely studied. Maister [1976] studied safety stock costs and cycle inventory costs, presenting the well-known “Square Root Law” for safety stock inventory which states the system wide safety stock will be reduced by $\sqrt{n}$ when the inventory is consolidated from $n$ facilities to one facility. This rule requires demands to be independent and each location must have the same proportion of demand. Eppen [1979] also shows this result in his study of the multi-location newsboy problem. Erkip et al. [1990] present an expression for safety stock with correlated demands between locations and between successive time periods. This research assumes normally-distributed demand with equal coefficients of variation across customers, and focuses on finding an optimal ordering policy assuming a given allocation of inventory locations to customers. More recently, Das and Tyagi [1999] study correlated demand patterns between multiple products and offer guidelines on how to group products to minimize safety stock levels when both positive and negative correlations are present. The inventory aggregation work of Eppen [1979] is more recently reviewed by Chopra and Meindl [2004] who address the impact of aggregation on safety stock for situations with independent and correlated customer demand. Using the notation of Chopra and Meindl, the aggregated safety stock is given by

$$I_{SS} = \left( F^{-1}(CSL) \right) \sigma^F_L,$$

(1)

where $I_{SS}$ is the safety stock required for a given facility, $F^{-1}$ is the inverse cumulative distribution function (cdf) of demand allocated to this facility over the lead time to replenish inventory at this facility, $CSL$ is the cycle service level (i.e., the expected in-stock probability level across inventory cycles), and $\sigma^F_L$ is the standard deviation of demand during the replenishment lead time at the facility. Furthermore, Chopra and Meindl show that

$$\sigma^F_L = \sqrt{L \left( \sum_{i=1}^{k} \sigma^2_i + 2 \sum_{i<j} \rho_{ij} \sigma_i \sigma_j \right)},$$

(2)

where $\sigma^2_i$ is the variance of demand at location $i$, $\rho_{ij}$ is the correlation coefficient between demands at locations $i$ and $j$, and $L$ is the replenishment lead time.
where $L$ is the supplier lead time, $\sigma_i$ and $\sigma_j$ are the standard deviations of demand at customers $i$ and $j$, respectively, and $\rho_{ij}$ is the correlation coefficient of the demands of customers $i$ and $j$.

Traditional location problem formulations such as the multi-source Weber problem [Brimberg et al. 2000], $P$-median problem [Daskin 1995], and the uncapacitated fixed charge facility location problem [Mirchandani and Francis 1990] are widely studied models whose objective is to locate facilities to serve a known set of demands at given customer locations to minimize transportation cost (if the number of facilities is given) or the sum of transportation costs and fixed facility costs (if the number of facilities is not specified). Both exact methods and heuristics are used to solve these NP-hard problems (Daskin [1995], Hansen and Mladenovic [1997]). More recent research has extended these formulations to include inventory costs in the objective function. Nozick and Turnquist [1998] describe a method to approximate safety stock costs in a fixed-charge location model. Their work assumes independent customer demands and that the inventory costs follow the “Square Root Law” for safety stock inventory. They then approximate these costs using fixed and linear cost terms and solve the problem with a constructive ADD procedure [see Daskin 1995] followed by an improvement procedure. Daskin et al. [2002] and Shen et al. [2003] introduce joint location-inventory formulations that extend the uncapacitated fixed charge model to explicitly include the impact of demand variability with independent demands on inventory pooling costs. They both use Lagrangian-relaxation based exact algorithms to solve the problem, utilizing weighting factors to examine several levels of location, transportation, and inventory costs. Snyder et al. [2007] extend this work with a formulation that minimizes the sum of fixed facility, transportation, and inventory costs assuming a constant variance to mean ratio and independent customer demands.

This line of research was extended further by Vidyarthi [2007], who considers a more complex multiproduct, two-echelon facility location model that includes fixed facility, transportation, and inventory costs. Assuming independent demand and a piecewise linear approximation of the inventory costs, he uses a Lagrangian relaxation to obtain lower bounds and a heuristic to find feasible solutions. Assuming independent demands, Ozsen et al. [2008] study a capacitated location model with risk pooling (CLMRP) that is solved using a Lagrangian algorithm. Sourirajan et al. [2008] analyze a two-stage supply chain network that includes explicit calculation of safety stock costs and congestion costs at the distribution center. Uster et al. [2008] study a three-tier supply chain, considering both inventory and transportation costs, to locate a single warehouse and determine a distribution strategy from the supplier.
to the retailers. Their integrated solution heuristic compares favorably with the traditional sequential models that locate facilities and then determine an inventory strategy. Using a general concave cost to reflect the scale economies of combined fixed and variable network costs (e.g., fixed facility costs, variable transportation costs, and variable production and/or inventory costs), Bucci et al. [2009] develop and evaluate several metaheuristics that combine algorithmic construction, allocation, and location techniques, including alternate location–allocation (ALA), variable neighborhood search (VNS), and Tabu search, to solve large-scale problems. They also determine combinations of heuristic techniques that can provide near optimal solutions with moderate computational effort.

Finally, a few authors have proposed approaches for generating correlated demand data sets. The principal difficulty is in obtaining a correlation matrix that reflects correlation between many pairs of demands, yet is still statistically valid (i.e., positive semi-definite). Xu and Evers [2003] review inventory pooling effects with correlated demands and specifically address the challenge and importance of creating statistically valid correlation matrices, which they demonstrate to be a non-trivial task even for small problems. They offer methods to generate small correlation matrices (up to a 3-by-3 matrix) and procedures to assess the validity of a given correlation matrix. Using a different approach, Corbett and Rajaram [2006] describe a copula-based method that can be used to generate correlation values between a single pair of normally distributed demands as well as pairs of demands that follow any one of several non-normal distributions. Dorey and Joubert [2005] also review the difficulty of generating statistically valid correlation matrices and discuss the use of the copula method to generate correlation values between a pair of values. What is lacking in these papers, however, is a method to generate correlation values between more than three customer demands. This is the issue we address in this paper.

Our research integrates ideas from these three streams of literature to study the impact of inventory pooling on solutions to facility location problems with uncorrelated or correlated demands. We present the model formulation in the next section.

3. Model Formulation and Solution Methodology

Given \( n \) uncaptacitated distribution facilities that serve \( m \) customer locations, our formulation of the problem is as follows:
$$\min z = \sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} X_{ij} + K \sum_{i=1}^{n} \left( \sum_{j=1}^{m} X_{ij} \sigma_{j}^{2} + 2 \sum_{j<y} X_{ij} X_{iy} \rho_{jy} \sigma_{j} \sigma_{y} \right)$$  \(3\)

subject to

$$\sum_{i=1}^{n} X_{ij} = 1 \text{ for } j = 1, 2, \ldots, m$$  \(4\)

$$X_{ij} \in \{0, 1\} \text{ for } i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m$$  \(5\)

$$X_{iy} \in \{0, 1\} \text{ for } i = 1, 2, \ldots, n; \ y = 1, 2, \ldots, m,$$  \(6\)

In (3), $C_{ij}$ reflects the total transportation cost to move goods from facility $i$ to cover the annual demand at customer $j$. This transportation cost is modeled as a one-way full truckload shipment cost and is calculated as the product of the freight transport charge ($$/mi$), the estimated road distance (mi) between the customer and facility locations, and the total annual demand (truckloads) at the customer. $K$ is a user-specified constant that combines the annual cost of holding a unit in inventory at any facility in the network, the constant and common replenishment lead time across all facilities, and the common CSL across all facilities. The decision variables are the binary integer variables $X_{ij}$, which take a value of 1 if customer demand at $j$ is supplied by facility $i$ and 0 otherwise. Thus, the first term in the objective function computes the total annual cost of transporting goods from the facilities to the customers. The second term computes the total safety stock costs for all facilities using the formulation of Chopra and Meindl [2004] as given in expressions (1) and (2). Constraints (4) require that each customer be served by only one facility. Constraints (5) and (6) are integer constraints.

### 3.1 Metaheuristic Solution Approach

It is important to state at the outset that the objective of this paper is not to provide a specific solution algorithm for location problems with correlated demand, but to make the case that explicit consideration of correlations is important since even quite low levels of correlation can substantially alter the structure of the desired solution. Given the computational burden of exact procedures, especially for location problems incorporating the explicit calculation of safety stock costs, we opt for the “H10” heuristic described in Bucci et al. [2009]. This heuristic has been shown to consistently produce solutions within 1% of the optimal total cost in modest CPU times (approximately 1300 seconds to solve a facility location problem with 400 customers and 400 potential facility locations). As our primary motivation is the
impact of correlated demands on the solution cost and structure, obtaining an exact optimal solution is not
critical to our analysis. The H10 heuristic combines a constructive ADD procedure with a discrete ALA
improvement procedure as shown in Figure 2 [Daskin 1995, Whitaker 1985]. The problems were
generated and the heuristic solution procedures were coded and executed in MATLAB 7.2, utilizing the
Matlog toolbox developed by Kay [2006]. All tests were run on a 3.2 GHz Intel Pentium 4 computer with
1527 MB of memory running Windows XP.

Figure 2: General Solution Procedure for “H10” Heuristic (Bucci et al. 2009)

4. Computational Study: Common Correlation Coefficient for all Customers Pairs

In our computational experiments we model a single-source, single-tier distribution network with a single
product—or “aggregated product,” i.e., one that reflects average product costs and characteristics and
aggregates customer demand across the product portfolio for each customer location—that may have
inter-customer correlated demands. The customer locations and the potential facility locations are the
five-digit ZIP codes in the continental U.S. that have non-zero population and are located between −80
and −90 degrees longitude (9,744 in total). To obtain the customer locations in any given problem
instance we select customers randomly without replacement using population-weighted selection
probabilities. We also assign customer demand weights randomly using population-weighted
probabilities, where a demand weight specifies a customer’s portion of the total demand. The intent here
is to generate realistic clusters of customer locations and demand weights for the continental United States—i.e., to generate insights from our analysis that are relevant to real-world distribution problems, in which demands are likely to cluster around population centers that are dispersed, however randomly or “non-randomly,” throughout a given geographic region. In generating the cost data, we assume the resulting supply chains should mirror existing U.S. retail networks, which typically contain between five and twenty distribution facilities to serve most of the continental U.S. For example, Lowe’s [2006] and Walgreens [2006] respectively use 11 and 13 regional distribution centers to serve the continental U.S. The distance between facilities and customer locations is computed using great circle distances in statute miles, magnified by a circuity factor of 1.2 to reflect the road network effects that inflate “crow-flies” distances [Ballou et al. 2002]. Regardless of the number of customers in the network, the customer weights are normalized to maintain a total demand of 10,000 units per year. An inventory cost weighting factor $\beta$, which adjusts the inventory cost relative to the transportation cost, was set to obtain solutions with the desired number of facilities (between 5 and 20). Although we lack detailed empirical data, discussion between the authors and industry consultants suggests that correlation values up to 0.2 are reasonable for real customer demand data. Thus, the range of correlation coefficients values in our study ranges between $0 \leq \rho_{ij} \leq 0.3$.

We let $N = \{1, 2, \ldots, 9744\}$ denote the (arbitrarily) ordered set of all 9,744 customer locations. For each problem instance that is formulated and solved, we define $N_n \subseteq N$ ($|N_n| = n$) to be the set of customer locations for the problem, with $n \in \{100, 200, 400\}$. Further, we define $M \subseteq N_n$ to be the set of $m$ candidate facility locations (i.e., $|M| = m$) for a particular problem instance.

In the following two sections we assess the impact of a constant level of inter-customer correlation by solving the location problem assuming independent customer demands, and then, while keeping the solution fixed, we introduce a common correlation coefficient for all customer pairs and recalculate the inventory cost. For these problems, we separately assess the impact on the results of an increasing number of customers in the network ($n = 100$, 200, or 400) and an increasingly decentralized network (i.e., increasing values of $m$, such that $m = 1, 4, 6, \text{or} 8$). For all of these problems, we report aggregate statistics for thirty instances at each experimental setting.

4.1 Effect of the Number of Customers in the Network

To assess the effect the number of customers in the network has on safety stock costs in the presence of
inter-customer correlation we compare solutions with $n = 100, 200, 400$ while keeping $m = 4$. Table 1 shows the facility sizes for each of these solutions as a percentage of the total demand and as a percentage of the total number of customers. We can see in Table 1 that all solutions have facilities of roughly equal size (i.e., around 25% of the total demand) and the percentage of customers allocated to each facility is also roughly the same. Figure 3 shows the percentage increase in inventory, relative to the independent demand solution for each value of $n$, as inter-customer correlation is introduced. Figure 4 shows this same data, but it is presented as the increase in safety stock required— as a percentage of the expected weekly demand. We observe from the 100-customer problem that even low levels of correlation have a significant impact on the safety stock required (e.g., a 16.2% increase for $\rho = 0.02$). As the number of customers in the problem increases, the impact of correlation grows more dramatically due to the compounding effect that allocating more customers to a facility has on the covariance term in the safety stock calculation. Figure 4 shows that the inventory cost for the independent demand case is nearly identical regardless of the number of customers in the problem (recall that the total demand in the network is held constant regardless of the number of customers), showing that the solutions for these problems are not skewed or biased as the number of customers in the network increases.

Table 1: Characteristics of solutions with 100, 200, and 400 customers

<table>
<thead>
<tr>
<th># of customers:</th>
<th>Percentage of demand allocated to each facility</th>
<th>Percentage of customers allocated to each facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility 1</td>
<td>32.2% 32.4% 36.3%</td>
<td>33.1% 33.0% 34.2%</td>
</tr>
<tr>
<td>Facility 2</td>
<td>27.1% 27.4% 27.1%</td>
<td>27.7% 27.6% 28.5%</td>
</tr>
<tr>
<td>Facility 3</td>
<td>23.1% 23.1% 21.3%</td>
<td>23.3% 23.7% 22.4%</td>
</tr>
<tr>
<td>Facility 4</td>
<td>17.6% 17.1% 15.4%</td>
<td>15.9% 15.8% 14.9%</td>
</tr>
</tbody>
</table>
Figure 3: Percentage increase in safety stock, as inter-customer demand correlation is introduced

Figure 4: Safety stock required, as a percentage of expected weekly demand, as inter-customer demand correlation is introduced
4.2 Effect of the Number of Facilities in the Solution

To assess the effect of the number of facilities in the network on safety stock cost in the presence of inter-customer correlation we compare solutions with \( m = 1, 4, 6, \) or 8 while keeping the number of customers \((n = 400)\) and the total customer demand constant. Table 2 shows the facility sizes for each of these solutions as a percentage of the total demand and as a percentage of the total number of customers. As the number of facilities in the solution increases, a few facilities maintain a fairly large portion of the demand while others are fairly small in size. This likely stems from several factors, such as the natural clustering of customer locations and customer demands present in the dataset, and perhaps from the exclusion of pooling effects beyond safety stock (e.g., cycle stock) and from our assumption of linear outbound transportation costs (i.e., not reflecting economies of scale, which would put small facilities at a transportation cost disadvantage). Figure 5 shows the percentage increase in safety stock, relative to the independent demand solution for each value of \( m \), as inter-customer correlation is introduced. Figure 6 shows this same data, but it is presented as the increase in safety stock required—as a percentage of the expected weekly demand. For the independent demand case \((\rho = 0)\), we observe from Figure 6 that the inventory required for the four facility solution compared to the single-facility solution (2.0 times larger) equals the 2.0 times increase predicted by the “Square Root Law” [Maister 1976]. This also holds reasonably well for the change from one facility to eight facilities. (The “Square Root Law” predicts a multiple of \( \sqrt{8} = 2.82 \) while our model shows an inventory level 2.61 times larger.) Comparing this to the four-facility, independent demand case to the four-facility, correlated demand case with \( \rho = 0.1 \), we see that the inclusion of correlation results in 2.8 times more safety stock. As this change is comparable to the impact of changing the number of facilities in the solution from one to four, or one to eight, it is important to consider correlation in formulating and solving these problems.

Table 2: Characteristics of solutions with 4, 6, and 8 facilities

<table>
<thead>
<tr>
<th># of facilities in solution</th>
<th># of facilities in solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% of total demand allocated to each facility</th>
<th>% of customers allocated to each facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.8%</td>
<td>33.3%</td>
</tr>
<tr>
<td>26.9%</td>
<td>27.6%</td>
</tr>
<tr>
<td>22.7%</td>
<td>23.2%</td>
</tr>
<tr>
<td>17.7%</td>
<td>15.9%</td>
</tr>
<tr>
<td>10.5%</td>
<td>11.7%</td>
</tr>
<tr>
<td>6.0%</td>
<td>6.6%</td>
</tr>
<tr>
<td>2.5%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% of total demand allocated to each facility</th>
<th>% of customers allocated to each facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.8%</td>
<td>25.0%</td>
</tr>
<tr>
<td>21.5%</td>
<td>22.7%</td>
</tr>
<tr>
<td>18.9%</td>
<td>20.0%</td>
</tr>
<tr>
<td>14.4%</td>
<td>14.0%</td>
</tr>
<tr>
<td>10.1%</td>
<td>11.7%</td>
</tr>
<tr>
<td>7.8%</td>
<td>6.6%</td>
</tr>
<tr>
<td>6.2%</td>
<td>6.4%</td>
</tr>
<tr>
<td>2.5%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>
Figure 5: Percentage increase in safety stock, as inter-customer demand correlation is introduced

Figure 6: Safety stock required, as a percentage of expected weekly demand, as inter-customer demand correlation is introduced
The next section presents the methodology developed to generate valid correlation matrices for correlation coefficient values that vary based on the distance between customer locations. This procedure is then used to generate random test instances that we can study using the metaheuristic solution procedure.

5. Procedure for Generating Valid Correlation Matrices

Lacking actual demand data and guidance from the literature on data generation techniques for large correlation matrices that are non-random, we developed a methodology to determine correlation values based on the assumption that correlation of demand between customer demands are inversely proportional to the distance between customer locations (i.e. the correlation decreases as the distance between customer locations increases). For example, two customers located in Pennsylvania are likely to have more highly correlated demands for a given item than a customer in Pennsylvania and a customer in Arizona due to their common geography (e.g., common weather) and the fact that buying behaviors (e.g., tastes and preferences) are likely to be more consistent within a given geographic region. The methodology provides a valid correlation matrix by partitioning a large number of customers into a smaller number of groups, and establishes the relationships between the inter-customer demand correlations of the customer groups based on the structural properties of the matrix. In Appendix A, we establish the relationship between the inter-group correlation coefficients for the scenario in which there are three customer groups.

5. 1 Application of Correlation Matrix Generation Technique

We now use test instances generated using the correlation matrix generation technique to show the impact of inter-customer demand correlation on safety stock costs and solution structure. We geographically partition the solution space into three customer groups (i.e. geographic regions) based on their location. These three customer groups are then used to determine the correlation values for each pair of customer sites. For the analysis, the three groups were established by arbitrarily choosing two latitude values to partition the overall region, as shown in Figure 7. The correlation coefficients are then determined using the following steps. If all pairs of customer locations \((i,j)\) are in the same group (i.e. AA, BB, or CC) they have a common correlation value \(\rho_0\). If the customer locations \((i,j)\) are located in adjacent groups (i.e. AB or BC) they are assigned a common correlation value \(\rho_1\). If the customer locations \((i,j)\) are in distant groups (i.e., AC) they have a common correlation value \(\rho_2\). Based on the assumption that as distance
between a customer pair increases the level of correlation between the customer pair decreases, we assume $\rho_0 \geq \rho_1 \geq \rho_2$, consistent with the properties proved in Proposition 1 and Remark 1 in Appendix A.

![Figure 7: Partitioning of the Solution Space into three Geographic Groups](image)

Each problem is solved from the outset with the specified correlation values and compared to the solution that assumes independent customer demands (see Table 3). For all of these problems, thirty instances at each set of correlation coefficient settings are run with $n = 100$, and aggregate statistics for the thirty runs are reported. From Table 3, we observe a substantial increase in inventory levels even with small ($\rho = 0.025$) levels of inter-customer demand correlation. This result is consistent with our observations from studying networks with a common value of demand correlation across all pairs of customers. However, by introducing geographically-based differences in demand correlation values–using our method to ensure the statistical validity of the correlation matrix for the customer network–we are able to more extensively, and more precisely, demonstrate the effect of increasing levels of correlation on safety inventory levels in the network, even as the demands of customers that are geographically distant remain...
independent. Finally, we note the “anti-pooling” effect that results from increasing correlation levels; producing an increasingly more decentralized network (i.e., one with more facilities in the solution).

Table 3: Average Impact of Inter-customer Correlation on Solution Structure

<table>
<thead>
<tr>
<th>ρ₀</th>
<th>0</th>
<th>0.025</th>
<th>0.050</th>
<th>0.100</th>
<th>0.100</th>
<th>0.300</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ₁</td>
<td>0</td>
<td>0.025</td>
<td>0.025</td>
<td>0.050</td>
<td>0.050</td>
<td>0.100</td>
</tr>
<tr>
<td>ρ₂</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

# of Facilities in Solution:
- Average: 8.2, 9.7, 9.7, 10.8, 10.9, 14.2
- Safety Stock (% of Expected Weekly Demand):
  - Average: 4.9, 5.7, 6.2, 7.1, 7.1, 9.7
- % Increase in Safety Stock:
  - Average: 18.3, 27.4, 46.5, 46.5, 99.6
  - Min.: 9.3, 15.0, 29.5, 29.5, 78.8
  - Max.: 35.2, 44.0, 60.2, 60.2, 132.4

6. Conclusions and Directions for Future Work

We investigate large-scale facility location problems that account for the risk pooling effects associated with centralizing safety stocks in the presence of customer demands that may be correlated across customer locations. Through an initial study, assuming a constant correlation coefficient for all customer pairs, we show the significant impact inter-customer correlation has on safety stock inventory levels. We then present and demonstrate a method to generate valid correlation matrices where multiple inter-customer correlation values are present. This technique is applied to large datasets under the premise that correlation of demand between customer locations decreases as the distance between customer locations increases. Here, we show safety stock cost and solution structure are significantly affected by the presence of even modest levels of demand correlation, illustrating the importance of considering demand correlation in location problems.

There are several important avenues for future research. Demonstrating how the matrix generation technique can be extended to more than three groups of customers will allow the study of more practical industry applications. Studying problems that incorporate additional costs, such as inbound transportation costs, cycle stock-based pooling effects, and scale economies in transportation costs, will offer additional insights into the impact of correlated demands on the solution costs and the underlying network design.
Finally, the study of approximation techniques for the explicit calculation of safety stock costs could offer less computationally complex means to capture the impact of safety costs when inter-customer correlations are present.

Appendix A

In general we assume that the correlation $\rho_{ij}$ between the demands of two customers, one in group $i$ and the other in group $j$, has the form $\rho_{ij} = \rho_{|i-j|}$. To characterize the feasible combinations of $\rho_0$, $\rho_1$, and $\rho_2$, we let $m_i$ denote the number of customers in group $i$, with $m_i \geq 2$ for $i = 1, 2, 3$; and let $I_{m_i}$ denote the $m_i \times m_i$ identity matrix for $i = 1, 2, 3$. Similarly we let $1_{m_i}$ and $0_{m_i}$ denote the $m_i$-dimensional column vectors of ones and zeros, respectively; and $\mathbf{U}_{m_i} = 1_{m_i}'1_{m_i}$ the $m_i \times m_i$ matrix of ones for $i = 1, 2, 3$. We let $0_{m_i, m_j} = 0_{m_i, m_j}'$ denote the $m_i \times m_j$ matrix of zeros for $i, j = 1, 2, 3$. To simplify some of the notation, we let $m = m_1 + m_2 + m_3$ denote the total number of customers whose demands are to be modeled; and because $m_j \geq 2$ for $i = 1, 2, 3$, we always have $m \geq 6$. Finally for $i = 1, 2, 3$, we let $Q_i$ denote an $m_i \times (m_i-1)$ matrix such that $[m_i^{-1/2}1_{m_i}, Q_i]$ is an $m_i \times m_i$ orthogonal matrix. In particular, note that for $i = 1, 2, 3$ we can always start from the vector $m_i^{-1/2}1_{m_i}$ in $m_i$-dimensional Euclidean space $\mathbb{R}^{m_i}$ to select a set of $m_i \times 1$ vectors

\[ \{m_i^{-1/2}1_{m_j}; V_j: j = 1, \ldots, m_i - 1\} \]

that form an orthonormal basis of $\mathbb{R}^{m_i}$; and thus we can take $Q_i = [V_1, \ldots, V_{m_i-1}]$. For a specific example of such a construction, see p. 71 of Searle [1982].

Proposition 1. The $m \times m$ correlation matrix $R$ of customer demands has the form

\[
\begin{bmatrix}
(1-\rho_0)I_{m_1} + \rho_0 \mathbf{U}_{m_1} & \rho_1 1_{m_1}'1_{m_2} & \rho_2 1_{m_1}'1_{m_3} \\
\rho_1 1_{m_1}'1_{m_2} & (1-\rho_0)I_{m_2} + \rho_0 \mathbf{U}_{m_2} & \rho_1 1_{m_2}'1_{m_3} \\
\rho_2 1_{m_1}'1_{m_3} & \rho_1 1_{m_3}'1_{m_2} & (1-\rho_0)I_{m_3} + \rho_0 \mathbf{U}_{m_3}
\end{bmatrix};
\]
and \( \mathbf{R} \) is positive definite if and only if the following conditions hold simultaneously:

\[
- \frac{1}{m_1 - 1} < \rho_0 < 1,
\]

\[
|\rho_1| < \sqrt{\frac{[1 + (m_1 - 1)\rho_0][1 + (m_2 - 1)\rho_0]}{m_1 m_2}},
\]

and given the values of \( \rho_0 \) and \( \rho_1 \) satisfying (7) and (8) we take

\[
\rho_2^* < \rho_2 < \rho_2^{**},
\]

where \( \rho_2^* < \rho_2^{**} \) and \( \rho_2^*, \rho_2^{**} \) are the roots of the following quadratic equation in \( \rho_2 \):

\[
\det(\mathbf{B}) = 0,
\]

with

\[
\mathbf{B} \equiv \begin{bmatrix}
1 + (m_1 - 1)\rho_0 & \rho_1 \sqrt{m_1 m_2} & \rho_2 \sqrt{m_1 m_3} \\
\rho_1 \sqrt{m_2 m_2} & 1 + (m_2 - 1)\rho_0 & \rho_1 \sqrt{m_2 m_3} \\
\rho_2 \sqrt{m_3 m_3} & \rho_1 \sqrt{m_3 m_3} & 1 + (m_3 - 1)\rho_0
\end{bmatrix}
\]

**Proof.** It is easy to check that the \( m \times m \) matrix
is orthogonal so that $HH' = I_m$ and that

$$HRH' = \begin{bmatrix} B & 0_{3,m-3} \\ 0_{m-3,3} & (1 - \rho_0)I_{m-3} \end{bmatrix},$$

where $B$ is given by Equation (11). Since $H$ is orthogonal, we see that $R$ is positive definite if and only if $HRH'$ is positive definite. Because $HRH'$ is symmetric, it is positive definite if and only if its leading principal minors are all positive; see, for example, p. 205 of Searle [1982]. Moreover, from p. 266 of Searle [1982], we see that

$$(1 - \rho_0)^{m-3} \det B = (1 - \rho_0)^{m-3} \det B'$$

(recall that $m \geq 6$); and thus $R$ is positive definite if and only if $1 - \rho_0 > 0$ and the leading principal minors of $B$ are positive. Combining the inequality $1 - \rho_0 > 0$ with the requirement that the first-order leading principal minor of $B$ must be positive yields Equation (7). Requiring the second-order leading principal minor of $B$ to be positive yields Equation (8). Finally if we take $\rho_0$ and $\rho_1$ as given quantities satisfying (7) and (8), then Equation (10) is a quadratic equation in $\rho_2$ having the form

$$a \rho_2^2 + b \rho_2 + c = 0,$$

where the quadratic coefficient is given by $a = -m_1m_3[1 + (m_2 - 1)\rho_0]$, the linear coefficient is given by $b = 2m_1m_2m_3\rho_1^2$, and the intercept is given by

$$c = \frac{\prod_{t=1}^{3}[1 + (m_t - 1)\rho_0] - \rho_2^2m_2m_3[1 + (m_1 - 1)\rho_0] - \rho_2^2m_1m_3[1 + (m_2 - 1)\rho_0]}{\rho_2^2m_1m_2m_3[1 + (m_1 - 1)\rho_0] - \rho_2^2m_1m_2[1 + (m_3 - 1)\rho_0]}.$$

Because $a < 0$, we must have $\det(B) > 0$ for all $\rho_2 \in (\rho_2^*, \rho_2^{**})$, where $\rho_2^*$ and $\rho_2^{**}$ are the roots of Equation (10) and $\rho_2^* < \rho_2^{**}$. This completes the proof of Proposition 1.
Remark 1. In the proof of Proposition 1, we are of course assuming that the quadratic equation
\[ a\rho_2^2 + b\rho_2 + c = 0 \]
has discriminant \( b^2 - 4ac > 0 \). On the other hand if \( b^2 - 4ac \leq 0 \), then we cannot find any feasible real values of \( \rho_2 \) for which \( R \) will be positive definite. In the special case that all groups have the same size so that \( m = m_1 + m_2 + m_3 \), we see that
\[
b^2 - 4ac = 4m_1^2 \left[ 1 + (m_1 - 1)\rho_0 \right] \left[ \rho_0^2 - \rho_1 \rho_2 \right] > 0
\]
if Equation (8) is satisfied; and thus when all groups have the same size, Proposition 1 is guaranteed to provide a means of choosing feasible real values of \( \rho_0 \), \( \rho_1 \), and \( \rho_2 \) for which \( R \) will be positive definite. Moreover in the case of unequal group sizes, if we pick \( \rho_0 \) to satisfy Equation (7) and if we take \( \psi \in (-1, 1) \) such that \( \rho_1 = \psi \rho_0 \) satisfies Equation (8), then it can be shown that the choice \( \rho_2 = \psi \rho_1 = \psi^2 \rho_0 \) always ensures that \( B \) (and hence \( R \)) is positive definite. Although we are unable to prove that \( b^2 - 4ac \) must be positive for all possible values of \( m_i \) with \( m_i \geq 2 \) for \( i = 1, 2, 3 \) and for all possible values of \( \rho_0 \) and \( \rho_1 \) that satisfy Equations (7) and (8), in our computational experience we have always been able to find feasible real values of \( \rho_0 \), \( \rho_1 \), and \( \rho_2 \) for which \( R \) is positive definite using Equations (7) through (11) as given in Proposition 1.

7. References


