Article

Integrated Inventory Transshipment and Missing-Data Treatment Using Improved Imputation-Level Adjustment for Efficient Cross-Filling

Hang Thi Thanh Vu and Jeonghan Ko *

Department of Industrial Engineering, Ajou University, 206 Worldcup-ro, Suwon-si 16499, Korea
* Correspondence: jko@ajou.ac.kr; Tel.: +82-31-219-2335

Abstract: This research investigates an integrated problem of transshipment for cross-filling and imputation for missing demand data. Transshipment for cross-filling has proved effective in mitigating shortages with relatively low inventory, thus reducing resource consumption in inventory management. Although accurate demand data are critical for cross-filling decision making, some demand data are inevitably incomplete. These missing data should be treated for effective transshipment operations. Despite the importance, these missing data issues have not been adequately studied for transshipment problems. This paper addresses how transshipment can be conducted under missing demand conditions. A novel integrated problem is established to combine demand-data imputation processes and transshipment decisions. Imputation strategies and new algorithms suitable for transshipment are developed to handle missing demand data. Diverse demand and transshipment cases are analyzed for cost-effectiveness. The analysis uncovers that conventional straightforward imputation methods result in inferior transshipment decisions. The study also reveals that imputed values should be adjusted to appropriate levels for transshipment to be effective. The strong interplay between imputation processes and shortage prevention is also discovered for transshipment with missing demand. This study demonstrates how inventory transshipment can be carried out successfully with appropriate treatment of missing demand data in practice.

Keywords: transshipment; cross-filling; inventory management; pooling; missing data; imputation; resource reduction; sustainability

1. Introduction

This paper investigates a novel problem of integrating inventory transshipment for conducting cross-filling among distribution centers and data imputation for handling incomplete demand data in transshipment. Although accurate demand data are critical in cross-filling decision making, missing values in demand data are inevitable. These incomplete values should be handled properly to generate transshipment decisions, because transshipment decisions are strongly based on the demand data. However, conventional imputation methods often lead to incorrect transshipment decisions. In addition, the impact of inventory and imputation parameters on transshipment has not been fully investigated. It is also challenging to streamline statistical imputation processes and the transshipment optimization process involving vehicle routing and inventory management. This study presents an integrated approach to solve these problems.

Cross-filling among distribution centers (DCs) has demonstrated its effectiveness in preventing shortages with comparatively low inventory levels at DCs [1–3]. The shortage at DCs often leads to a shortage at retail stores, resulting in sales loss and customer dissatisfaction. A high inventory level often incurs additional storage and handling activities as well as high holding costs and a substantial risk of inventory obsolescence. These often result in great financial loss, energy consumption, greenhouse gas emissions, and waste disposal. This is especially true for cold-chains or perishable items such as medicine or fresh...
food. A high inventory level is also not desirable for short-product-lifecycle items such as high-technology products. To avoid shortage without high inventory levels, cross-filling has often been used as one of the effective strategies. Cross-filling is to transship product items of a DC to other DCs experiencing a shortage of the same product items. Figure 1 illustrates the transshipments among DCs and their routes. In Figure 1, DCs 1 to 4 cross-fill inventory each other by transshipping. DC 5 does not participate in this specific transshipment, because shortage mitigation by transporting goods to or from DC 5 is probably not cost-effective. Transshipment (cross-filling) functions as a pooling strategy at DC levels to reduce the overall inventory levels for both financial and environmenta sustainability.

Figure 1. Illustration of transshipment operations among distribution centers.

Accurate demand data are critical in cross-filling decision making and operations. All strategic and operation plans are based on demand data. In particular, for cross-filling, the quantity and route of transshipment is determined by the amount of demand to fill and possible inventory and shortage. Thus, correct and timely demand data are indispensable for optimal transshipment operations.

Inevitable missing values in demand data should be handled properly for effective transshipment operations. Each DC receives orders from its retailers typically on a regular basis, e.g., daily. These demand data for a DC may contain missing values due to many reasons, such as monitoring failures or data-logging errors. A late rush order may arrive after an ordering time window has closed. These missing demand values should be treated properly for effective transshipment planning; simply ignoring them will lead to possible shortages and flawed transshipment routes. Thus, these missing demand data should be replaced with appropriate values to be included in optimal transshipment planning.

Although missing demand data strongly affect the quality of transshipment decisions, the missing data issues in the transshipment problem have not been thoroughly investigated. The previous studies conducted extensive research on transshipment policies or optimal transshipment routes. However, few of these studies have fully addressed how to handle the missing demand data cases for transshipment. Preliminary studies revealed that conventional simple imputation resulted in high shortage penalties. In addition, this problem is not a simple shortage versus transshipment cost problem. Missing demand values and replaced ones affect transshipment routes and amount. Conversely, vehicle routing routes and inventory parameters affect how much the imputed values influence transshipment. Therefore, either of the imputation or transshipment problems should not be
This study investigates how inventory transshipment can be carried out effectively under missing demand data conditions. In other words, the primary goal of this study is to answer the question of what the most effective inventory and imputation strategies are to improve transshipment decisions on the condition that some of the demand data are missing. First, this paper characterizes the problem of inventory transshipment with missing demand data. This research also compares different approaches to handling missing demand data. Several imputation methods are proposed to handle the missing data, considering different inventory strategies. This paper also demonstrates how the imputation methods can be incorporated into transshipment decisions and operations for cross-filling. In addition, this paper analyzes the qualitative and quantitative impact of the data missingness and imputation methods on the cost saving by transshipment.

One of the primary contributions of this study is to show how imputation approaches can be integrated into transshipment decisions and operations for cross-filling. The newly developed integrated approach combines a statistical imputation process and adjustment algorithm with transshipment decision procedures, thus overcoming the limitations of the current separated methods. This study also uncovers the characteristics of different demand imputation and adjustment methods for the application to inventory transshipment. The proper imputation levels and adjustment are also obtained to achieve cost-effective transshipment. In addition, this study provides managerial insights into the effect of missing data and imputation methods on the inventory transshipment.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. Section 3 describes the transshipment model and treatment methods with missing demand data. Section 4 reports case studies and analyzes the results. The conclusions and future research directions are presented in Section 5.

2. Review of the State-of-the-Art Research

Transshipment or cross-filling has been considered important inventory pooling strategies [1,3,4] for supply chains. Diverse transshipment problems have been studied, depending on inventory conditions and mathematical models [2]. Rim and Vu developed a genetic algorithm to solve transshipment vehicle routing with simultaneous pickup and delivery (TVRSPD) problems [1]. A mixed-integer program was suggested for a generalized transshipment problem [5]. These studies seek optimal transshipment routes and amounts in response to shortage caused by high demands.

The transshipment problems have been studied for a variety of aspects and properties. The impact of lateral transshipment on the bullwhip effect was investigated in a two-tiered supply chain system [6]. Heuristic policies were suggested for a multilocation system with combined transshipment and production control [7]. A problem of group configuration was investigated for within-groups transshipment [8]. DCs’ behaviors were examined for their willingness to share all inventory in a complete network [9] or complete pooling condition [10]. A combined control policy for production and transshipment was proposed for two location firms [11]. A transshipment policy with two thresholds for a logistics service system was suggested [12]. Transshipment between two newsvendor retailers was investigated for the impact on the optimal inventory [13]. Heuristic algorithms were also proposed for a transshipment inventory-routing problem [14]. Other diverse aspects of supply chains have been described, including sustainability and new technologies in supply chain network design [15]. Other topics include closed-loop supply chains [16] and healthcare logistics and service networks [17].
Extensive research has been conducted to address demand uncertainty. Demand is not always completely observable, or inventory records may contain errors [18]. Firms cannot accurately forecast demand or inventory in a volatile market. The newsvendor problem with incomplete demand information has been considered for a long time [19]. Extensive studies related to demand uncertainty exist in the literature, e.g., a method based on order statistics [21], stocking policy based on bootstrapping [22], and demand rate uncertainty model using relative entropy [23]. Other examples include a transshipment policy described by demand uncertainty and transfer lead time fuzziness [24].

A missing data problem is one of the common real-world data uncertainty issues. For example, missing data appear in survey research due to inappropriate survey frames, as well as null or irrelevant responses [25]. Incomplete data also appear in sensor networks owing to power outages, integrity attacks, or bit errors in transmission [26,27]. Financial data also contain missing or partial data due to market closures for holidays, failure to collect the data within the prescribed time frame, or recording errors [28]. Missing data issues also occur in many other fields, such as industrial database processes [29], telecommunications and computer network management [30], and transportation management systems [31].

Missing data can be handled using various approaches. The arrays containing missing values may be deleted, or missing values may be imputed [32]. Numerous approaches to impute missing data exist, such as mean [33], regression [34], and multiple imputations [35]. Different machine learning approaches have been employed for forecasting and backcasting missing demand data due to improper database management systems [36]. The application of clustering and missing data imputation were examined for a large electricity-demand dataset [37]. The mechanisms of missing data are often categorized into the following three types and considered accordingly for imputation: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR) [38].

Although transshipment problems, data uncertainty, and missing-data treatment have been investigated extensively, none of the above studies have properly addressed the problem of transshipment under missing demand data. This study is performed to address this gap.

3. Transshipment Model and Missing-Data Treatment

Section 3.1 defines the problem of this paper: an integrated missing-data treatment and transshipment decision problem. Section 3.2 describes the steps of data imputation and transshipment. Section 3.3 presents the two stages of the missing data treatment approaches. Section 3.4 describes the combined mathematical representation of the problem.

3.1. Problem Definition and Assumptions

This section describes the transshipment processes and the treatment of missing demand data for the transshipment decisions. We assume that a central decision maker (headquarters, HQ) collects all information and makes a cross-filling decision. The HQ acquires the demand information for DCi from the retailers located in the region served by DCi. A missing demand case occurs when a retailer does not provide demand data within the ordering time window. If a case involving missing demand data occurs for item k at DCi, we set the indicator variable Mik = 1. We assume that n = Σk Σi Mik demand values are missing. For planning transshipment, the missing demand data cases may be ignored, or the missing values are to be estimated (imputed).
This study assumes an inventory policy by which all DCs replenish daily up to a certain level \((U)\), because it is known that such policies result in optimal long-term performance \([39]\). The order-up-to level for item \(k\) at DC \(i\) is set as \(U_{ik} = \mu_{ik} + z_{\alpha} \sigma_{ik}\), where \(z_{\alpha}\) is the z-score in the standard normal distribution corresponding to in-stock probability \(\alpha\). In this paper, the in-stock probability is referred to as the service level, which is often defined as the probability of not being stock out during a time period \([40]\). We assume that the upstream suppliers (manufacturers) provide sufficient product items to cover the order-up-to level. Using the target in-stock probability, order-up-to level, and ignored or estimated demand, the expected inventory and shortage levels before transshipment \((I_{ik}^-\) and \(S_{ik}^-\), respectively) are generated.

Using the inventory and shortage information as well as the demand data, the HQ plans the optimal transshipment among DCs. Then, the DCs carry out the transshipment planned by the HQ. The HQ also places a replenishment order to the upstream suppliers for the DCs. After the transshipment is completed, each DC delivers the product items to its retailers. Figure 2 illustrates the relationship in transshipment operations and information exchange.

**Figure 2.** Illustration of transshipment operations and information exchange.

The main decisions for the transshipment include the imputation of missing demand values and planning of the optimal transshipment amounts and routes. Thus, this problem can be considered an integrated problem of statistical treatment of missing values and transshipment vehicle routing with simultaneous pickup and delivery (TVRSPD) \([1,5]\).
Notations

The symbols used in this paper are described in Table 1.

Table 1. Notations.

| Sets and Indexes | Description |
|------------------|-------------|
| $\mathcal{O}$   | Set of product items ($k = 1, 2, \ldots, K$) |
| $\mathcal{J}$   | Set of distribution centers ($i, j = 1, 2, \ldots, Y$) |
| Parameters      | Description |
| $U_{ik}$        | Order-up-to level of product item $k$ at DC $i$ |
| $I_{ik}^-, S_{ik}^-$ | Expected inventory and shortage, respectively, of product item $k$ at DC $i$ if transshipment is not carried out; calculated by allocating on-hand inventory to the daily demand; non-negative integer |
| $p_k$           | Unit price of product item $k$ |
| $m$             | Handling cost per transshipped product item |
| $h$             | Daily inventory holding cost rate ($= 0.25/365 = 0.000684$) |
| $c_{ij}$        | Trucking cost per round trip between DCs $i$ and $j$ |
| $\gamma$        | Shortage cost rate of product item $k$ |
| $v_k$           | Unit volume of product item $k$ |
| $v_0$           | Volume capacity of a truck |
| $t_{ij}$        | Time required for one transshipment round trip between DCs $i$ and $j$ including loading and unloading items |
| $T$             | Maximum time allowed for a transshipment round trip |
| $d_{ik}$        | Daily demand information of product item $k$ received by DC $i$ |
| $d_{ik}^{N1} := d_{ik}^c \cup d_{ik}^m$ | Demand data including both non-missing values ($d_{ik}^c$) and imputed ones ($d_{ik}^m$) when $\sum I_{ik}^- - \sum S_{ik}^- \geq 0$ |
| $d_{ik}^{N2} := d_{ik}^c \cup d_{ik}^m$ | Demand data including both non-missing values ($d_{ik}^c$) and imputed ones ($d_{ik}^m$) when $\sum I_{ik}^- - \sum S_{ik}^- < 0$ |
| $d_{ik}^c$      | Non-missing demand of product item $k$ at DC $i$ |
| $D = \{d_{ik}\}$ | Matrix of demand data including both non-missing values ($d_{ik}^c$) and imputed ones ($d_{ik}^m$) |

Decision variables

| Description |
|-------------|
| $X_{ijk}$   | Quantity of product item $k$ to be transshipped from DC $i$ to $j$; non-negative integer |
| $N_{ij}$    | Number of trucks used for transshipment between DCs $i$ and $j$, $1 \leq i \neq j \leq Y$; non-negative integer |
| $I_{ik}^+, S_{ik}^+$ | Expected inventory and shortage, respectively, of product item $k$ at DC $i$ if transshipment is carried out; non-negative integer |
| $M_{ik}$    | Indicator of demand information missingness for product item $k$ at DC $i$; $M_{ik} = 0$ if $d_{ik}$ is given, and $M_{ik} = 1$ if $d_{ik}$ is missing |
| $d_{ik}^{m*}$ | Temporary imputed data of product item $k$ at DC $i$ |
| $d_{ik}^m$  | Imputed demand of product item $k$ at DC $i$ |

Other notations

| Description |
|-------------|
| $\alpha$  | Target in-stock probability |
| $\beta$   | Rate of inventory increases in imputation, reflecting demand fluctuation |
| $c$       | Coefficient of variation (CV) in demand, reflecting the relative variability of the demand compared with its mean |

3.2. Timeline of the Operations and Information

This section describes the timeline of the operations, information stream, and decisions for the inventory management with transshipment under the condition of missing demand data. The timeline clarifies the relationship among operations [1,5] and decisions as well as calculations of relevant values.

The timeline is organized in nine steps, as follows, and is illustrated in Figure 3. The time on the right-hand side of Figure 3 is only for illustrating relative times.
Figure 3. Timeline of the operations and information for inventory management with transshipment.

The nine steps of the timeline are defined as follows.

Step 1: The previous day’s replenishment order arrives at each DC from the upstream suppliers.

Step 2: Demand data for the next day are given for DC \( i \) from the retailers located in its region by late afternoon (e.g., 4pm).

Step 3: Missing data cases are handled.
   (a) Identify the missingness \( M_{ik} \) of daily demand data \( d_{ik} \).
   (b) Impute the incomplete demand data \( d_{im}^{ik} \) to obtain complete daily demand \( D = d_{ik}^{i} \cup d_{ik}^{m} \).

Step 4: Get the complete demand \( D \).

Step 5: Calculate \( I_{ik} \) and \( S_{ik} \).

Step 6: Calculate the optimal transshipment quantities \( X_{ik} \) and \( N_{ik} \).

Step 7: Place a replenishment order to suppliers.

Step 8: Have received the previously unknown data: \( d_{ik}^{m} \).

Step 9: Complete delivery to retailers.
Step 4: Expected inventory \( (I^-) \) and shortage \( (S^-) \) are evaluated for each product item at DC \( i \) by assigning its on-hand inventory to the daily demand: for \( d_{ik} > U_{ik} \), \( S^-_{ik} = d_{ik} - U_{ik} \) and \( I^-_{ik} = 0 \); otherwise, \( S^-_{ik} = 0 \) and \( I^-_{ik} = U_{ik} - d_{ik} \).

Step 5: Determine the optimal transshipment plan \((X_{ijk} \text{ and } N_{ij})\) and compute the expected \( I^+_k \) and \( S^+_k \) for the case in which transshipment is carried out.

Step 6: DC \( i \) places a replenishment order to the upstream supplier so that the inventory level after replenishment on the next day becomes \( U_{ik} \).

Step 7: Transshipments are carried out among DCs.

Step 8: The true value of the missing demand may become available.

Step 9: Delivery from DCs to the retailers are carried out.

Go to Step 1 for the next day’s operation.

Without loss of generality, this study assumes that the nine steps repeat daily. This time interval can be set to any length if necessary.

3.3. Treatment of Missing Demand Data

This section describes the imputation strategies and algorithm used for missing demand data. If demand data are missing for item \( k \) at DC \( i \), then indicator variable \( M_{ik} \) is set to 1. We assume that \( n = \sum \sum M_{ik} \) demand values are missing or incomplete.

In our transshipment problem, missing data occur when retailers do not provide demand data within the ordering time window. Missing data are the result of external factors (lateness of demand information), indicating that the missing mechanism in this case is MAR.

The missing demand data cases can be handled in a variety of ways. They may be simply disregarded during transshipment planning. The missing values can be estimated and replaced (imputed) for a more effective transshipment planning. Without loss of generality, this study investigates four representative imputation methods corresponding to common protection levels against shortage.

The imputation process consists of two stages. The first stage is to set a temporary imputation value, depending on the shortage mitigation strategies. The second stage is to adjust the imputed value to a valid one.

3.3.1. First Stage of the Imputation Process: Value Selection and Calculation

As the first stage, temporary imputed value \( d_{ik}^{mp} \) is set using one of the following methods:

\[
\begin{align*}
    d_{ik}^{mp} &= \begin{cases} 
    0 & \text{(Method 1) No protection} \\
    \frac{\sum_{i \in J} d_{ik}}{1 - \sum_{i \in J} M_{ik}} & \text{(Method 2) Low protection} \\
    \frac{U_{ik}}{a} & \text{(Method 3) Medium protection} \\
    U_{ik}(1 + \beta) & \text{(Method 4) High protection}
    \end{cases}
\end{align*}
\]  

(1)

These imputation treatments are characterized by shortage protection levels. Facing missing demand data, a decision maker can select from diverse options. The decision maker may decide to disregard the missing values as if the demand does not exist (imputation with “0”). The decision maker may replace them with the average of the known data. The decision maker may impute them with higher values for better protection (reduced shortage).

Four main strategies and corresponding imputation methods are explained below:

(1) No protection imputation (imputation with “0”). In this method, we replace the missing values at \( M_{ik} = 1 \) with zeros, such that \( d_{ik}^{mp} = 0 \). In other words, we assume that no demand exists if these demand values are not available from the downstream retailers. This imputation has a consequence: it makes the calculated value of the expected shortage become zero. This is caused by the current positive inventory level at the time of imputation calculation. This imputation method is simple and easy
to use. However, it has a risk that an actual shortage case may be designated as a no-shortage case.

(2) Low protection (imputation with average): In this method, the missing data are imputed with the mean of the known demand data. For example, for item $k$, the values from all the other DCs having received demand data can be used to calculate the average value. Namely, $d_{ik}^{m*} = \frac{\sum_{i=1}^{Y} d_{ik}}{(Y - \sum_{i=1}^{Y} M_{ik})}$. Although positive demand values are used to impute the missing values, the resulting imputed values are usually lower than $U_{ik}$. Thus, similar to the zero-imputation method above, this imputation method also often leads to evaluating no or little expected shortage for the missing data points. Using this imputation method, some actual shortages may not be prepared adequately.

(3) Medium protection (imputation considering target in-stock probability $\alpha$): This imputation method is based on attempting to approximately satisfy the target in-stock probability. The missing values are imputed as $d_{ik}^{m*} = \frac{U_{ik}}{\alpha}$. Thus, the values exceeding the order-up-to level ($U_{ik}$) are assigned to all missing data points. Please note that there is still possibility of actual shortage if the real demand fluctuates to a high value.

(4) High protection (imputation considering $\alpha$ and demand fluctuation rate $\beta$): This treatment method reflects a strategy that a DC seeks a very low shortage possibility. Here, we assume that a missing data case has an underlying highly fluctuating demand, and shortage is highly expected with the current on-hand inventory. Thus, the imputed values are set higher than those in the previous methods. For example, with the target in-stock probability $\alpha$ and fluctuation level $\beta$, the missing value can be imputed as $d_{ik}^{m*} = \frac{U_{ik}(1 + \beta)}{\alpha}$. Parameter $\beta$ can be determined using a certain ratio to the demand variability or in-stock probability. This imputation method incurs extra transshipment costs but allows more shortage cases to be taken care of.

These four imputation methods are representative cases of common inventory strategies in cross-filling and are chosen to reveal the efficacy of the imputation methods. Although numerous imputation methods exist, it is not possible that this single study covers all such cases. The details of these methods are illustrated with numerical examples in Section 4.

3.3.2. Second Stage of the Imputation Process: Adjustment and Validation

As the second stage, the temporary imputed value is adjusted to a valid one, if necessary. In particular, the imputed value needs to be limited below an upper bound so that the total expected inventory ($I^-$) is sufficient to cover the total shortage ($S^-$) of item $k$ at a DC.

The algorithm presented below finds such an upper bound and adjusts the imputed value not to exceed the bound. The flowchart shown in Figure 4 depicts the logic of the imputation adjustment algorithm. The algorithm adjusts imputed values for all the missing data points of item $k$ at all DCs instead of adjusting only one random point. Computing $d_{ik}^{m*}$ starts from examining a temporary imputed value $d_{ik}^{m*}$ that takes one of the values by Equation (1). Subsequently, an upper bound is calculated, and $d_{ik}^{m*}$ is compared with the bound. The number of missing points is computed for each inventory item. Then, the inventory to transship is divided as evenly as possible but maintaining number integrity. Thus, if necessary, $d_{ik}^{m*}$ is adjusted to obtain a valid $d_{ik}^{m*}$. (Numerical examples are provided in Section 4.5.2).
3.4. Calculation of Transshipment Quantity

Given the imputed demand data and the expected shortage among DCs, the transshipment plan among DCs is determined. Transshipment planning mainly involves the calculation of transshipping amounts and routes. This calculation itself is not the main concern of this study. Thus, various methods, such as those presented in [1,5], can be used, provided that the transshipping amounts and routes are correctly calculated. The following mathematical model is a combination of a mixed-integer program (MIP) by Rim and Jiang [5], selected for convenience, and the new constraints for processing the missing demand data (Equations (5) to (8)), as suggested in Section 3.3.

The integrated optimal transshipment and missing-data treatment is expressed as follows:

$$
Min Z = \sum_{i \in I} \sum_{j \in J} c_{ij} N_{ij} + m \sum_{i \in I} \sum_{j \in J} \sum_{k \in O} X_{ijk} + \sum_{i \in I} \sum_{k \in O} p_k \gamma S_{ik}^+ + \sum_{i \in I} \sum_{k \in O} h p_k U_{ik}. \tag{2}
$$

The objective function in Equation (2) aims to minimize the total costs, including trucking (the first term on the right-hand side), handling (second term), resulting shortage...
(third term), and inventory holding (fourth term) costs. Please note that the objective function can be extended to include various measures.

\[
\sum_{j \neq i} X_{ijk} \leq U_{ik} \quad \forall i, j \in J, \forall k \in O,
\]

\[
S_{ik}^+ \geq d_{ik} - U_{ik} + \sum_{j \neq i} X_{ijk} - \sum_{j \neq i} X_{jik} \quad \forall i, j \in J, \forall k \in O.
\]

Equations (3) and (4) represent that a DC can simultaneously ship-out and ship-in each product item through a series of DCs at most as much as the order-up-to level \(U_{ik}\). Equation (3) expresses that the total number of items transshipped from location \(i\) cannot exceed the maximum inventory amount restricted by the order-up-to level.

\[
d_{ik} - d_{ik}^{N1} = 0 \quad \forall i \in J, \forall k \in O \text{ such that } \max(U_{ik} - d_{ik}, 0) - \sum_{i \in J} \max(d_{ik} - U_{ik}, 0) \geq 0,
\]

\[
d_{ik}^{N1} := d_{ik}^c \cup d_{ik}^m \quad \forall i \in J, \forall k \in O \text{ such that } \max(U_{ik} - d_{ik}, 0) - \sum_{i \in J} \max(U_{ik} - d_{ik}, 0) = 0,
\]

\[
d_{ik} - d_{ik}^{N2} = 0 \quad \forall i \in J, \forall k \in O \text{ such that } \max(U_{ik} - d_{ik}, 0) - \sum_{i \in J} \max(d_{ik} - U_{ik}, 0) < 0,
\]

\[
d_{ik}^{N2} := d_{ik}^c \cup d_{ik}^m \quad \forall i \in J, \forall k \in O \text{ such that } \frac{\sum_{i \in J} \max(U_{ik} - d_{ik}, 0) - \sum_{i \in J} \max(U_{ik} - d_{ik}, 0)}{\sum_{i \in J} M_{ik}} = 0.
\]

Equations (5)–(8) update \(d_{ik}\) of Equation (4), considering the upper bound following the procedures in Section 3.3. Equations (6) and (8) determine valid imputed values \(d_{ik}^m\) from temporary imputed values \(d_{ik}^m^*\), depending on the shortage mitigation strategies. The imputed value \(d_{ik}^m\) is combined with non-missing value \(d_{ik}^c\) for updating \(d_{ik}\). (Note: These equations calculate approximate but very closed upper bounds. The exact bounds are obtained by the algorithm in Figure 4).

\[
\sum_{k \in O} v_k X_{ijk} \leq v^0 N_{ij} \quad \forall i, j \in J,
\]

\[
N_{ij} = N_{ji} \quad \forall i, j \in J.
\]

Equations (9) and (10) restrict the transshipment volumes not to exceed the trucking volume capacity in either direction between a pair of DCs. In Equation (9), \(v_k\) represents the volume of product item \(k\). The left-hand side expresses the total volume of all items. The right-hand side expresses the total transportation capacity between locations \(i\) and \(j\).

\[
(t_{ij} - T) N_{ij} \leq 0 \quad \forall i, j \in J.
\]

Equation (11) prohibits the transshipment between DCs of too long distances.
4. Numerical Analysis and Discussion

This section analyzes the imputation scenarios suggested in Section 3 using numerical simulations and discusses the impact of the imputation methods and managerial implications.

4.1. Parameters and Demand Data

To illustrate the characteristics of the problem and suggested imputation approaches, this study investigates sets of the cases with seven DCs and five inventory items. Although there can be many DCs and a DC may manage numerous inventory items, these case sets can represent the transshipment of critical items and key DCs for a supply chain. The size of these cases would also fit illustrating the effectiveness of different imputation approaches.

Realistic cost parameters for TVRSPD [1] were adopted to evaluate the transshipment performance under practical conditions. We use representative cost factors listed in Table 2 to simulate the transshipment cases and the suggested data imputation approaches. The terms and symbols are defined in Section 3.1. Some parameters are to be changed for sensitivity analysis in later sections. Many cases with different parameter values are to be generated for statistical analysis as shown in later sections.

Table 2. Simulation parameters with seven DCs and five items.

|           | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $v_k$     | 0.26| 0.62| 0.45| 0.62| 0.26|     |     | 1   | 0   | 134.5| 67.1| 70.7| 70.7| 92.2| 143.2|
| $p_k$     | 369 | 481 | 326 | 265 | 165 |     |     | 2   | 0   | 72.1 | 155.2| 148.7| 80.0| 72.1| 80.0 |
| $t_{ij}$  |     |     |     |     |     | 1   | 2   | 3   | 4   | 5   | 6   | 7   |     |     |
| $c_{ij}$  |     |     |     |     |     | 2   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| $p_k$     | U($10 \sim $500) |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $v_k$     | U(0.03 m$^3 \sim 1$ m$^3$) |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $v_0$     | 12.7 m$^3$ |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $h$       | 0.25/365 = 0.00068 per day |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $m$       | $\$0.5 |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $T$       | 3 h  |     |     |     |     |     |     |     |     |     |     |     |     |     |
| $\gamma$  | 0.3 per shortage |     |     |     |     |     |     |     |     |     |     |     |     |     |

The demand of item $k$ at DC $i$ is assumed to follow a normal distribution $N(\mu_{ik}, \sigma_{ik}^2)$, where $\mu_{ik}$ is generated from a uniform distribution with (50, 60). The normality assumption is reasonable enough for many practical cases. In order to analyze more generalized cases not restricted by a specific standard deviation value, demand variability is represented by the ratio to the mean, the coefficient of variation (CV). This paper considers the demand variability that is not so high, meaning a CV not exceeding 0.5. Thus, the representative standard deviation $\sigma_{\bar{y}}$ is set as $c \cdot \mu_{ik} = 0.3\mu_{ik}$, where $c$ is the CV. Using these parameters, the demand values were simulated. The target in-stock probability is set as $\alpha = 0.85$. This specific probability was chosen based on the transshipment problem characteristics as well as our experience and preliminary studies. If the probability is one of the common values in industry such as 95% or 99% in service levels, then transshipment is probably not necessary. This is because high safety stock levels virtually prevent shortages, and thus a very slim chance of cross-filling exists. Recall that the primary purpose of transshipment is to reduce inventory levels. Thus, a slightly lower value among the common probabilities...
was chosen. In fact, we conducted preliminary studies with different probabilities, and the results verified the appropriateness of such number choices. This paper presents only such meaningful cases, and a broader analysis would be beyond the scope of this single paper. The corresponding order-up-to level is calculated $U_{ik} = \mu_{ik} + \sigma_{ik}z_\alpha$. Subsequently, using $U_{ik}$ and the demand value, the expected inventory and shortage before transshipment ($I^-$ and $S^-$) are computed. In this numerical example, parameter $\beta$ for the high-protection imputation method is chosen to be equal to the CV of the demand ($\beta = c = 0.3$). This parameter choice enables us to examine the effect of the inventory increase comparable to the demand variability.

4.2. Simulation with Imputation Scenarios

Although simulations were conducted for a large number of cases in this study, the analysis in this section presents a few important cases representing the main characteristics of the problem. A variety of cases are explained in Sections 4.3 and 4.5 for their distributed characteristics. In addition, for an illustrative purpose to compare the four imputation methods, the examples in this section follow the branch satisfying $\sum I_{ik}^- - \sum S_{ik}^- \geq 0$ on the flowchart in Figure 4 ($d^{NH}_1$). The simulations are conditioned by the location of DCs with missing demand data. We categorize the missing value cases into three main types by their DC locations in relation to the transshipment route. The categories are missing demand at (1) possible transshipment starting DCs sending transshipment items, (2) shortage DCs receiving transshipped items (where $S_{ik}^- > 0$), and (3) DCs not on the transshipment route. In this section, we focus on the cases of the missing data at the shortage DCs, since shortage prevention is the hallmark in cross-filling. The effect of missing data is minimal or irrelevant in the other two DC types.

Table 3 summarizes the data used for the simulations and analyses. In Table 3, the row represents DCs, and column expresses inventory items. The shaded matrix elements indicate that these are on the transshipment routes. In the upper left corner of Table 3, the order-up-to level is presented. Below the order-up-to levels, the original demand data are shown. Below these values, two matrices provide the values of estimated inventory ($I_{ik}^-$) and shortage ($S_{ik}^-$) before transshipment. In the matrix of the estimated shortage before transshipment ($S_{ik}^-$), elements $S_{51}^-$, $S_{12}^-$, $S_{34}^-$, and $S_{75}^-$ present shortage values, as shown at the bottom left of Table 3. Namely, in the original data without missing demand values, the shortage before transshipment is expected at four locations: $S_{51}^-$, $S_{12}^-$, $S_{34}^-$, and $S_{75}^-$. The remaining part of Table 3 indicates the location of the missing demand, and presents the imputed values by the four methods and the corresponding expected inventory and shortage before transshipment. The simulation assumes that some of the demand data are missing at the shortage DCs. For $K = 5$ and $Y = 7$, three ($n = 3$) random missing values exist among those data points. In Table 3, the imputed data are indicated by the bold boxes. In the simulation, $M_{51}$, $M_{34}$, and $M_{75}$ are randomly chosen as missing points, as shown in Table 3.

For the original non-missing demand data case and four imputed demand data cases, the optimal transshipment routes and amounts were determined. For these optimization problems, exact solutions were found using commercial MIP optimization software CPLEX®. Because the problem sizes are not huge, the solution times were short in usual workstations. If the problem size is large, other algorithms [1] could be used to find near optimal solutions. The optimal costs were also computed. The comparison of these results is presented in the subsequent sections.
Table 3. Demand and imputed values by the four methods with estimated inventory and shortage before transshipment.

\[ D \]

\[ \begin{array}{cccccc}
D_{ik} & 1 & 2 & 3 & 4 & 5 \\
1 & 51 & 90 & 65 & 60 & 71 \\
2 & 76 & 55 & 45 & 53 & 17 \\
3 & 60 & 48 & 37 & 90 & 64 \\
4 & 43 & 50 & 65 & 47 & 66 \\
5 & 105 & 33 & 56 & 60 & 65 \\
6 & 48 & 75 & 44 & 51 & 62 \\
7 & 37 & 37 & 62 & 74 & 79 \\
\end{array} \]

\[ D \]

\[ \begin{array}{cccccc}
D_{ik} & 1 & 2 & 3 & 4 & 5 \\
1 & 51 & 90 & 65 & 60 & 71 \\
2 & 76 & 55 & 45 & 53 & 17 \\
3 & 60 & 48 & 37 & 90 & 64 \\
4 & 43 & 50 & 65 & 47 & 66 \\
5 & 105 & 33 & 56 & 60 & 65 \\
6 & 48 & 75 & 44 & 51 & 62 \\
7 & 37 & 37 & 62 & 74 & 79 \\
\end{array} \]

\[ S_{ik} \]

\[ \begin{array}{cccccc}
S_{ik} & 1 & 2 & 3 & 4 & 5 \\
1 & 0 & 14 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 \\
\end{array} \]

In transshipment

Missing data \( M_{ik} = 1 \)

4.3. Result of Each Imputation Scenario

In this section, for each of the four imputation methods, the simulated results in Section 4.2 are analyzed in terms of shortage protection and cost. The differences are also summarized. Analyses of more diverse cases are shown in Section 4.4.

4.3.1. Zero Imputation

Consider the case of imputation with “0” in Table 3 (b). Only one shortage case, \( S_{12} = 90 - 76 = 14 \), is expected if transshipment is not carried out. Using the replaced values (zeros), no shortage is suggested at the three missing demand cases that actually have shortage. \( S_{51}, S_{34}, \text{and } \overline{S}_{25} \) are wrongly set to zero. Thus, the transshipment calculation considers only the suggested shortage at \( S_{12} \) for cross-filling, but not the other real shortage \( S_{51}, S_{34}, \text{and } \overline{S}_{25} \). The transshipment cost for the 14 units of shortages at \( S_{12} \) is approximately \$623. The true shortage penalty is around \$6,046 for the 56-unit shortages in \( S_{51}, S_{34}, \text{and } \overline{S}_{25} \).

The imputation with “0” at the shortage DCs makes the true shortages hidden. This causes the true shortage cases not to be considered in transshipment planning. This incurs the shortage penalty when the true demand is realized. Although this imputation method is a fast and easy way, applying this will induce a negative effect on the fulfillment and customer satisfaction. Thus, disregarding the missing demand data may not be a favorable approach if the shortage cost rate is high.
4.3.2. Low-Protection Imputation

The case of low-protection imputation is shown in Table 3 (c). The missing data points are imputed by the mean of the known demand value of item $k$ among the other DCs (e.g., $d_{m51}^{ik} = (d_{011}^{ik} + d_{021}^{ik} + d_{031}^{ik} + d_{041}^{ik} + d_{061}^{ik} + d_{071}^{ik}) / (7 - 1) = 53$). After imputing the values, only one shortage $S_{12} = 14$ is expected before transshipment. This imputation suggests no expected shortage at the missing data cases, because the imputed demand values are lower than the order-up-to level $U_{ik}$. The true shortages at the missing data points remain unprotected after the transshipment. Thus, in the specific simulations presented in this section, the transshipment and shortage costs are not different from those by the method of imputation with zero (Section 4.3.1). The cases with different costs are presented in a later section. This average imputation approach may not be suitable for handling the missing data at the shortage DCs if the shortage cost rate is not low.

4.3.3. Medium-Protection Imputation

The case of the medium-protection imputation is shown in Table 3 (d). The missing demand data are imputed by values that are higher than $U_{ik}$ based on the target in-stock probability $\alpha$, i.e., $d_{mik}^{ik} = U_{ik}/0.85$, where 0.85 is the target in-stock probability. The following two possible situations may occur in this imputation.

Situation 1: The estimated shortage before transshipment is higher than the actual amount. This triggers transshipping an extra unnecessary amount of inventory. Thus, an additional transshipment cost is incurred. In Table 3, the expected value of $S_{75}$ is 89 but the actual one is 79. A total of 10 extra units of shortages are to be covered, compared with the true value.

Situation 2: The estimated shortage before transshipment is lower than the actual amount. A shortage penalty occurs when the true demand is realized. In Table 3, even though shortages are expected at missing points $S_{51}$ and $S_{34}$, they are still 21 and 6 units fewer compared with the true values, respectively. These remaining shortages incur a shortage penalty of approximately $2800.

In general, this imputation method may cover significant shortages in the missing demand cases, unless the demand fluctuates significantly. However, this imputation cannot guarantee the prevention of all shortage cases.

4.3.4. High-Protection Imputation

The high-protection imputation case is shown in Table 3 (e). The missing points will have the highest imputed values among the four imputation methods. The imputed values are set to $d_{mik}^{ik} = U_{ik}(1 + 0.3)/0.85$. This means that the calculated value for the shortage before transshipment is mostly higher than the true one. An extra transshipment cost can be incurred if the expected shortage amount is higher than the true value. Even in this case, the cost is not high compared with the shortage penalty of the true shortage amount. This will not incur large additional inventory holding cost, because the inventory level remains $U$ until the time delivery to retailers commences.

4.3.5. Summary of Cost Values

Table 4 shows the transshipment cost, actual shortage penalty cost when the true demand appears, and computed total cost. No cost differences occur between the first two methods, because all the missing data cases are considered to have no shortages, as explained in Section 4.3.2. By the high-protection imputation, the transshipment cost is higher than those of the other methods due to the extra shortage amount attempted to cover. However, this case results in the lowest total cost.
Table 4. Summary of the costs for the imputation methods in Table 3.

|                      | Transshipment Cost | True Shortage Penalty | Total     |
|----------------------|--------------------|-----------------------|-----------|
| No imputation        | $623.48            | $5422.80              | $6046.28  |
| Low protection       | $623.48            | $5422.80              | $6046.28  |
| Medium protection    | $786.80            | $2801.70              | $3588.50  |
| High protection      | $979.62            | -                     | $979.62   |

Please note that the decision in this study is not simply comparing the shortage cost and transportation cost. Because vehicle routing is involved and imputation introduces uncertainty, cost alone cannot determine which imputation method is better. Further details are explained in Sections 4.4 and 4.5.

4.4. Trend and Statistical Analysis

Sections 4.2 and 4.3 described the cases in which transshipment has the strong effect on cost saving. This section describes the trend and statistical analysis of the cost reduction by the four imputation methods with a variety of random demand values. The cost savings are compared with those of the cases of no transshipment. For each imputation level, 20 samples were simulated to consider the diverse demand conditions. Figure 5 shows the cost savings using the transshipment and missing data treatment approaches.

Figure 5. Distributions of cost-saving by transshipment at different imputation levels. (Note: The trend showing average cost savings is indicated by the black line).
As shown in Figure 5, as the imputation level increases, so does the average cost saving. The cost saving is generally the largest with the high-value imputation method. While the low imputation value cases suffer significant shortage costs, the high imputation cases have lower shortage costs but additional trucking and handling costs. Because the shortage cost rate is higher than the other rates, the cost saving trend in Figure 5 is consistent with the expected trend.

In addition to the overall trend, subtle differences exist between the neighboring imputation methods. There are trivial differences in the average cost savings between the imputation with "0" and low-protection imputation methods. Slight differences are also observed between the medium and high imputation ones. The cost savings of the first two are from the transshipment itself, not from the additional shortage coverage based on the imputation values. This suggests that an inventory level equivalent to the average demand may not effectively hedge shortage possibility. This is usually the reason why safety stock is necessary. A detailed comparison of these two cases is presented in Section 4.5.

The results also indicate the effectiveness of the medium protection compared with the other methods. In many cases, the cost saving exceeds 70% in methods (3) and (4). However, between the medium- and high-protection approaches, the differences are not significant. The cost saving by method (4) is only 3% higher than that by method (3). Highly imputed values lead to predicting larger shortages beyond the nominal variation range, leading to possibly superfluous transshipment volume. This indicates a possibly negative effect due to extra transshipment or at least the possibility of diminishing returns on inventory investment. The cost saving does not always increase as imputation level increases, even with the high shortage cost. A detailed comparison of these two cases is also provided in Section 4.5.

Figure 5 also shows the distributed values in the cost savings from the 20 samples. The dispersed values reveal the variation from each imputation method. In fact, wide ranges are observed in the cost saving values. The variation is caused by randomness and a variety of reasons: random demand values, resulting various shortage quantities, product item prices, missing-data locations, etc. among the 20 samples. Owing to the wide statistical variability, there are even a few cases in which the highest cost-saving value from the first two methods is higher than the lowest value of the medium- and high-protection methods. This means that in some low-imputation method cases, the demands of missing data cases were high by chance, and they were handled by transshipment. On the contrary, in some high-imputation method cases, the demands of missing data cases were low by chance, and they did not have to be handled by transshipment. Thus, this high imputed value case leads to limited cost saving.

The variations within each imputation method, however, decrease as the imputation level increases. In the low-protection imputation methods, there are high chances that at least some of the actual shortages are not covered by transshipment. Thus, more variability exists in the sense that high shortage costs may or may not occur. The standard deviation is 16% among the 20 samples. In contrast, in the high-protection case, the variations in cost savings are relatively low; the standard deviation is 9.7%, which is lower than the abovementioned value. This variation reduction occurs because of the increased effectiveness of the medium and high imputation methods. As the imputation level increases, the cost savings approach the possible limit. Namely, the shortages have more chances to be considered in transshipment planning and covered by cross-filling. Thus, it is less likely that a large shortage cost occurs. This result implies that the high imputation method can facilitate more stable operations throughout DC networks.

To examine the robustness of the trend, a sensitivity analysis was conducted. The results are summarized in Figure 6.

Figure 6 shows the sensitivity analysis of cost savings by the transshipment with different shortage cost rates and missing data treatment approaches. The shortage cost rate ($\gamma$) is defined as the ratio of the shortage cost to the item price. The trend appears to be consistent. Regardless of the shortage cost rate, the cost saving average exceeds...
60% in methods (3) and (4). The cost savings in methods (3) and (4) are higher than those in methods (1) and (2). One special case is with $\gamma = 0.2$: the cost saving by method (4) is lower than that by method (3). This probably include cases in which highly imputed values lead to predicting too much shortage and consequent extra transshipment volume, and thus net negative cost saving. However, no significant differences exist between the medium and high-protection methods. The sensitivity analysis result also implies that the medium-protection method can be effective.

![Figure 6. Sensitivity analysis of cost-saving by transshipment at different shortage cost rates and imputation levels.](image)

4.5. Detailed Comparative Analysis

This section provides a more detailed analysis of the cost values by the different imputation methods. Whereas Section 4.4 has provided the analysis of the overall trends, this section presents more detailed analyses. Section 4.5.1 compares the zero and low-protection imputation methods. Section 4.5.2 analyzes the high imputation method and imputation adjustment algorithm.

4.5.1. Comparison of the Zero- and Low-Protection Imputation Methods

In the numerical simulation results of Section 4.3, little difference was observed between the imputation methods with zero and the average. This occurred because of several reasons. The individual demands are usually lower than the order-up-to level $U_k$. For the imputation with average, these demand values are averaged to impute the missing data of item $k$. Thus, the imputed value (the average) is also most likely to be lower than $U_k$. Thus, no shortage is suggested for the most of the missing demand cases. Hence, usually the mean imputation method ends up indicating the same no-shortage cases as the zero-imputation method. In other words, in many cases, the average-imputation method will generate a similar result to the zero-imputation method.

However, although the amount is not large, the two methods usually generate different cost savings. Table 5 illustrates the cases of cost saving improvement by the average-imputation, compared with zero-imputation. In Table 5, the missing demand are at locations
With the zero-imputation method, no shortage is suggested at the three missing points that actually have shortages. Thus, no transshipment operations are performed for these locations. By contrast, with the low-protection method (imputation by average), the missing values are replaced with the average value. Thus, in matrix $D$, $d_{74}^n$ becomes 72 and this is greater than $U_{74} = 68$ by 4. This leads to the calculated shortage of $S_{74}^- = 4$. Thus, the transshipment is carried out to prevent this expected shortage. Although the difference is not large, the two methods lead to different transshipment operations and cost values. Table 6 summarizes the improved cost saving of the average imputation method compared with the zero-imputation method.

**Table 5. Example of different results in imputation methods (1) and (2).**

| $u_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1 | 68 | 73 | 71 | 79 | 77 |
| 2 | 68 | 73 | 76 | 76 | 79 |
| 3 | 72 | 79 | 79 | 75 | 76 |
| 4 | 73 | 68 | 68 | 73 | 67 |
| 5 | 77 | 68 | 71 | 66 | 79 |
| 6 | 66 | 71 | 69 | 77 | 69 |
| 7 | 67 | 76 | 73 | 68 | 76 |

| $D$ | 1 | 2 | 3 | 4 | 5 |
|-----|---|---|---|---|---|
| 1   | 65 | 93 | 51 | 71 | 55 |
| 2   | 43 | 36 | 71 | 79 | 52 |
| 3   | 36 | 46 | 35 | 72 | 63 |
| 4   | 55 | 48 | 34 | 78 | 38 |
| 5   | 67 | 47 | 42 | 64 | 43 |
| 6   | 63 | 32 | 46 | 73 | 33 |
| 7   | 67 | 63 | 65 | 74 | 78 |

| $D$ | 1 | 2 | 3 | 4 | 5 |
|-----|---|---|---|---|---|
| 1   | 65 | 0 | 51 | 71 | 55 |
| 2   | 43 | 36 | 71 | 0 | 52 |
| 3   | 36 | 46 | 35 | 72 | 63 |
| 4   | 55 | 48 | 34 | 78 | 38 |
| 5   | 67 | 47 | 42 | 64 | 43 |
| 6   | 63 | 32 | 46 | 73 | 33 |
| 7   | 67 | 63 | 65 | 0 | 78 |

| $l_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|-------|---|---|---|---|---|
| 1     | 3 | 0 | 20 | 8 | 22 |
| 2     | 25 | 37 | 5 | 0 | 27 |
| 3     | 36 | 33 | 44 | 3 | 13 |
| 4     | 18 | 20 | 34 | 0 | 29 |
| 5     | 10 | 21 | 29 | 2 | 36 |
| 6     | 3 | 39 | 23 | 4 | 36 |
| 7     | 0 | 13 | 8 | 0 | 0 |

| $l_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|-------|---|---|---|---|---|
| 1     | 3 | 73 | 20 | 8 | 22 |
| 2     | 25 | 37 | 5 | 76 | 27 |
| 3     | 36 | 33 | 44 | 3 | 13 |
| 4     | 18 | 20 | 34 | 0 | 29 |
| 5     | 10 | 21 | 29 | 2 | 36 |
| 6     | 3 | 39 | 23 | 4 | 36 |
| 7     | 0 | 13 | 8 | 0 | 0 |

| $s_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|--------|---|---|---|---|---|
| 1     | 0 | 20 | 0 | 0 | 0 |
| 2     | 0 | 0 | 0 | 3 | 0 |
| 3     | 0 | 0 | 0 | 0 | 0 |
| 4     | 0 | 0 | 0 | 5 | 0 |
| 5     | 0 | 0 | 0 | 0 | 0 |
| 6     | 0 | 0 | 0 | 0 | 0 |
| 7     | 0 | 0 | 0 | 6 | 2 |

| $s_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|--------|---|---|---|---|---|
| 1     | 0 | 0 | 0 | 0 | 0 |
| 2     | 0 | 0 | 0 | 0 | 0 |
| 3     | 0 | 0 | 0 | 0 | 0 |
| 4     | 0 | 0 | 0 | 0 | 0 |
| 5     | 0 | 0 | 0 | 0 | 0 |
| 6     | 0 | 0 | 0 | 0 | 0 |
| 7     | 0 | 0 | 0 | 0 | 2 |

Known data | No imputation | Low protection
Table 6. Cost saving for the example in Table 5.

|                           | Transshipment Cost | True Shortage Penalty | Total    | % Saving |
|---------------------------|-------------------|-----------------------|----------|----------|
| No imputation             | $699.38           | $3601.50              | $4300.90 | 8%       |
| Low protection            | $701.38           | $3283.50              | $3984.88 | 14%      |

4.5.2. Analysis of the High-Protection Imputation Method and Adjustment Algorithm

In this section, we describe (1) a more detailed analysis of the high-protection cases and (2) how to adjust the imputation values using the algorithm depicted on the flowchart in Figure 4.

As can be seen in Figures 5 and 6, the medium- and high-protection approaches do not have significant differences. In some cases, the higher imputed values even lead to worse cost saving. This trend reflects the fact that the problem of this study is not simply a function of transportation and shortage costs. Because the shortage cost rate is higher than the approximate transportation cost, it is tempting to say that high shortage prevention is preferred so that high imputed values are preferred. The problem condition is not so simple. High imputed values lead to estimating large shortages and transshipping large volumes. However, the large shortage may not be realized. In other words, high imputed values may result in unnecessary cost increase due to extra transshipment but no effective cost reduction by shortage mitigation.

In addition, this analysis concerns the cases in which, for some imputed demand data, the total inventory \( I^− \) of an item \( k \) from all DCs may not be enough to cover the estimated shortage. In a naïve transshipment planning, most stocks are transshipped to the shortage locations just by the least-cost plan. Such transshipment may incur an additional shortage penalty in the sending DCs. If some shortage values by imputation are arbitrarily high, the situation becomes worse. Thus, an upper bound of the imputed values must be specified so that the total shortage to cover is at most equal to the available inventory.

Even though this case can occur in any of the four imputation methods, it has much more chances to exist in the high-protection method due to the high imputed values. The example in this section demonstrates these cases. As shown in Table 7, the missing demand data are at locations \( M_{12}, M_{24}, \) and \( M_{74} \). Using the high-protection method, the temporary imputed values are \( d_{m}^{\text{im}}=112, d_{m}^{\text{im}}=116, \) and \( d_{m}^{\text{im}}=113 \). For item 2, \( \sum_{i} I_{i}^{−} - \sum_{i} S_{i}^{−} = 163 - 39 = 124 \geq 0 \); thus, \( d_{ik}^{m} = d_{21}^{im} = 112 \) (no adjustment). For item 4, \( \sum_{i} I_{i}^{−} - \sum_{i} S_{i}^{−} = 63 - 84 = -21 < 0 \). This means that there are 21 shortages that even the stock of all DCs cannot cover. In Table 7, with the temporary imputed values, \( S_{12}^{+} = 16 \) and \( S_{44}^{+} = 5 \) (known shortage). Thus, we deduct the following values from the temporary imputed values so that the total deducted value is equal to 21. We compute the number of missing points for inventory item 4: \( v = \sum_{i} M_{i4} = 2 \). We compute the division 21/2 as quotient 10 with remainder 1 (\( q = 10 \) and \( l = 1 \) after 21 is divided by \( v = 2 \)). For the first \( l = 1 \) index, where the missing data are located, \( g = 2 \) (DC 2), and because \( q > 0 \), \( d_{24}^{m} = d_{24}^{im} - (1 + q) = 116 - 11 = 105, d_{24}^{m} = d_{24}^{im} - q = 113 - 10 = 103 \). Thus, \( d_{24}^{im} \) and \( d_{44}^{im} \) are adjusted. After adjusting the imputed values to valid ones, no longer shortages after transshipment occur at all DCs.
Table 7. Example of calculating valid $d_{ik}^{mp}$ for the high-protection method using the algorithm in Figure 4.

| $U_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1        | 68| 73| 71| 66| 77|
| 2        | 68| 73| 76| 76| 79|
| 3        | 72| 79| 79| 75| 76|
| 4        | 73| 68| 68| 73| 67|
| 5        | 77| 68| 71| 66| 79|
| 6        | 66| 71| 69| 68| 69|
| 7        | 67| 76| 73| 74| 76|

| $D_{ik}$ | 1   | 2  | 3  | 4  | 5  |
|----------|-----|----|----|----|----|
| 1        | 65  | 93 | 51 | 58 | 55 |
| 2        | 43  | 36 | 71 | 79 | 52 |
| 3        | 36  | 46 | 35 | 72 | 63 |
| 4        | 55  | 48 | 34 | 78 | 38 |
| 5        | 67  | 47 | 42 | 23 | 43 |
| 6        | 53  | 32 | 46 | 59 | 33 |
| 7        | 67  | 63 | 65 | 82 | 78 |

| $I_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1        | 3 | 0 | 20| 8 | 22|
| 2        | 25| 37| 5 | 0 | 27|
| 3        | 36| 33| 44| 3 | 13|
| 4        | 18| 20| 34| 0 | 29|
| 5        | 10| 21| 29| 43| 36|
| 6        | 3 | 39| 23| 9 | 36|
| 7        | 0 | 13| 8 | 0 | 0 |

| $S_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1        | 0 | 20| 0 | 0 | 0 |
| 2        | 0 | 0 | 0 | 3 | 0 |
| 3        | 0 | 0 | 0 | 0 | 0 |
| 4        | 0 | 0 | 0 | 5 | 0 |
| 5        | 0 | 0 | 0 | 0 | 0 |
| 6        | 0 | 0 | 0 | 0 | 0 |
| 7        | 0 | 0 | 0 | 8 | 2 |

| $S_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1        | 0 | 0 | 0 | 0 | 0 |
| 2        | 0 | 0 | 0 | 16| 0 |
| 3        | 0 | 0 | 0 | 0 | 0 |
| 4        | 0 | 0 | 0 | 5 | 0 |
| 5        | 0 | 0 | 0 | 0 | 0 |
| 6        | 0 | 0 | 0 | 0 | 0 |
| 7        | 0 | 0 | 0 | 0 | 0 |

| $S_{ik}$ | 1 | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|---|
| 1        | 0 | 0 | 0 | 0 | 0 |
| 2        | 0 | 0 | 0 | 16| 0 |
| 3        | 0 | 0 | 0 | 0 | 0 |
| 4        | 0 | 0 | 0 | 5 | 0 |
| 5        | 0 | 0 | 0 | 0 | 0 |
| 6        | 0 | 0 | 0 | 0 | 0 |
| 7        | 0 | 0 | 0 | 0 | 0 |

Known data Imputation with $d_{ik}^{mp}$ Imputation with $d_{ik}^{mp}$

4.6. Insight for Practice

This study provides insights on how to make transshipment more effective by proper treatment of missing data in practice. First, the results imply that demand-data imputation in transshipment decision-making is not simply to recover the missing values. Sophisticated data imputation methods exist in the literature. The majority of them focus on recovering the missing values as close to the original ones as possible, especially around mean values.
The numerical studies in this paper suggest that such an approach may fail for specific applications such as transshipment for cross-filling. Transshipment decisions based on average values were inferior to those based on the higher values. This means that a common straightforward imputation did not improve decision making; we should have a better and customized approach to handle missing values. Therefore, this study indicates that managers should not focus on simply how to recover the demand values themselves, but how to set appropriate shortage values for effective transshipment.

The results also suggest that a simple trade-off between inventory and transshipment costs does not always work for cross-filling. Although a higher cost in shortage than transportation may prefer a higher inventory, this trend may stop at a certain point. In cross-filling under a missing-demand condition, a high inventory may lead to unnecessary transshipment. This incurs extra transportation cost but seldom contributes to shortage prevention. This means that managers should find appropriate points from which diminishing returns on inventory investment become apparent. These points should be estimated by a combination of optimization and statistical analysis because of the complex interrelation of the parameters.

This study also implies that handling missing data is not just a one-step imputation for practitioners. The imputed values should be adjusted before being used. The algorithm in Figure 4 was devised for such a case. Without such algorithms, the optimal solution may generate infeasible solutions in practice. This indicates that practitioners should be advised on the practical validity of the statistically processed numbers.

5. Conclusions

This study investigated an integrated problem of demand data imputation and transshipment operation decisions, and the result analysis verified that the integrated approach enhances transshipment performance successfully. First, this paper established a combined problem of inventory transshipment for cross-filling among distribution centers and data imputation for treating missing demand data. This research formulated imputation methods corresponding to principal inventory management strategies. This research also developed algorithms keeping imputed values within practical inventory bounds. In addition, by combining a mathematical program and the developed algorithm, this research provided a mathematical model that allows simultaneous transshipment decisions and demand imputation. Therefore, this study contributes to establishing a theoretical foundation for advanced modeling of the transshipment problem.

Second, this paper uncovered key theoretical characteristics of the integrated imputation-transshipment problem and uncovered the effect of imputation parameters on transshipment performance. Numerical simulations were conducted for diverse demand and imputation cases with practical parameters. A comprehensive analysis of the results reveals that imputed values above a particular level are critical for effective shortage mitigation and cost reduction by transshipment. Moreover, the statistical analysis shows that the consistency of transshipment performance improves as the imputation level increases.

Third, this paper confirms that in practice the transshipment effectiveness can be improved by proper data treatment as suggested in this study. This study shows that demand data imputation can be incorporated into a holistic transshipment process in industry. The method in this paper is straightforward to implement for practical application. This study also demonstrates that transshipment can be performed efficiently even under missing demand data conditions.

The results imply that the novel method proposed in this study performs better for transshipment decision-making than the conventional imputation methods recovering simply the missing values. For example, whereas imputation with the demand average did not result in significant cost savings, imputed and adjusted values sufficiently higher than the average help reduce the shortage cost by 80% on average. This means that specialized guidelines and algorithms are necessary for a specific transshipment problem. This paper also provides managerial insights into how cost parameters, inventory levels, and impu-
tation levels jointly influence transshipment operations and cost reduction in inventory management. A simple trade-off does not exist between inventory and transshipment costs in cross-filling. This means that an appropriate balance should be struck for a specific transshipment condition by the combination of optimization and statistical analysis.

Hence, this study lays the groundwork for sophisticated transshipment operations with data treatment, and the integrated approach can be extended to a variety of directions. This study focused on the key characteristics of the integrated transshipment and imputation problem, but this single paper does not address all possible extended topics. For example, this paper did not consider diverse probability distributions of the demand. A variety of vehicle routing conditions were not considered. These limitations can be addressed in the future work that investigates diverse problem conditions. A wide range of demand properties can be considered with different missing data conditions. Diverse imputation methods can be employed for large demand-data sets to take advantage of advanced statistical learning methods. Diverse cross-filling situations can also be reflected in transshipment modeling. The examples include vehicle availability and distance restriction. The future research may also consider a variety of performance measures such as financial risks or various sustainability indexes.

Author Contributions: Conceptualization, H.T.T.V. and J.K.; methodology, H.T.T.V. and J.K.; software, H.T.T.V.; validation, H.T.T.V. and J.K.; investigation, H.T.T.V. and J.K.; data curation, H.T.T.V.; writing—original draft preparation, H.T.T.V. and J.K.; writing—review and editing, H.T.T.V. and J.K.; visualization, H.T.T.V.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1095569) and the Ajou University research fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Simulated demand data are available from the authors upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Rim, S.-C.; Vu, H.T.T. Transshipment Vehicle Routing with Pickup and Delivery for Cross-Filling. Math. Probl. Eng. 2021, 2021, 6667765. [CrossRef]
2. Chiou, C.-C. Transshipment problems in supply chain systems: Review and extensions. In Supply Chain; BoD: Berlin, Germany, 2008; pp. 427–448.
3. Ballou, R.H.; Burnetas, A. Planning multiple location inventories. J. Bus. Logist. 2003, 24, 65–89. [CrossRef]
4. Evers, P.T. Filling customer orders from multiple locations: A comparison of pooling methods. J. Bus. Logist. 1999, 20, 121.
5. Rim, S.C.; Jiang, J. Multi-item, multi-location transshipment model for cross-filling. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Riyadh, Saudi Arabia, 26–28 November 2019.
6. Wei, Y.; Chen, F.; Xiong, F. Dynamic complexities in a supply chain system with lateral transshipments. Complexity 2018, 2018, 3959141. [CrossRef]
7. Bhatnagar, R.; Lin, B. The joint transshipment and production control policies for multi-location production/inventory systems. Eur. J. Oper. Res. 2019, 275, 957–970. [CrossRef]
8. Herer, Y.T.; Tzur, M.; Yücesan, E. Transshipments: An emerging inventory recourse to achieve supply chain leagility. Int. J. Prod. Econ. 2002, 80, 201–212. [CrossRef]
9. Lien, R.W.; Iravani, S.M.R.; Smilowitz, K.; Tzur, M. An efficient and robust design for transshipment networks. Prod. Oper. Manag. 2011, 20, 699–713. [CrossRef]
10. Paterson, C.; Kiesmüller, G.; Teunter, R.; Glazebrook, K. Inventory models with lateral transshipments: A review. Eur. J. Oper. Res. 2011, 210, 125–136. [CrossRef]
11. Dhahri, A.; Gharbi, A.; Ouhimmou, M. Integrated production-transshipment control policy for a two-location unreliable manufacturing system. Int. J. Prod. Econ. 2022, 247, 108440. [CrossRef]
12. Cheng, L.; Guo, X.; Li, X.; Yu, Y. Data-driven ordering and transshipment decisions for online retailers and logistics service providers. Transp. Res. Part E Logist. Transp. Rev. 2022, 161, 102684. [CrossRef]
13. Rudi, N.; Kapur, S.; Pyke, D.F. A two-location inventory model with transshipment and local decision making. *Manag. Sci.* 2001, 47, 1668–1680. [CrossRef]

14. Coelho, L.C.; Cordeau, J.F.; Laporte, G. The inventory-routing problem with transshipment. *Comput. Oper. Res.* 2012, 39, 2537–2548. [CrossRef]

15. Fathollahi-Fard, A.M.; Dulebenets, M.A.; Tian, G.; Hajiaghaei-Keshteli, M. Sustainable supply chain network design. *Environ. Sci. Pollut. Res. 2022, 1–3.* [CrossRef] [PubMed]

16. Samadi, A.; Mehranfar, N.; Fard, A.M.F.; Hajiaghaei-Keshteli, M. Heuristic-based metaheuristics to address a sustainable supply chain network design problem. *J. Ind. Prod. Eng.* 2018, 35, 102–117. [CrossRef]

17. Fathollahi-Fard, A.M.; Ahmadi, A.; Karimi, B. Sustainable and robust home healthcare logistics: A response to the COVID-19 pandemic. *Symmetry 2022, 14, 193.* [CrossRef]

18. Frankel, R.; Keller, S.B.; Keller, B.C.; Gibson, B.J.; Hanna, J.B.; Defee, C.C.; Chen, H.; Tate, W.; Sanders, N.R.; Goldsby, T.J.; et al. *Definitive Guides for Supply Chain Management Professionals (Collection)*; Pearson: London, UK, 2014.

19. Karlin, S.; Arrow, K.; Scarf, H. A min–max solution of an inventory problem. In *Studies in the International Theory of Inventory and Productions*; Stanford University Press: Stanford, CA, USA, 1958.

20. Lu, Y.; Song, M.; Yang, Y. Joint inventory and pricing coordination with incomplete demand information. *Prod. Oper. Manag. 2016, 25, 701–718.* [CrossRef]

21. Lordahl, A.E.; Bookbinder, J.H. Bookbinder. Order-statistic calculation, costs, and service in an (s, Q) inventory system. *Nav. Res. Logist. 1994, 41, 81–97.* [CrossRef]

22. Fricker, R.D., Jr.; Goodhart, C.A. Applying a bootstrap approach for setting reorder points in military supply systems. *Nav. Res. Logist. 2000, 47, 459–478.* [CrossRef]

23. Lim, A.E.B.; Shanthikumar, J.G. Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res. 2007, 55, 198–214.* [CrossRef]

24. Hmiden, M.; Ben Said, L.; Ghedira, K. Transshipment problem with uncertain customer demands and transfer lead time. In *Proceedings of the 2009 International Conference on Computers & Industrial Engineering*, Troyes, France, 6–9 July 2009.

25. Brick, J.M.; Kalton, G. Handling missing data in survey research. *Stat. Methods Med. Res.* 2003, 12, 132–142. [CrossRef]

26. Gruenwald, L.; Halatchev, M. Estimating missing values in related sensor data streams. In *Proceedings of the Eleventh International Conference on Management of Data*, Goa, India, 6–8 January 2005.

27. Berjab, N.; Le, H.H.; Yokota, H. Recovering Missing Data via Top-k Repeated Patterns for Fuzzy-based Abnormal Node Detection in Sensor Networks. *IEEE Access 2022, 10, 61046–61064.* [CrossRef]

28. John, C.; Ekpenyong, E.J.; Nworu, C.C. Nworu. Imputation of missing values in economic and financial time series data using five principal component analysis approaches. *CBN J. Appl. Stat. 2019, 10, 3.*

29. Lakshminarayan, K.; Harp, S.A.; Samad, T. Imputation of missing data in industrial databases. *Appl. Intell. 1999, 11, 259–275.* [CrossRef]

30. Ji, C.; Elwalid, A. Measurement-based network monitoring: Missing data formulation and scalability analysis. In *Proceedings of the 2000 IEEE International Symposium on Information Theory* (Cat. No. 00CH37060), Sorrento, Italy, 25–30 June 2000.

31. Smith, B.L.; Scherer, W.T.; Conklin, J.H. Exploring imputation techniques for missing data in transportation management systems. *Transp. Res. Rec. 2003, 1836, 132–142.* [CrossRef]

32. Lang, K.M.; Little, T.D. Little. Principled missing data treatments. *Prev. Sci. 2018, 19, 284–294.* [CrossRef]

33. Von Eye, A. (Ed.) *Statistical Methods in Longitudinal Research: Principles and Structuring Change*; Elsevier: Amsterdam, The Netherlands, 1990; Volume 1.

34. Brown, C.H. Asymptotic comparison of missing data procedures for estimating factor loadings. *Psychometrika 1983, 48, 269–291.* [CrossRef]

35. Rubin, D.B. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. *Proc. Surv. Res. Methods Sect. Am. Stat. Assoc. 1978, 1, 20–28.*

36. Twumasi, C.; Twumasi, J. Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana. *Int. J. Forecast. 2022, 38, 1258–1277.* [CrossRef]

37. Inman, D.; Elmore, R.; Bush, B. A case study to examine the imputation of missing data to improve clustering analysis of building electrical demand. *Build. Serv. Eng. Res. Technol. 2015, 36, 628–637.* [CrossRef]

38. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Volume 793.

39. Herer, Y.T.; Tzur, M.; Ucesan, E.Y. The multilocation transshipment problem. *IIE Trans. 2006, 38, 185–200.* [CrossRef]

40. Chen, F.Y.; Krass, D. Inventory models with minimal service level constraints. *Eur. J. Oper. Res. 2001, 134, 120–140.* [CrossRef]