8-2017

Essays in Measuring, Controlling, and Coordinating Supply Chain Inventory and Transportation Operations

Payam Parsa
University of Arkansas, Fayetteville

Essays in Measuring, Controlling, and Coordinating Supply Chain Inventory and Transportation Operations

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Engineering

by

Payam Parsa
Islamic Azad University
Bachelor of Science in Industrial Engineering, 2009
Southern Illinois University Edwardsville
Master of Science in Industrial and Manufacturing Engineering, 2013

August 2017
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dr. Manuel Rossetti
Dissertation Director

Dr. Edward Pohl
Committee Member

Dr. Shengfan Zhang
Committee Member

Dr. Brian Fugate
Committee Member

Abstract

Supply chain collaboration programs, such as continuous replenishment program (CRP), is among the most popular supply chain management practices. CRP is an arrangement between two partners in a supply chain to share information on a regular basis for lowering logistics costs while maintaining or increasing service levels. CRP shifts the replenishment responsibility to the upstream partner to avoid the bullwhip effect across the supply chain. This dissertation aims to quantify, measure, and expand the benefits of CRP for the purpose of reducing logistics cost and improving customer service. The developed models in this dissertation are all applied in different case studies supported by a group of major healthcare partners. The first research contribution, discussed in chapter 2, is a comprehensive data-driven cost approximation model that quantifies the benefits of CRP for both partners under three cost components of inventory holding, transportation and ordering processing without imposing assumptions that normally do not hold in practice. The second contribution, discussed in chapter 3, is development of a verifiable efficiency measurement system to ensure the benefits of CRP for all partners. Multi-functional efficiency metrics are designed to capture the trade-off in gaining efficiency between multiple functions of logistics (i.e. inventory efficiency, transportation efficiency, and order processing efficiency). In addition, a statistical process control (SPC) system is developed to monitor the metrics over time. We discuss suitable SPC systems for various time series behaviors of the metrics. The third contribution of the dissertation, discussed in chapter 4, is development of a multi-objective decision analysis (MODA) model for multi-stop truckload (MSTL) planning. MSTL is becoming increasing popular among shippers while is experiencing significant resistance from carriers. MSTL is capable of reducing the shipping cost of shippers substantially but it can also disrupt carriers’ operations. A MODA model is developed for this problem to

incorporate the key decision criteria of both sides for identifying the most desirable multi-stop routes from the perspective both decision makers.

Acknowledgement

First and foremost, I would like to sincerely thank my Ph.D. advisor, Dr. Manuel Rossetti for influencing my character and career in so many ways. He has changed me from a fresh graduate student to a critical thinker and experienced academician in my discipline. I cannot be more grateful for his fairness, flexibility and understanding during the most challenging times in my Ph.D. studies. I would also like to thank my dissertation committee members, Dr. Ed Pohl, Dr. Shengfan Zhang, and Dr. Brian Fugate, for their valuable time to support me during my job search and serving in my Ph.D. committee.

I would like to thank the faculty and staff of the department of Industrial Engineering for allowing me to thrive with their gracious support. Last by not least, I am very thankful to all my friends who have supported me and shaped my journey in Arkansas. Especially, I express my sincere gratitude to Paiman Farrokhvar, Amirali Ghahari, Mitra Panahipour, Ridvan Gedik, Furkan Oztanriseven, Luisa Janer, Alireza Sheikhzadeh, and Taha Havakhor.

Dedication

I would like to dedicate this dissertation to my family for their tremendous support and unconditional love throughout my life. I am certain that none of my success would have been achieved without the sacrifices that my parents made. Time and again they have picked me up and I hope someday I will pick them up too.

Table of Contents

List of Figures

List of Tables

List of Papers

Parsa, P., Rossetti, M. D., Zhang, S. & Pohl, E. A., (2017) “Quantifying the Benefits of Continuous Replenishment Programs for Partner Evaluation” *International Journal of Production Economics, Volume 187 Pages 229-245, ISSN 0925-5273, (chapter 2 is published here)*

Parsa, P., Rossetti, M. D., Zhang, S., (2017) “Multi-Stop Truckload Planning” *Proceedings of the Industrial and Systems Engineering Research Conference (chapter 4 is partially published here)*

1 INTRODUCTION

We are currently in the era of constant and fast change in consumer behaviors and expectations, explosion of new products that seek to fulfill customer's new needs, intense competition between companies in gaining market share, and using data analytics as a leverage to boost sales and lower costs. In such an environment, customer satisfaction cannot be achieved easily. They no longer accept out of stocks, excessive sales prices, and delayed deliveries. Supply chain management is truly challenged at the present time to appropriately cope with these changes in consumer behavior. It is fair to say that the retail industry is a pioneer in elevating customer expectations and innovating solutions. However, other industries, such as healthcare, need to catch up quickly because consumers, soon or later, will expect the same level of standard from every provider. Supply chain collaboration and multi-stop trucking have been amongst the effective solutions that companies use to cope with many of such challenges. In chapter one and two, we shed light on cost modeling and performance measurement of supply chain collaboration programs, while the third chapter devotes to a fairly recent concept in the transportation sector, multi-stop truckload planning.

Different types of supply chain collaboration programs have been proposed and evolved over time since late 1980's when Walmart introduced the concept. Continuous Replenishment Program (CRP), or sometimes known as Vendor Managed Inventory (VMI), is the most common form of such programs. CRP is a collaboration initiative between two partners, normally a supplier and a retailer, to lower the logistics cost and increase service performance. It requires the supplier (upstream partner) to manage the replenishment process using inventory and demand information shared by the retailer (downstream partner). Decision making within supply chain collaboration programs, especially at the managerial level, oftentimes relies on a leader's "gut instinct" rather than data-based analytics. Decisions such as whether to start/continue a CRP relationship with a partner are oftentimes made very simplistically based on past experiences. Likewise, evaluation of current relationships often consists of a simple comparison of the relationship with similar ones. In many cases, collaboration programs are initiated between two partners for a very specific reason, which mostly has minimal improvement impact, while the full potential of the program is yet to be realized. The first chapter of this dissertation contributes to this area by developing a multi-echelon model to approximate the cost of supply chain within a CRP relationship for both partners. The model approximates three cost components of inventory holding, transportation, and order processing for a multi-product, one supplier one retailer system subject to service level constraints. The model is applied on healthcare supply chain network that supports replenishment of healthcare distributors from a manufacturer. The model answers three key questions: 1) What is the cost savings impact of CRP on each partner? 2) How does it vary across a distribution network? 3) How does it vary across different cost components?

The original benefit of CRP or VMI systems is reducing inventory levels by removing the bullwhip effect from the replenishment process in the supply chain (Lee et al., 1997). However, various academic studies and industry practices have revealed that significant transportation and order processing savings can also be achieved in VMI by proper shipment consolidation and timely replenishment (Çetinkaya et al., 2008; Parsa et al., 2017). The challenge begins when the partners care more about some of the benefits than others. This behavior, which is somewhat inevitable, could become destructive and oftentimes leads to the failure of collaborative relationships. It is essential to consider the objectives of all involved partners in order to build trust, maintain a VMI relationship, and utilizing its full potential. Our collaboration with a group of major healthcare partners in the U.S. confirmed that maintaining a successful VMI relationship needs a verifiable efficiency measurement system to ensure the benefits of VMI for all partners.

The crucial point is that an effective system for monitoring VMI relationships needs a group of efficiency metrics that can capture the trade-off in gaining efficiency between various functions of logistics such as inventory holding, transportation and order processing. Since VMI shifts the replenishment responsibility to upstream partner, it can be manipulated to favor a party or at least be conceived for doing so. As Gunasekaran and Kobu (2007) identified through a multi-faceted literature review on supply chain and logistics performance metrics, there are numerous overlapping metrics with 85% of them being quantitative, mostly concerned with financial performance, and focused on a single function of logistics operations (i.e. inventory, transportation, etc.). There has not been a significant work to design metrics that explore the relationship between functions or propose a statistical screening framework to monitor them over time. The second chapter of this dissertation proposes multi-functional efficiency metrics that can capture the trade-off in gaining efficiency between multiple functions of logistics. There is a trade-off between gaining efficiency in inventory holding and in transportation, which mostly concerns with maintaining an optimal level of shipment consolidation. The same is true for inventory holding efficiency versus order processing efficiency. In the second chapter, we develop metrics that can illustrate the status of a system with respect to such tradeoffs over time. We also determine optimal trade-off levels for the developed metrics. In addition, a statistical process control (SPC) system is developed to monitor them over time. The SPC system suggests whether the system is acting normal or if a significant shift has happened. We elaborate the application of metrics and suitable statistical methods to develop SPC systems using datasets obtained from a major healthcare manufacturer.

VMI arrangements empower upstream partners by giving them the full responsibility of replenishing the downstream partners. They can adjust replenishment quantity and timing of their customers in order to use a single vehicle for delivering to multiple locations. This mode of transportation is called multi-stop truckload (MSTL) and is rapidly gaining market share, mainly from the less-than-truckload (LTL) business. Besides VMI, other factors, such as congestion in urban areas and achieving more certain delivery time, have led shippers to use MSTL more frequently. While shippers have shown great interests in offering MSTLs to carriers, carriers have become more reluctant to accepting them. MSTLs can impose extra costs to carriers and disrupt their operations by making drivers unavailable for longer periods and interrupting their network flow balance. On the other hand, rejections from carriers cause shippers to go to their next preferred carrier, which is often more expensive. Therefore, the problem is a multi-objective decision analysis from the perspective of shippers. They should offer multi-stop routes that not only maximize their cost savings but have a desired level of acceptance chance from carriers. In a recent empirical study, Chen and Tsai Yang (2016) has identified the key properties of MSTLs that contribute to their acceptance chance from carriers.

The chapter three of this dissertation proposes a multi objective decision model to identify the best two-stop routes that maximize the cost savings for the shipper and fulfill the most important load acceptance criteria of the carriers, which are *out-of-route miles* and *proximity of stops*. The model provides a trade-off capability for selecting routes with more appeal to either shipper or carrier. The application of the model is discussed for a healthcare supply network. We use weekly forecast data at the SKU level along with shipping and distance information of the distribution network to compute the potential savings of every possible two-stop route via an exhaustive search. The routes with positive savings will be subject to a multi objective decision model that selects the best routes given the load acceptance criteria of carriers. The chapter provides an insightful sensitivity analysis that can help shippers to wisely offer multi-stop routes that maximize their savings and acceptance rate.

References

Çetinkaya, S., Tekin, E., Lee, C.-Y., 2008. A stochastic model for joint inventory and outbound shipment decisions. IIE Trans. 40, 324–340. doi:10.1080/07408170701487989

Chen, X., Tsai Yang, S. lin, 2016. Multi-Stop Trucking : A Study on Cost and Carrier Acceptance. MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

Gunasekaran, A., Kobu, B., 2007. Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. Int. J. Prod. Res. 45, 2819–2840. doi:10.1080/00207540600806513

Lee, H.L., Padmanabhan, V., Whang, S., 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. Manage. Sci. 43, 546–558. doi:10.1287/mnsc.43.4.546

Parsa, P., Rossetti, M.D., Zhang, S., Pohl, E.A., 2017. Quantifying the benefits of continuous replenishment program for partner evaluation. Int. J. Prod. Econ. 187, 229–245. doi:10.1016/j.ijpe.2017.02.017

2 CHAPTER 1

2.1 Introduction

Introduced and popularized in the late 1980's by Walmart and Proctor & Gamble, Continuous Replenishment Programs (CRPs) are one of the most commonly used supply chain collaboration (SCC) initiatives. Variants of CRPs are sometimes represented as Vendor Managed Inventory (VMI) in academia and among practitioners. CRP requires the manufacturer (upstream partner) to manage the replenishment process by using inventory and demand information shared by the distributor (downstream partner). Hence, CRP resembles a centralized inventory control system. CRP has been adopted by many organizations from different sectors, including the healthcare sector, and continues to be one of the best practices for improving supply chain performance (Waller et al. 1999, Krichanchai & MacCarthy 2016). This research is motivated by a leading healthcare manufacturer who is interested in quantifying the benefits of CRP relationships with healthcare distributors in order to select the most profitable partners for their continuous replenishment programs and to set fair profit sharing contractual terms. The distributors are separate entities and are not internal parts of the manufacturer's organization. In a typical healthcare supply chain, distributors are supplied directly from manufacturers and hospitals are supplied directly from distributors.

The initial motivation behind collaboration in supply chains was reducing the bullwhip effect, which reduces the required amount of inventory across the chain (Lee et al. 1997). In addition, industrial practices and academic research revealed other benefits of CRP that are available in transportation and ordering costs (Disney et al. 2003, Zhang et al. 2007, Van der Vlist et al. 2007). However, implementing a CRP between two partners requires significant investments for the technological requirements of the arrangement. The partners need to have

Warehouse Management Systems (WMS) and Electronic Data Interchange (EDI) capabilities to share accurate demand and inventory information on a daily basis (Angulo et al. 2004). In a complicated supply chain (SC), in which the relationship between two partners could easily grow into a network with a few hundred lanes and thousands of SKU's, the decision of entering a CRP relationship or terminating one can have substantial cost and service impacts. As discussed in Sabath & Fontanella (2002), SCC has been difficult to implement and part of the problem has been the failure in differentiating the most profitable customers from the rest. Furthermore, a successful CRP relationship should include a fair revenue or profit sharing contact that incentivizes both partners to work together to optimize the entire SC (Giannoccaro & Pontrandolfo 2004, Cachon & Lariviere 2005, Li et al. 2015). The first step to set a fair and motivating contract is to have an accurate estimation of future cost savings.

This paper presents a cost approximation model with service level constraints under stationary stochastic demand. Service levels to end customers (i.e. hospitals) and to a distributor are measured by the probability of meeting the demand immediately from stock (i.e. type 1 service measure). We considered a two-echelon serial supply chain in which a product is shipped from a manufacturer distribution center (DC) to a distributor DC. A CRP between a manufacturer and distributor normally contains a network of various demand and supply locations (nodes) that are connected through channels (links). This model is able to accurately estimate the cost savings at the channel level, location level, and network level. The cost savings are computed for the inventory holding, transportation and order processing cost components for both partners. We demonstrate the model on a case study in a healthcare supply chain. To analyze a multi-product system, we developed the concept of a "standardized item" as a representation of the product mix on each channel. Therefore, the present paper considers a problem setting of a manufacturer that supplies a distributor with a single product (standardized item). To summarize, the model answers the following fundamental questions about a CRP relationship: *1) What is the cost savings impact of CRP on each partner? 2) How does it vary across a distribution network? 3) How does it vary across different cost components?*

The remainder of this paper is organized as follows. The next section reviews the related literature on SCC modeling with emphasis on the impact of CRP on the total cost of SC. Section 3 provides the assumptions of the model and an overview of the model structure. Sections 4 through 6 discuss the cost components of the model and impact of CRP on each. Section 7 presents the results of the case study, followed by the final section that presents conclusions and future work.

2.2 Literature Review

SCC models started with the emergence of Quick Response (QR) in late 1980s and early 1990s when J.C. Penny implemented this model in the apparel industry to shorten the long lead times (Iyer & Bergen 1997). During the same time frame, Walmart introduced the Retail Link platform that connects suppliers with the end customer demand data (Stalk et al. 1992). The next version of SCC model was introduced in the grocery industry as Efficient Consumer Response (ECR). Spartan Stores (Schiano & McKenney 1996), HEB (Clark et al. 1994), Campbell Soup (Cachon & Fisher 1997) and Proctor and Gamble (Keh & Park 1997) were pioneers in implementing ECR strategies (Sahin & Robinson 2002). The main purpose of ECR was to improve the responsiveness of the SC to consumer demand in the grocery sector. The largest reported benefits under ECR have come from Continuous Replenishment Programs (CRP), which were eveloped as a new mechanism for managing the flow of information and products between a supplier and a group of retailers (Cachon & Fisher 1997). CRP accounted for 38% of the total grocery industry ECR savings, and since then CRP has been implemented in many other SC sectors often under the name of Vendor Managed Inventory (Waller et al. 1999, Keh & Park 1997). CRP rearranges the conventional system of ordering and replenishment characterized by the transfer of purchase orders from the distributor to the manufacturer. CRP is a process of restocking where the manufacturer ships to the distribution center full loads that are supposed to satisfy a prearranged level of stock (Derrouiche et al., 2008). Oftentimes in CRP, the distributor is responsible for providing demand forecasts to the manufacturer, while in VMI the manufacturer generates forecasts using the demand data shared by the distributor. In VMI, the manufacturer is the primary decision-maker in order placement and inventory control, by determining the appropriate inventory levels of each of the products (within agreed upon bounds) and the appropriate inventory policies to maintain the levels (Derrouiche et al., 2008). In summary, CRP is a relationship with a more balanced distribution of power between a vendor and a retailer but VMI shifts the power more to the vendor. Finally, the most advanced form of SCC is Collaborative Planning, Forecasting and Replenishment (CPFR) where all members of a SC jointly develop demand forecasts, production plans, and inventory replenishments. The conditions that rationalize upgrading from CRP to CPFR are investigated by Sari (2008) and Kamalapur et al. (2013). The focus of this study is technically on CRP but since VMI and CRP are often used interchangeably in both academia and industry, the literature review covers both types of collaboration programs.

In positioning this paper, the broad framework proposed by Torres et al. (2014) is very useful. They categorize published academic literature on VMI into five groups of: *strategic*, *statistical characterization*, *analytical modeling* (deterministic and stochastic), *simulation,* and *game theory* (Table 1). This paper fits into the third group, where VMI arrangements are modeled by either deterministic or stochastic approaches. For completeness, we review representative papers in two groups most relevant to this paper: analytical modeling and simulation modeling. Another classification of VMI literature is presented in Govindan (2013), which suggests six dimensions of: *inventory*, *transportation*, *manufacturing*, *general benefits*, *coordination/collaboration*, and *information sharing* (Table 1).

Table 1: Two classifications of published academic literature on VMI based on Torres et al. (2014) and Govindan (2013)

#	Research methodology		#	Problem focus
1	Strategy		1	Inventory
2	Statistical Characterization		2	Transportation
3	Analytical Modeling		3	Manufacturing
4	Simulation		4	General benefits
5	Game Theory and Contracts		5	Coordination/Collaboration
			6	Information Sharing

In this section, we review the literature by combining the two classification schemes shown in Table 1. Govindan (2013) classifies the VMI literature based on problem focus and Torres et al. (2014) categorizes them based on research methodology. Figure 1 presents a Venn diagram to visualize the position of papers within the problem focus areas. The research methodology used in each paper is indicated by a number next to each paper. To further clarify the contribution areas, we add *order processing* as a new dimension to the dimensions proposed in Govindan (2013). Order processing includes activities in both ends of a CRP or VMI program such as picking, packing, loading, unloading, receipt verification, sorting, and putting away. As illustrated in Figure 1, this paper is located at the intersection of inventory, transportation, and order processing where analytical and simulation models have been proposed to study the impacts of VMI arrangements.

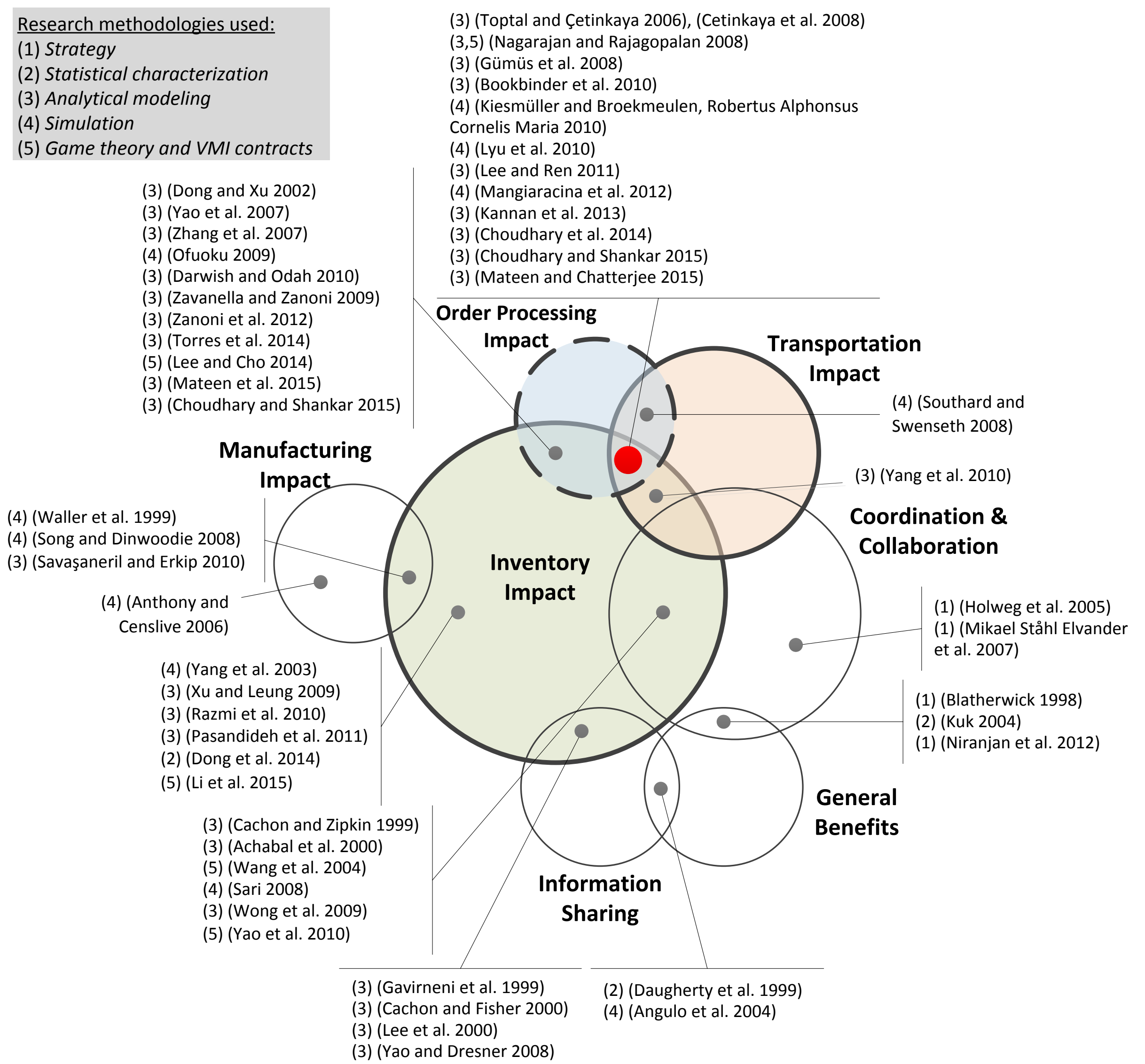

Figure 1: A graphical representation of published literature on VMI based on focus area and methodology

We remark that the papers that share the same spot in the literature with this paper have not investigated the true impact of CRP on transportation and order processing costs. Almost all of these papers assume constant ($/item) transportation and order processing costs in their models. The constant rate cost structures overlook the more complicated cost structures of FTL, LTL, and parcel transportation modes. The same is true for order processing cost because handling activities at both vendor and retailer are driven by shipment size and variety of products to process. Thus, a constant rate cost structure does not capture the impact of CRP on this cost component. In addition, a vendor normally uses a combination of the three transportation modes to transport the demand to a retailer; however, the developed models in the literature assume a single type of truck (i.e. full truckload) as the only means of transportation. In summary, the literature includes simplified models that show the shipment consolidation opportunities of CRP and indicate the associated transportation and order processing benefits without proposing a representative model that considers realistic cost structures.

Toptal & Cetinkaya (2006) study the benefits of VMI in transportation by using transport vehicles (i.e. having full truckloads) in a single-period setting. The paper studies a VMI relationship between a vendor and a buyer with stochastic demand and a single item with a short life cycle. Transportation cost is modeled as a combination of a fixed cost rate component, which represents ordering/replenishment cost, and a step-wise cost component, which represents the truck cost. A single type of truck (i.e. full truckload) is assumed as the only means of transportation. The results indicate that the vendor's expected profit is not necessarily increasing in buyer's order quantity, which is against the traditional belief on economies of scale in VMI. This work is extended in Cetinkaya et al. (2008) by modeling the vendor's demand as a stochastic bulk arrival process. Assuming a fixed cost of ordering, dispatching and transportation, and using cost approximation expressions, they show that the use of quantity based consolidation policies for outbound shipments can result in up to 57% cost savings.

Gumus et al. (2008) examine the benefits of Consignment Inventory (CI) for a single item system with deterministic demand using a joint economic lot sizing framework. The total cost function includes three cost components of inventory, ordering and transportation while ordering and transportation costs are assumed to be fixed. Bookbinder et al. (2010) develop a deterministic demand model to quantify the benefits of VMI in comparison with a regular independent decision-making system and a central decision-making system (i.e., both vendor and customer are members of the same corporate entity). The assumption is one vendor supplies a single product to one customer and similar to Gumus et al. (2008), ordering/ material handling and transportation costs are assumed to be fixed. Under a constant demand rate, the paper identifies the conditions that develop three possible outcomes of VMI, named as efficient system (both partners benefit), potentially-efficient system (total cost decreases but only one partner benefits), and inefficient system (system's total cost increases). Kannan et al. (2013) further extend this work to an analysis of one-vendor, multiple-buyer, and stochastic demand, motivated by a pharmaceutical case study. The paper provides useful insights on cost savings by making assumptions about possible impacts of VMI. Similar to many studies, ordering, handling, and transportation costs are assumed to be independent of shipment size.

Kiesmüller & Broekmeulen, Robertus Alphonsus Cornelis Maria (2010) consider a multi-product serial two echelon inventory system with stochastic demand and use a simulation study to determine the benefit of VMI from economies of scale in order picking activities. Order picking cost is modeled with two components: i) a fixed ordering cost per order line and ii) traveling distance of order picker in warehouse. They assume VMI enables the vendor to enlarge the preferred order quantities of the retailer to benefit from economies of scale and increase the utilization of transportation trucks and order picker. Similar to Toptal & Cetinkaya (2006) they assume a single type of truck as the only means of transportation. The results show if inventory holding costs are relatively low compared to the handling and transportation costs, a reduction in the number of order lines can reduce the total cost at the vendor. This study did not include the handling cost at the retailer side and is limited to a specific warehouse layout.

In two different case studies, Lyu et al. (2010) and Mangiaracina et al. (2012) collaborated with international grocery manufacturers and retailers to quantify the value of VMI in different cost components. The critical impact of VMI is enabling the manufacturer to plan the replenishment for optimizing transportation using a multi-stop policy. The results indicate that manufacturer's benefits are always greater but retailers have higher savings percentage rates.

Lee and Ren (2011), proposed a periodic-review stochastic inventory model to examine the benefits of VMI in a global environment. The paper uses a (s, S) policy for the supplier and shows that VMI provides an opportunity for the supplier to achieve economies of scale in production and delivery. Production/delivery cost is modeled together with a fixed component and a variable component.

Choudhary et al. (2014) compared the value of VMI with the value of Information Sharing via a two-echelon analytical model that assumes the transaction of a single item between a supplier and a retailer under time varying deterministic demand. Fixed ordering, setup, handling, and transportation costs are assumed with negligible lead times. The results emphasize the importance of transportation savings in VMI by indicating that when handling and transportation costs are negligible the cost benefits of shifting from IS to VMI is not significant. Mateen & Chatterjee (2015) developed a similar model with the same assumptions for one vendor and multiple retailers with a focus on modeling the transportation savings using an efficiency factor. Choudhary & Chatterjee (2015) extended the previous model by considering a non-stationary stochastic demand process and multiple retailers. However, the assumptions for ordering, handling and transportation costs remained unchanged.

2.3 Model Assumptions and Structure

The model described in this paper estimates the cost of inventory holding, order processing and transportation for each SC channel separately. A channel is a pairing of a manufacturer distribution center (DC) and a distributor DC. The difference between one channel and another could be significant due to a variety of factors such as product mix, transportation requirements, demand characteristics, etc. The followings are the critical assumptions of the model along with the logic behind using them:

- We considered a two-echelon serial supply chain in which a product is shipped from a manufacturer DC to a distributor DC. The concept of "standardized item" is developed to analyze a multi-item system in the case study.
- Manufacturer and distributor are two separate entities and are not internal parts of each other's organizations.
- This paper does not investigate the impact of CRP on the manufacturing plant level, which in fact can be very significant. The focus of the model is on the cost of distribution (i.e. inventory holding, transportation, and order processing) from a manufacturing DC to a distributor DC under service level requirements (Figure 3).
- The demand process at the distributor DC and manufacturer DC is modeled as Poisson process and compound Poisson process respectively. Usage of Poisson process for demand modeling is very common, especially when the underlying arrival process is unknown, and is used numerously in the context of SCC modeling (Cheung & Lee 2002, Lyu et al. 2010). Stationary stochastic process is very suitable for modeling the demand in CRP relationships because as Krichanchai & MacCarthy (2016), Mateen & Chatterjee (2015), and Niranjan et al. (2012) suggest, the most suited items for CRP have stable and

low variability demand. In the case of non-stationary demand, our model can be used as a piece-wise stationary function for approximation.

- All the cost components of the model are shared between the manufacturer and distributor excluding the transportation cost which is the responsibility of manufacturer only. This assumption is used by many papers including Mangiaracina et al. (2012) and Kannan et al. (2013). The main reason is the fact that CRP requires the manufacturer to take the responsibility of replenishment and since replenishment pattern directly affects the transportation cost, CRP partners oftentimes agree on transferring the cost of transportation to the manufacturer.
- Transportation cost contains three components of FTL, LTL and Parcel costs. FTL unit cost is $/mile, while LTL and Parcel costs are on a per load basis using a cost structure based on distance and weight. The determination of transportation mode in the model is based on the demand size (i.e. combination of volume and weight).
- Order processing cost contains two components of order generation cost and handling cost. Order generation cost component is assumed to be fixed ($/order), which implies an independency from order size. This is a realistic assumption because orders are normally generated electronically and automatically based on inventory levels and replenishment parameters with minimal manual labor intervention. In contrast, the handling cost component, which includes picking, packing, loading, unloading, receipt verification, sorting, and putting away, heavily involves manual labor and is greatly dependent on the order size and the variety of items on a purchase order. Therefore, handling cost is assumed to be variable and a function of order size and variety of items.

The model estimates the SC cost through a historical demand data analysis, which provides the basic demand characteristics of a channel such as product mix, demand variability, and order frequency. Figure 2 illustrates the model structure and its major components. Inventory holding cost of each channel is estimated using a (r, Q) system and channel-specific inventory policy parameters. The transportation cost of each channel is estimated by selecting the appropriate transportation mode (i.e. FTL, LTL or Parcel) using the demand characteristics, and channel specific transportation data such as rate, distance, order frequency, product mix and item dimensions. Order processing cost, which includes both order generation cost and handling cost, is driven by the ordering frequency of each channel. The impact of CRP on each cost component will be discussed in the following sections.

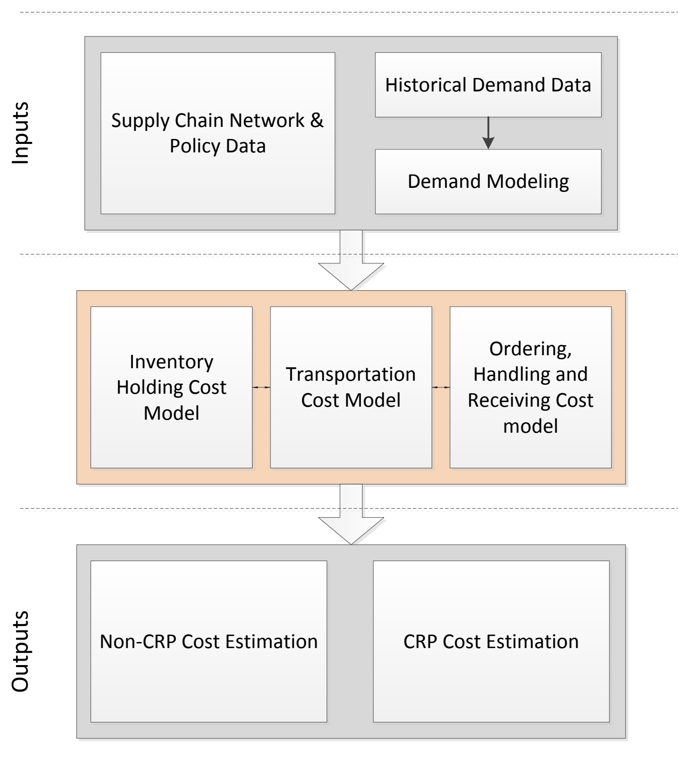

Figure 2: Model structure

2.4 Inventory Holding Cost

The first challenge of inventory holding cost modeling is the fact that thousands of different SKUs are regularly being ordered by the distributors while each SKU's demand has a different underlying stochastic process, different unit cost and possibly different holding charge. The problem becomes even more challenging because each shipment (e.g., a full truckload) contains several different SKUs and each SKU has different characteristics such as unit weight and volume that are important in transportation cost modeling. Thus, we simplify the problem in such a way that does not hurt the accuracy of cost approximations but enables the efficient analysis of system costs. We focused on the "significant few" instead of the "trivial many" items as a well-established strategy to cope with the size of the problem. An analysis of the demand data on all channels revealed that the Pareto rule clearly defines the demand pattern of items for each channel. As a general rule among all channels, the top 20% of items in terms of monetary value (demand × price) make up 80% of the volume, weight and monetary value of the shipped items on a channel. Therefore, we reduced the size of the demand data for each channel to only the top items that represent at least 80% of the volume, weight and monetary value of the channel demand. We call this set of important items that is different for each channel the *"standardized item set"* which is representative of the product mix for each channel. In the demand and cost modeling, the characteristics of the standardized item set (e.g., unit price, unit volume, and unit weight) are considered as the weighted average characteristics of the entire set. The mathematical illustration of standardized item set and its characteristics are as follows:

$d_{i,t}$ demand of item i in period t

p_i unit price ($) of item i

w_i unit weight (lb) of item i

v_i unit volume (ft^3) of item i

N set of all items ordered in period T

M standardized item set

P weighted average unit price ($) of the standardized item set

W weighted average unit weight (lb) of the standardized item set

V weighted average unit volume (ft^3) of the standardized item set

Note that t and T should be appropriately selected based on the planning strategies. In the case study section (section 2.7), t and T are assumed to be a week and a year. T should be sufficiently long (e.g., a year) to represent the real demand of the channel. The relationship between the total demand of a channel and the demand of the standardized item set is defined as follows:

$$\sum_{t=1}^{T}\sum_{i\in M} p_i \times d_{i,t} \geq 0.8 \times \sum_{t=1}^{T}\sum_{j\in N} p_j \times d_{j,t} \tag{1}$$

$$\sum_{t=1}^{T}\sum_{i\in M} w_i \times d_{i,t} \geq 0.8 \times \sum_{t=1}^{T}\sum_{j\in N} w_j \times d_{j,t} \tag{2}$$

$$\sum_{t=1}^{T}\sum_{i\in M} v_i \times d_{i,t} \geq 0.8 \times \sum_{t=1}^{T}\sum_{j\in N} v_j \times d_{j,t} \tag{3}$$

The characteristics of the standardized item set for each channel are computed as follows:

$$P = \frac{\sum_{t=1}^{T}\sum_{i\in M} p_i d_{i,t}}{\sum_{t=1}^{T}\sum_{i\in M} d_{i,t}}, \qquad W = \frac{\sum_{t=1}^{T}\sum_{i\in M} w_i d_{i,t}}{\sum_{t=1}^{T}\sum_{i\in M} d_{i,t}}, \qquad V = \frac{\sum_{t=1}^{T}\sum_{i\in M} v_i d_{i,t}}{\sum_{t=1}^{T}\sum_{i\in M} d_{i,t}} \tag{4}$$

2.4.1 Demand Modeling

Figure 3 illustrates the dynamics of ordering between the echelons of a supply chain in both CRP and non-CRP relationships. In a non-CRP relationship, the manufacturer DC does not have visibility of the end customer demand information and only receives the ordering information from the distributor DC. In a CRP relationship, the manufacturer DC replenishes the distributor DC automatically and on a regular basis without receiving any order information. In return, the manufacturer DC ensures an agreed upon service level and inventory level over time at the distributor DC. To enable such an arrangement, they share demand, forecast, and inventory level information on a regular basis (normally daily) through EDI transactions.

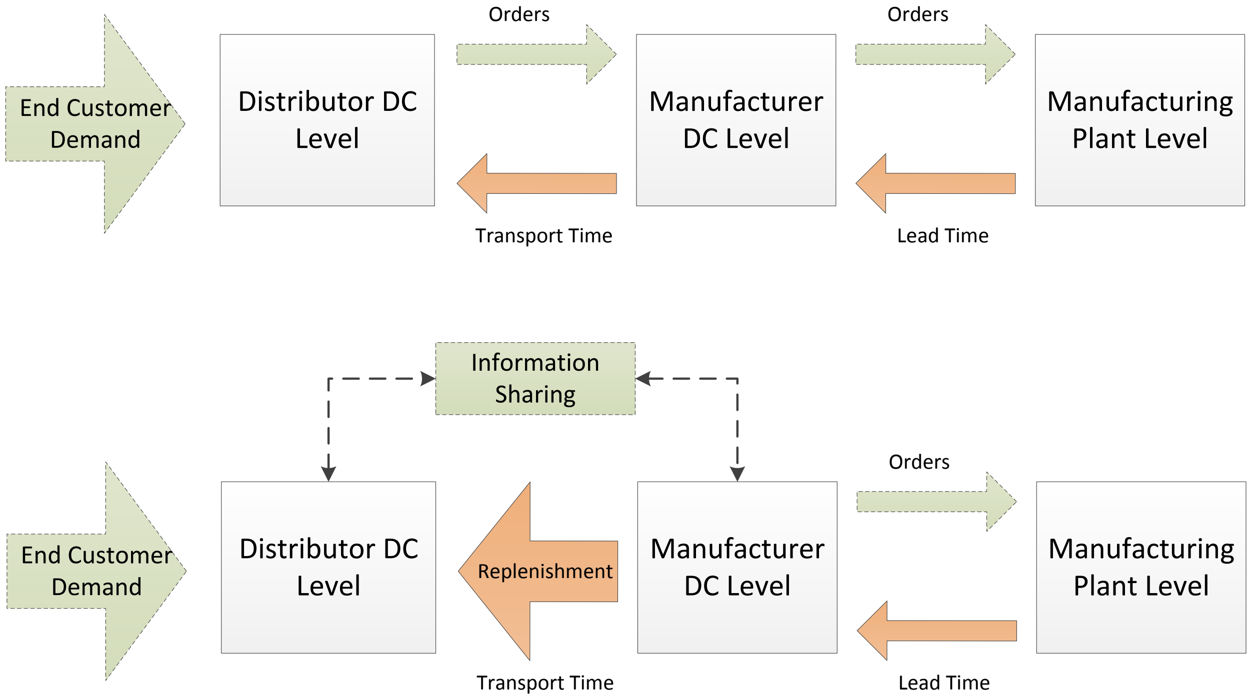

Figure 3: Ordering dynamics in non-CRP (top diagram) and CRP (bottom diagram). Dashed arrows indicate information flow and solid arrows indicate physical flow

In order to estimate the demand parameters that a distributor realizes from end customers, we study the orders that the distributor sends upstream to the manufacturer. The historical order data of the standardized item set for each channel contains the required information about the orders that a distributor DC sends to a manufacturer DC. The orders are realized over time for the entire standardized item set in order to compute order frequency, size and variance of the orders associated with the distributor. The sequence of standardized orders is mathematically defined in Equation (5) where $d_{i,t}$ is the amount of item *i* ordered in period *t.*

$$D_t = \sum_{i \in M} d_{i,t} \quad \text{for t = 1, … , T} \tag{5}$$

Equation (5) indicates that the demand at each period *t* (D_t) is realized as the accumulation of the amounts ordered for all the items in period *t* (i.e. items that are identified in the standardized item set).

$\overline{OF}_D$	the average order frequency of a distributor DC to a manufacturer DC (i.e. average time between positive demand occurrences $D_t > 0$ over period T).
$E[D_t]$	the mean order size (i.e. average of positive demand values $D_t > 0$ over period T).
$Var[D_t]$	the variance of order size (i.e. variance of positive demand values $D_t > 0$ over period T).
μ_c	the mean demand rate at the distributor DC level
σ_c^2	the variance of demand rate at the distributor DC level
μ_w	the mean demand rate at the manufacturer DC level
σ_w^2	the variance of demand rate at the manufacturer DC level

The arrival rate of demand is assumed to follow a Poisson process with rate λ_c, where $\lambda_c = \overline{OF}_D$, which means the time between demand occurrences are independent and follow an exponential distribution with parameter $1/\lambda_c$. The mean and variance of the demand rate (units/time) is derived as $\mu_c = \sigma_c^2 = \overline{OF}_D \times E[D_t]$. In order to clarify with an example, if $\overline{OF}_D$ is 3 times per week and $E[D_t]$ is 300 units then the weekly mean and variance of demand is 900 units.

The demand at the manufacturer level is a complicated stochastic process that depends on the distribution of the time between replenishments from a manufacturer DC to a distributor DC, and the size and variance of the replenishment orders. Here we assume the demand at the manufacturer level also follows a compound Poisson process. The basis for this assumption is appropriate in many situations because the distributor DC receives the end customer demand on a regular basis ($\overline{OF}_D$) with certain size and variance and they just transfer the demand to the upstream echelon with much more variability due to the bullwhip effect. The constructed compound Poisson process for the manufacturer level has the same arrival rate (i.e. $\lambda_w = \overline{OF}_D$) with the following mean and variance of demand rate (units/time):

$$\mu_w = \lambda_w \times E[D_t] = \overline{OF}_D \times E[D_t] = \mu_c \tag{6}$$

$$\sigma_w^2 = \lambda_w \times [Var(D_t) + (E[D_t])^2] = \overline{OF}_D \times [Var(D_t) + (E[D_t])^2] \tag{7}$$

The last piece of demand modeling is to represent the demand during lead time at the manufacturer DC. This lead time represents the time required for the manufacturing plant to resupply the manufacturer DC.

L_w	lead time from the manufacturing plant to manufacturing DC
$D(L_w)$	demand during lead time at manufacturing DC

according to Svoronos & Zipkin (1988) and assuming that:

$$D(L_w) \sim Gamma\ (E[D(L_w)], Var[D(L_w)]) \tag{8}$$

we can approximate the following:

$$E[D(L_w)] \cong \overline{OF}_D \times E[D_t] \times E[L_w] \tag{9}$$

$$Var[D(L_w)] \cong E[L_w] \times \sigma_w^2 + \mu_w^2 \times Var[L_w] \tag{10}$$

The reason behind approximating $D(L_w)$ using the Gamma distribution is the robustness of this distribution in accurate estimation of demand during lead time when the actual distribution is unknown (Rossetti & Ünlü 2011). In the next section, the inventory performance metrics at both levels of distributor and manufacturer are discussed.

2.4.2 Manufacturer Inventory Holding Cost & Performance

In order to quantify the performance of inventory management at the manufacturer, the following relevant inventory metrics are considered:

$\bar{I}_w$ the average inventory level at the manufacturer DC.

$\bar{B}_w$ the average back order level at the manufacturer DC.

$\overline{RR}_w$ the average ready rate at the manufacturer DC

The inventory systems at both levels of manufacturer and distributor are modeled assuming a (r, Q) system. r_w and q_w denotes the reorder point and reorder quantity at the manufacturer DC.

$$r_w = E[D(L_w)] + N^{-1}(\tau_w) \times \sqrt{Var[D(L_w)]} \tag{11}$$

$$q_w = \sqrt{\frac{2 \times k_w \times \mu_w}{\tau_w \times h}} \tag{12}$$

In Equation (11) & (12), k_w and τ_w denote the ordering cost and the required service level at the manufacturer DC, respectively. Equation (12), which computes q_w subject to a fill rate constraint (τ_w), is an approximation of a lower bound that is obtained in Agrawal & Seshadri (2000) and a heuristic approach discussed on page 226 of Zipkin (2000). A safety factor requirement is normally set above 90%. In the case study, τ_w is set to be 98%. $N^{-1}(\tau_w)$ denotes the safety stock factor, which is the inverse of the standard normal cumulative distribution with in stock probability of τ_w. Having r_w and q_w computed, the performance metrics at the manufacturer DC are calculated as follows (Zipkin 2000):

$$\bar{I}_w(r_w, q_w) = \frac{1}{2}q_w + r_w - E[D(L_w)] + \bar{B}_w(r_w, q_w), \tag{13}$$

$$\bar{B}_w(r_w, q_w) = \frac{1}{q_w}[F^2(r_w) - F^2(r_w + q_w)], \tag{14}$$

$$\overline{RR}_w(r_w, q_w) = 1 - \frac{1}{q_w}[F^1(r_w) - F^1(r_w + q_w)], \tag{15}$$

where F^1 and F^2 are the first and second order loss functions of the demand during lead time distribution (assumed to be Gamma here). The total holding cost at the manufacturer DC (Equation (17)) can be computed using the average inventory estimation ($\bar{I}_w$) and unit holding cost (h_w). While $\bar{I}_w$ is already computed (Equation (13)), unit holding cost can be computed by multiplying the manufacturer holding charge (i_w) to the standardized item cost (P) (Equation (16)).

$$h_w = i_w P \tag{16}$$

$$HC_w = h_w \bar{I}_w \tag{17}$$

In-transit inventory cost, which usually is part of the manufacturer total cost, should be computed. First, average in-transit inventory level should be estimated by multiplying the mean demand rate to the transportation time (i.e., T_c) (Equation (18)). Then, the cost of in-transit inventory can be computed as shown in Equation (19):

$$\bar{I}_\eta = E[D_t] \times T_c \tag{18}$$

$$HC_\eta = h_\eta \bar{I}_\eta \tag{19}$$

2.4.3 Distributor Inventory Holding Cost & Performance

In order to calculate the same performance metrics for the distributor DC, we need to first model the demand during lead time at the distributor level ($D(L_c)$). The lead time at the distributor level contains two components: order processing/transportation time from the manufacturer DC to the distributor DC (i.e., T_c) and the back order waiting time (T_{BW}).

$$L_c = T_c + T_{BW} \tag{20}$$

$$E[L_c] = E[T_c + T_{BW}] = E[T_c] + E[T_{BW}] \tag{21}$$

$$Var[L_c] = Var[T_c + T_{BW}] = Var[T_c] + Var[T_{BW}] \tag{22}$$

The expected value and variance of transport time ($E[T_c]$, $Var[T_c]$) depend on the carrier and order processing time. However, the expected value of back order waiting time is computed as follows:

$$E[T_{BW}] = \frac{\bar{B}_w(r_w, q_w)}{\lambda_w} \tag{23}$$

Using the approximation provided in Hopp & Spearman (2011) on page 619 we can also approximate the variance of back order waiting time as follows:

$$Var(T_{BW}) \cong \frac{1 - \overline{RR_w}}{\overline{RR_w}} \times (E[T_{BW}])^2 \tag{24}$$

The standard approach to model the demand during lead time is to fit a distribution to the mean and variance of the demand during lead time.

$$E[D(L_c)] = E[L_c] \times E[D_t] \tag{25}$$

$$Var[D(L_c)] = E[L_c] \times \sigma_c^2 + (E[D_t])^2 \times Var(L_c) \tag{26}$$

Now, r_c and q_c which denote the reorder point and reorder quantity at the distributor DC can be estimated:

$$r_c = \; E[D(L_c)] + N^{-1}(\tau_c) \times \sqrt{Var[D(L_c)]} \tag{27}$$

$$q_c = E[D_t] \tag{28}$$

In Equation (27), τ_c represents the required service level at the distributor DC. Knowing r_w and q_w, the performance metrics at the distributor DC such as average inventory level $\bar{I}_c(r_c, q_c)$, average back order level, $\bar{B}_c(r_c, q_c)$, and average ready rate $\overline{RR}_c(r_c, q_c)$ can be computed using the same equations used for the manufacturer DC (Equations (13), (14)& (15)). The only difference is that the parameters of the gamma distribution for computing first and second order loss functions are $E[D(L_c)]$ and $Var[D(L_c)]$ here. Once the metrics are computed, the total holding cost at the distributor DC (Equation (30)) can be computed using the average inventory estimation ($\bar{I}_c$) and unit holding cost (h_c):

$$h_c = i_c P \tag{29}$$

$$HC_c = h_c \bar{I}_c \tag{30}$$

2.4.4 Impact of CRP

Supply chain coordination programs such as CRP align different stages of the supply chain by exchanging useful and accurate information between them and allowing the manufacturer to control the flow of products throughout the supply chain. This collaboration among players improves the accuracy of forecasting, reduces lead times, and ultimately reduces the variability of demand in the supply chain and is known as the Bullwhip Effect (Lee et al. 1997).

The first impact of CRP is the reduction in the mean order lead time and its variance ($E[L_c]$& $Var[L_c]$ in Equation (21) & (22)) which result in a reduction of the inventory levels at both distributor's DC and in-transit. CRP essentially enables the distributor to reduce their own inventory by lowering the reorder point (Equation (27)) since the mean and variance of demand during lead time is reduced. The reorder point protects the distributor against the variability of demand during lead time, and CRP justifies a lower reorder point that is sufficient for maintaining the fill rate. Obviously, in-transit inventory will be reduced since $E[L_c]$ is reduced, which means the transportation period is shorter.

The second impact of CRP is the reduction in the required level of inventory at the manufacturer DC. This is achieved through lower variability of demand that manufacturer DC typically realizes in CRP. As Equation (13) shows, inventory level at the manufacturer DC (i.e. $\bar{I}_w(r_w, q_w)$) is depended on the reorder point (r_w), which is itself dependent on the variance of demand during lead time (Equation (11)). In addition, as Equation (10) indicates, the variance of demand during lead time (i.e. $Var[D(L_w)]$) is dependent on the variance of demand at the manufacturer DC level (i.e. σ_c^2). Finally, as Equation (7) indicates, σ_c^2 is dependent on the variance of order sizes coming from the distributor DC (i.e. $Var(D_t)$). Therefore, in the model $Var(D_t)$ is the root cause of the required inventory level for satisfying the fill rate. Although in CRP, the manufacturer DC does not receive orders from downstream and has a full control of replenishment process, $Var(D_t)$ is still a good indicator of reduction in σ_c^2.

2.5 Transportation Cost

This cost component represents the cost that the carrier charges the manufacturer to move the freight from the manufacturer's DC to the distributor's DC. We assumed that the freight is shipped directly from the manufacturer's DC to the customer's DC via ground transportation modes.

2.5.1 Cost Estimation

The first critical parameter in estimating the transportation cost on each channel is the size of shipments on the channel, which is a function of the customer's demand. Another important parameter is average shipment frequency, which normally has an indirect relationship with shipment size. In general, the combination of average shipment frequency and average shipment size should satisfy the demand over time. Transportation cost normally depends on the volume or the weight of shipments. The following equations estimate the volume and weight of the standardized orders for a channel. Subscript "*ij*" represents a channel from a manufacturer's DC *i* to a distributor's DC *j*.

$$\hat{S}_{ij}^{v} = \hat{E}[D_t]_{ij} \times V_{ij} \tag{31}$$

$$\hat{S}_{ij}^{w} = \hat{E}[D_t]_{ij} \times W_{ij} \tag{32}$$

These two parameters provide sufficient information on the size of each order, which is required for selecting a cost-effective shipping mode to transport the freight on a channel. Shipment frequency depends on what shipping mode will be selected to transport the demand. Selecting a more consolidated mode will result in less shipment frequency while utilizing the smaller shipping modes requires making frequent shipments. There are three shipping modes available for the manufacturer to choose from: full truckload (FTL), less than truckload (LTL) and parcel. While FTL is the most cost-effective shipping mode, it is not the best choice for low demand channels. This is also true for selecting a shipping mode between LTL and parcel. At this stage, we estimate the transportation cost of one standardized order considering the most cost-effective manner. Later in Section 2.5.2, we will adjust the cost estimation based on the ordering frequency of the channel and the historical transportation performance.

Depending on the shipping mode, carriers charge their customers using different rules and in return, customers set certain shipping mode selection rules to select the most cost-effective mode. These rules normally have lower and upper limits. In this case study, the manufacturer determined volume limits for declaring a shipment as an FTL. Any standardized order that has volume size ($\hat{S}_{ij}^{v}$) below the lower limit will be shipped via LTL except for those that are less than 150 lbs. which are considered small package and will be shipped via parcel. In addition, there is a maximum weight limit for FTL shipments, which is 45000 lbs. in most of the states in the U.S.

V_{min}^{FTL} lower volume limit for FTL shipments (ft^3)

V_{max}^{FTL} upper volume limit for FTL shipments (ft^3)

W_{max}^{FTL} upper limit for the weight of a FTL shipment (lb)

ω_{ij} distance from channel $i-j$ (mile)

R_{ij}^{FTL} FTL rate on channel $i-j$ (\$/mile)

R_{ij}^{LTL} LTL rate on channel $i-j$ (\$/lb)

R_{ij}^{Parcel} parcel rate on channel $i-j$ (\$/lb)

TC_{ij} transportation cost estimate for channel $i-j$ (\$)

ATC_{ij} adjusted transportation cost estimate for channel $i-j$ (\$)

The number of required shipments for each standardized order on any channel can be calculated using both weight and volume limits:

$$N_{FTL}^{W} = \left\lfloor \frac{\hat{S}_{ij}^{w}}{W_{max}^{FTL}} \right\rfloor \tag{33}$$

$$N_{FTL}^{V} = \left\lfloor \frac{\hat{S}_{ij}^{v}}{V_{min}^{FTL}} \right\rfloor \tag{34}$$

N_{FTL}^{W} and $N_{FTL}^{V}$are the required number of FTL shipments considering the weight limit and volume limit, respectively. Based on the characteristics of the standardized item set (unit weight and unit volume), the truck may either exceed the volume limit or weight limit first, which would result in different values for N_{FTL}^{W} and N_{FTL}^{V}. In such case, the required number of FTL shipments should be the larger value:

$$N_{ij}^{FTL} = Max\,(N_{FTL}^{W}\,, N_{FTL}^{V}) \tag{35}$$

If N_{ij}^{FTL} becomes zero for a channel, this means that the demand does not support FTL shipments. Therefore, LTL or/and parcel should be used for transportation. Parcel shipments normally have the maximum weight limit of 150 lbs. and this is a common rule across different industries (i.e. $W_{max}^{Parcel} = 150$ lbs.) If $N_{ij}^{FTL} \geq 1$, then one or multiple FTL shipments should be sent to satisfy the demand; however, there is a remainder that needs to be shipped via non-FTL shipping modes. Let $\hat{s}_{ij}^{w}$ be the weight estimate of the remainder (lb) for channel $i - j$. If $\hat{s}_{ij}^{w} \leq W_{max}^{Parcel}$, the remainder will be shipped by a parcel shipment; otherwise, an LTL shipment will be dispatched. This means that $\hat{s}_{ij}^{w}$ can only be shipped via either LTL or parcel. Thus, an indicator function is used to reflect this fact in the total cost function:

$$I(ij) \quad = \quad \begin{cases} 1 & if\ LTL\ is\ selected\ for\ channel\ i-j \\ 0 & if\ Parcel\ is\ selected\ for\ channel\ i-j \end{cases}$$

Carriers normally charge the shippers on a per mile basis ($/mile) for FTL shipments and on a per pound basis ($/lb) for LTL and parcel shipments. The transportation cost functions for each standardized order on channel $i - j$ are calculated using Equations (36).

$$\begin{aligned} TC_{ij} &= TC_{ij}^{FTL} + I(ij) \cdot TC_{ij}^{LTL} + \big(1 - I(ij)\big) \cdot TC_{ij}^{Parcel} \\ &= N_{ij}^{FTL} \cdot \omega_{ij} \cdot R_{ij}^{FTL} + \hat{s}_{ij}^{w} \cdot (I(ij) \cdot R_{ij}^{LTL} + \big(1 - I(ij)\big) \\ &\quad \cdot R_{ij}^{Parcel}) \end{aligned} \tag{36}$$

2.5.2 Cost Adjustment

The cost estimation obtained from section 2.5.1 is based on an ideal situation where the demand is 100% certain and transportation mode selection is made based on a set of rules which will result in a very efficient transportation. In reality, factors such as order size variability, arrival pattern of orders over a week, and expedited shipping requests disturb the efficiency of mode selection process. Therefore, when we look at the actual shipping pattern on a channel we observe inefficiencies that involve excessive LTL and Parcel shipments. In this section, we take those inefficiencies into consideration by adjusting the estimation. In order to adjust the value of

TC_{ij} based on the historical transportation efficiency of channels, a transportation efficiency metric (C_{ij}) is constructed. Later in section 2.5.4, the impact of CRP on the transportation cost, which can be summarized as improved shipment consolidation, is modeled using the same metric. The detail of the metric is discussed in the second chapter of Parsa (2017). The metric captures the cost efficiency of transportation on each channel by incorporating the shipping rates of each mode (i.e. FTL, LTL, Parcel), space utilization of dispatched FTL trucks, the distance on the channel, and the weight limit of shipping trucks. Mateen & Chatterjee (2015) also proposed an efficiency factor to model the impact of CRP on transportation but without discussing the contributing elements of the factor. C_{ij} is the transportation efficiency score of channel $i-j$ and it could vary between 0 and 100, where zero represents a channel on which the entire demand is shipped via parcel, and 100 indicates a channel where the entire demand is shipped via fully space-utilized FTL trucks. The metric can be computed for channel $i-j$ over a certain period of time using the linear function in Equation (37), where W_{ij}^{FTL}, W_{ij}^{LTL}, W_{ij}^{Parcel} are efficiency weights (on the scale of 100) to represent the cost efficiency of each mode based on their associated shipping rates. To account for empty space within FTL trucks, average space utilization of FTL trucks ($\bar{u}_{ij}$), is multiplied to W_{ij}^{FTL}. Lastly, P_{ij}^{FTL}, P_{ij}^{LTL}, P_{ij}^{Parcel} are the portions of the total demand shipped via FTL, LTL and Parcel respectively over the same period of time.

$$C_{ij} = \bar{u}_{ij} \times W_{ij}^{FTL} P_{ij}^{FTL} + W_{ij}^{LTL} P_{ij}^{LTL} + W_{ij}^{Parcel} P_{ij}^{Parcel} \qquad (37)$$

This metric (C_{ij}) is significantly dependent on the demand volume because high demand channels have a better potential of consolidating the shipments and gaining higher C_{ij} values. This fact is shown in Figure 4 where the metric is computed for 143 non-CRP and CRP channels and the relationship between demand and C_{ij} values is plotted.

We remark that TC_{ij} is a cost figure for the highest "achievable" transportation efficiency level for channel $i - j$. In order to adjust TC_{ij} estimation, we need to find the associated metric value for the TC_{ij} estimation (let it be C_{ij}^{max}) and compare it with C_{ij} to understand how close the transportation on channel $i - j$ is performed to the highest achievable efficiency level. The difference between C_{ij} and C_{ij}^{max} is how much our initial estimation (i.e. TC_{ij}) needs to be adjusted. Now, the problem is to determine representative C_{ij}^{max} values for channels. Figure 4 shows that high demand channels generally have higher C_{ij} values while they might not perform at the highest achievable efficiency level. In contrast, low demand channels tend to have lower scores whereas they might perform at their best possible level. In other words, not all channels can achieve the efficiency level of 100. For instance, a channel with the weekly demand volume of 200 ft^3 is not able to support FTL shipments therefore; the efficiency score of 100 is not achievable and should not be considered as an appropriate C_{ij}^{max} value. To overcome this issue, we can categorize the channels to multiple demand size categories and determine a different highest achievable efficiency level (C_{ij}^{max}) for each category. This allows setting a fair and achievable transportation efficiency level for the channels because channels that share the same demand category have close demand volume; therefore they can share a common C_{ij}^{max} value.

Table 2 shows the demand categorization that we used for our case study, where five categories are considered. To determine C_{ij}^{max} value for each, we initially look at the top performers; however, any dataset contains outliers; for example, a channel that only shipped once or twice in the past six months due to an irregular rapid demand surge is not representative of the channels of the category. Thus, we considered the 3rd quartile of the efficiency scores in each category as an achievable level (i.e. C_{ij}^{max} is the 3rd quartile of black points in Figure 4). As

Figure 4 indicates, CRP channels tend to perform more efficient in transportation; therefore we compute different achievable transportation efficiency levels for them using the 3rd quartile of CRP scores in each category (i.e. C_{ij}^{CRP} is the 3rd quartile of yellow points in Figure 4).

Table 2: Channel categories based on weekly demand (Case study example)

Demand Category ($\boldsymbol{ft^3}$) (x-axis of Figure 4)	$\boldsymbol{C_{ij}^{max}}$	$\boldsymbol{C_{ij}^{CRP}}$
[0, 800)	16.91	20.27
[800 , 1600)	28.14	57.28
[1600 , 2400)	60.07	78.60
[2400 , 3200)	57.30	76.61
[3200 , 12000)	67.53	85.06

Once C_{ij} and C_{ij}^{max} are computed for channel $i-j$, an adjustment multiplier (ρ_{ij}) can be computed to adjust the total transportation cost estimation (TC_{ij}):

$$\rho_{ij} = \begin{cases} \dfrac{C_{ij}^{max}}{C_{ij}} & if\ C_{ij} < C_{ij}^{max} \\ 1 & Otherwise \end{cases} \tag{38}$$

Therefore, total adjusted transportation cost for channel $i-j$ will be:

$$ATC_{ij} = \overline{OF}_D^{ij} \times \rho_{ij} \times TC_{ij} \tag{39}$$

2.5.3 Validation

The last step in the transportation cost modeling is to validate the output of the model. The goal is to compare the cost estimation of the model with the historical transportation cost. Historical shipping data along with the shipping rates ($R_{ij}^{FTL}, R_{ij}^{LTL}, R_{ij}^{Parcel}$) and distances enable the calculation of the historical transportation cost for each channel. The model uses Table 2, which is constructed using the data of 143 channels, to adjust the estimation. However, in order to avoid bias in the validation process, a new set of 57 channels is selected that do not overlap with the 143 channels that are used for determining the C_{ij}^{max} values in Table 2.

Let HTC_{ijk} be the historical transportation cost for channel $i-j$ at week k.

Let $\overline{HTC}_{ij}$ be the average weekly historical transportation cost for channel $i-j$.

In the validation process, we compare the values of $\overline{HTC}_{ij}$ and ATC_{ij} for each channel. They both represent an estimate of weekly transportation cost for channel $i-j$. While $\overline{HTC}_{ij}$ is based on the historical data, ATC_{ij} is the output of the model. The validation metrics used in this section are:

- Error (E_{ij}) $= ATC_{ij} - \overline{HTC}_{ij}$
- Relative error (RE_{ij}) $= \frac{E_{ij}}{\overline{HTC}_{ij}}$

Table 3 also summarizes the model validation results numerically. Overall the model performs very good considering average relative error of -4% which is equivalent of $65.

Table 3: Validation metrics summary

Validation Metric	Value
Average Error ($\frac{\sum_1^{57} E_{ij}}{57}$)	-$65
Average Relative Error ($\frac{\sum_1^{57} RE_{ij}}{57}$)	-4%

2.5.4 Impact of CRP

So far in Section 2.5, we have constructed a modeling framework to compute the transportation cost of a channel in a supply chain. We have also validated the results of the model using empirical data. This section investigates the impact of a CRP relationship on transportation cost. As discussed in the literature review section, supply chain collaboration programs have shown its capability in consolidating the shipments, which results in reduction in the transportation cost (Cetinkaya et al. 2008, Southard & Swenseth 2008) The transportation efficiency metric that is used in Section 2.5.2 to adjust the transportation cost estimation is also used here to compare the performance of CRP with non-CRP in terms of transportation (Figure 4) and quantify the cost of transportation in CRP. CRP clearly improves transportation efficiency due to the ability of the manufacturer in controlling the replenishment process, consolidating the shipments, and utilizing more cost-effective transportation modes. The metric captures the shipping consolidation effect of CRP using two parameters: i) average space utilization of trucks ($\bar{u}_{ij}$) and ii) usage rates of transportation modes (P_{ij}^{FTL}, P_{ij}^{LTL}, P_{ij}^{Parcel}). CRP generally shows higher $\bar{u}_{ij}$ values across the channels and higher usage rate of more consolidated modes ($P_{ij}^{FTL} > P_{ij}^{LTL} > P_{ij}^{Parcel}$). A combination of these two parameters is reflected in the metric and a clear difference between CRP and non-CRP is visible in Figure 4.

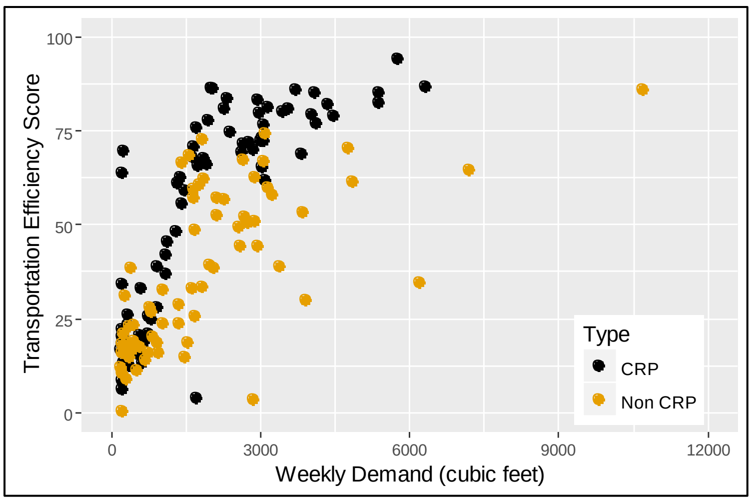

Figure 4: Transportation efficiency metric for 143 channels: CRP vs. Non-CRP

In order to predict the cost of transportation in CRP for non-CRP channels, we quantified the difference between CRP and non-CRP scores in Figure 4. The values of C_{ij}^{CRP} in Table 2 are the 3rd quartile of the CRP efficiency scores in each demand category for CRP channels and are considered as the achievable target levels.

The two required parameters to estimate the transportation cost of channel $i-j$ in CRP are C_{ij} *and* C_{ij}^{CRP}. We need to compute the CRP multiplier (ρ_{ij}^{CRP}) to adjust ATC_{ij} accordingly:

$$\rho_{ij}^{CRP} = \begin{cases} \dfrac{C_{ij}}{C_{ij}^{CRP}} & if\ C_{ij} < C_{ij}^{CRP} \\ 1 & Otherwise \end{cases} \tag{40}$$

Hence, the total transportation cost of channel $i - j$ in CRP (ATC_{ij}^{CRP}) will be:

$$ATC_{ij}^{CRP} = \rho_{ij}^{CRP} \times ATC_{ij} \tag{41}$$

2.6 Order Processing Cost

This cost component represents the cost of physical and technological activities associated with generating a purchase order (PO) at a distributor location, sending it to the manufacturer, handling it at a manufacturer DC, and receiving it at the distributor location. It also includes the cost of ordering and receiving associated with orders to the upstream manufacturing plants. Figure 5 illustrates the ordering mechanism graphically.

Let k_1 be the cost of ordering at the distributor level ($/PO)

Let z_1 be the cost of receiving at the distributor level ($/PO line)

Let k_2 be the cost of order processing at the manufacturer DC level ($/PO)

Let z_2 be the cost of handling the orders at the manufacturer DC level ($/PO line)

Let k_3 be the cost of ordering to the upstream at the manufacturer DC level ($/PO)

Let z_3 be the cost of receiving from upstream at the manufacturer DC level ($/PO line)

While the cost of ordering (k_1) and receiving orders (z_1) are the distributor's cost, the manufacturer is responsible for processing the incoming orders (k_2) from the distributor, handling cost at the DC (z_2), ordering (k_3) and receiving costs (z_3) from the upstream. While the ordering and order processing costs ($k_{1,2,3}$) are driven by number of POs, the receiving and handling costs ($z_{1,2,3}$) are driven by PO lines which are essentially different items on a PO.

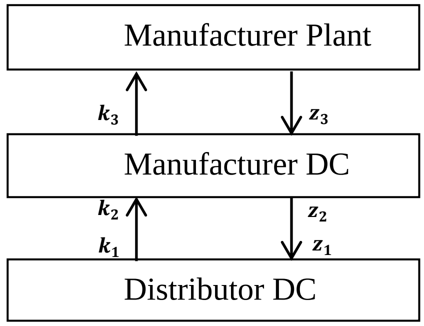

Figure 5: Ordering & handling cost components

The unit cost parameters that are defined above are estimated using an Activity Based Costing (ABC) approach. This methodology identifies the activities associated with each task and allocates cost to them by measuring the required time and labor. In this section, we present the modeling framework to compute the total cost of ordering and handling for both distributor and manufacturer knowing the unit cost parameters.

2.6.1 Distributor Cost

Ordering cost is driven by the number of orders that a distributor sends to a manufacturer and is not dependent on the size of orders. Therefore, the weekly ordering cost for channel $i-j$ can be computed as follows:

$$OC_{ij}^{d} = \overline{OF}_{D}^{ij} \times k_1 \tag{42}$$

On the other hand, receiving cost is driven by the number of different item types that a distributor receives on each PO (i.e. number of PO lines). The reason is that receiving activities such as receipt confirmation and put-away, are performed on a per-item-type basis. This means the variety of items on each PO (i.e. number of PO lines) is a main driver of the receiving cost. Assuming $\bar{l}_{ij}$ is the average weekly number of PO lines that are received on channel $i-j$, the weekly receiving cost is computed as follows:

$$RC_{ij}^{d} = \bar{l}_{ij} \times z_1 \tag{43}$$

Note that the units of the terms in Equation (43) are \$/period t = PO line/period t * \$/PO line where period t is assumed as a week in the case study. $\bar{l}_{ij}$ is not among the model inputs and needs to be estimated. We built a regression model to find the predictors of this parameter among the model inputs. We found two significant predictors. First is the size of the standardized item set for a channel (i.e. $|M|$), which is the number of items in the set, and second is the demand size of channel (i.e. $\overline{OF}_D \times E[D_t]$). The regression model was built using the data of 68 channels with R^2 of 0.9. We also cross validated the model by splitting the data into training and testing sets with the split ratio of 60%. The trained model performed acceptable on the testing set with mean absolute percentage error (MAPE) of 22% and R^2 of 0.83. Therefore, we estimated $\bar{l}_{ij}$ for each channel using the following regression model:

$$\hat{l} = 0.19836 \times |M| + 0.154511 \times (\overline{OF}_D \times E[D_t]) \tag{44}$$

Later in section 2.6.3, we will discuss how we built a similar model for predicting $\bar{l}_{ij}$ in CRP. To summarize, $OC_{ij}^d + RC_{ij}^d$ is the distributor's share of ordering and handling cost.

2.6.2 Manufacturer Cost

First, the exact number of items that a distributor receives each week needs to be shipped from a manufacturer DC. Therefore, the cost of handling at the manufacturer DC, which includes picking, packing and shipping activities, can be computed using the estimated $\bar{l}_{ij}$:

$$PC_{ij}^m = \bar{l}_{ij} \times z_2 \tag{45}$$

As discussed earlier, the manufacturer has the cost of processing the incoming orders in addition to the cost of ordering to the upstream.

$$OPC_{ij}^m = \overline{OF}_D^{ij} \times k_2 \tag{46}$$

$$OC_{ij}^m = \overline{OF}_W^{ij} \times k_3 \tag{47}$$

Equation (46) represents the processing cost of the incoming orders from downstream, and Equation (47) is the cost of ordering to the upstream. However, the order frequencies are not the same because the (r, Q) system of the upstream relationship requires different ordering parameters to satisfy the demand to the downstream. Therefore, $\overline{OF}_W^{ij}$ needs to be estimated using $E[X(2)]$ and q_w, which are the mean of the demand rate and the reorder quantity at the manufacturer DC (Section 2.4.1 and 2.4.2).

$$\overline{OF}_{W_{ij}} = \frac{\mu_w}{q_w} \tag{48}$$

Note that μ_w and q_w vary by channel but for notation consistency with Section 2.4, we don't add ij subscript to Equation (48).

The receiving cost of the incoming orders from the manufacturing plant can be computed in the same way the receiving cost is computed for the distributor (Equation (49)).

$$RC_{ij}^m = \bar{L}_{ij} \times z_3 \tag{49}$$

$\bar{L}_{ij}$ is the average weekly number of PO lines that are received from the upstream. To summarize, $PC_{ij}^m + OPC_{ij}^m + OC_{ij}^m + RC_{ij}^m$ is the manufacturer's share of ordering, handling and receiving cost.

2.6.3 Impact of CRP

In this section, we discuss the impact of CRP on the distributor and manufacturer DC's. The first benefit that a distributor immediately realizes in CRP is that ordering cost (OC_{ij}^d) becomes (essentially) zero because the manufacturer will automatically replenish the distributor by monitoring the inventory levels and using demand forecasts shared by the distributors (Kannan et al. 2013, Bookbinder et al. 2010). It is noteworthy that for small distributors that do not have information sharing mechanisms such as EDI in place, CRP requires a setup cost to provide the required system for information sharing.

Order processing cost for the manufacturer (OPC_{ij}^{m}) will also change in CRP because the manufacturer does not need to process any incoming PO's but instead should generate orders for the distributors. We found that order processing issues such as discrepancies, combining split orders etc. reduces virtually to zero in CRP. The process of order generation in CRP is very straight-forward and fast but requires higher skilled workers. By using the ABC approach, we found that ultimately k_2 reduces in CRP.

In order to quantify the impact of CRP on the distributor's handling cost (RC_{ij}^{d}) and manufacturer's handling cost (PC_{ij}^{m}), we need to look for any change in both $\bar{l}_{ij}$ and k values. A data analysis indicated that CRP distributors receive items in larger quantities but less frequently, as opposed to frequent and smaller quantities for non-CRP distributors. This implies significantly less put-away, sorting and receipt confirming activities in the receiving dock for CRP distributors and also less handling cost for the shipping DC. Figure 6 illustrates this fact for 158 different CRP and non-CRP channels. Obviously, as a channel becomes larger (x axis), it is likely to receive more variety of items (y axis). However, this happens at a faster rate for non-CRP channels. To quantify the impact of CRP on the distributor's receiving cost (RC_{ij}^{d}) and manufacturer's handling cost (PC_{ij}^{m}), we need to predict these costs in CRP. As discussed earlier and Equation (43) & (45) indicate, $\bar{l}_{ij}$ is the driver of both cost components. As Figure 6 illustrates, $\bar{l}_{ij}$, which is represented on the y axis, is expected to be lower for CRP channels. Thus, we fit a regression model (i.e., Equation (50)) to the existing 76 different CRP data points (lighter colored points) on Figure 6 to predict $\bar{l}_{ij}$ in CRP. We found the exact same predictors of regression Equation (44) as the significant ones here with R^2 of 0.98. We also cross validated the model by splitting the data into training and testing set with the split ratio of 60%. The trained model performed well on the testing set with MAPE of 16% and R^2 of 0.97.

$$\bar{l}_{ij}^{CRP} = 1.077136 \times |M| + 0.020362 \times (\overline{OF}_D \times E[D_t]) \tag{50}$$

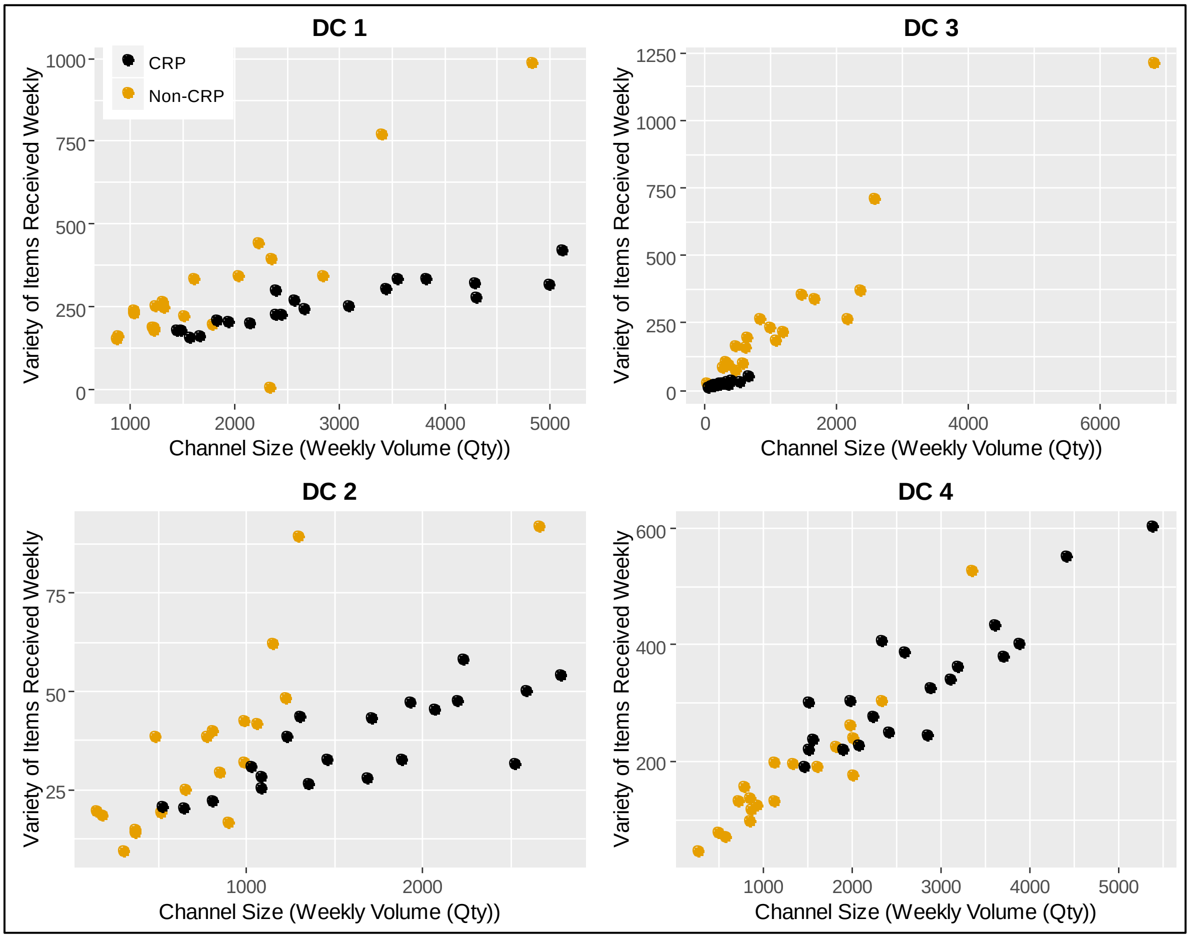

Figure 6: Variety of items ordered and received on each channel ($\bar{l}_{ij}$ vs. $\bar{l}_{ij}^{CRP}$) in a sample of 158 CRP and non-CRP channels on 4 different DCs where each DC supplies a specific product line

Any change in the handling unit cost values (z_1 and z_2) is not really a function of CRP because handling activities in DC's are still the same. They could change from a DC to another depending on many other factors such as DC layout and material handling devices. Therefore, (RC_{ij}^{d}) and (PC_{ij}^{m}) will change in CRP only due to a reduction in $\bar{l}_{ij}$ which can be computed for both CRP and non-CRP using regression models.

2.7 Model Application on a Case Study

In this section, we apply the model on a case study in which the benefits of a CRP relationship between a healthcare manufacturer and an independent distributor is quantified for both partners. The manufacturer has four DC's across the U.S. while the distributor has a network of 20 DC's mostly spread on the eastern half of the U.S. (Figure 7).

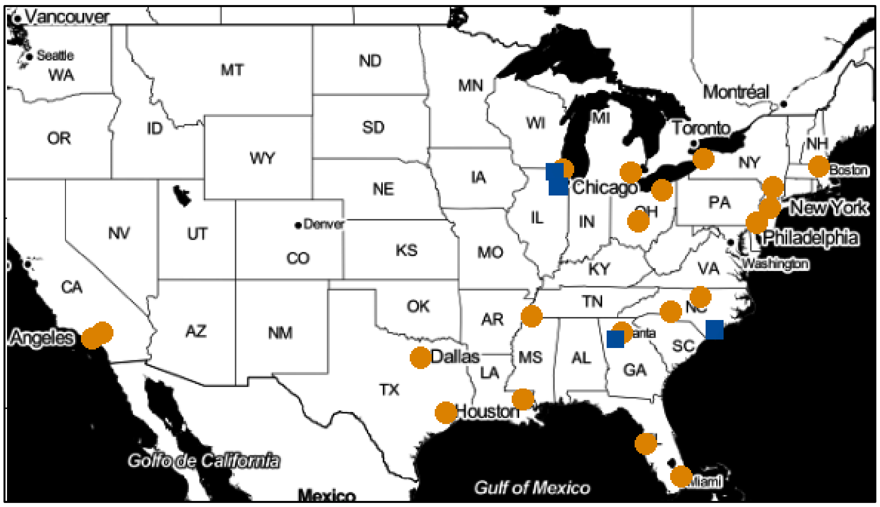

Figure 7: Case study illustration: 4 manufacturer DC locations (dark squares) and 20 distributor DC locations (lighter circles)

Although this business relationship could be as large as 80 different distribution channels (4×20), some of the channels are not practically active due to the demand and supply characteristics of the network. An initial data analysis showed that 72 channels regularly distribute the supply across the country and they will be the focus of this case study. We first illustrate the application of the model on one channel as an example and then show the output of the model on all the 72 channels.

2.7.1 Channel Instance

Consider a channel from the manufacturer's DC in Georgia (GA) to a distributor's DC in the Chicago area. Before computing the demand parameters, we should first set the basic inputs for the channel. These inputs are provided in appendix (Table 6). Then, we compute the demand-related inputs using the historical demand and shipping data of the channel. As discussed in the beginning of Section 2.4, the standardized item set has to be identified for the channel. The total number of items shipped on this channel is 1503 and per Equations (1)-(4), 301 items are identified as the set ($|M| = 301$) with the following characteristics:

$$P = \$149, \quad W = 8.5\ lbs., \quad V = 1.28\ ft^3$$

Once the standardized item set is defined, we can realize D_t over time and compute demand parameters. The followings are the weekly estimates:

$$\overline{OF}_D = 2.6, \quad E[D_t] = 1121, \quad Var[D_t] = (590)^2$$

Now we can use the model to calculate each cost component for non-CRP and CRP. Tables in the appendix show this calculation process.

As mentioned before, a reduction in $Var[D_t]$ and $Var[T_c]$ represents the impact of CRP on the holding costs of manufacturer and distributor, respectively. We approximate the impact of CRP by reducing $Sd[D_t]$ and $Var[T_c]$ by 50%. This is what we concluded after extensive investigation, including interviews and comparative data analysis of CRP and non-CRP channels.

As discussed in Section 2.5.4, transportation efficiency scores ($C_{ij}, C_{ij}^{max}, C_{ij}^{CRP}$) adjust the transportation cost estimation (TC_{ij}) for both CRP and non-CRP relationships.

In this example, we showed how the potential partners for CRP can use the model to predict the cost savings of a channel. Table 4 provides a summary of weekly costs and savings for the selected channel. It also shows how the distribution of costs and savings change by moving from non-CRP to CRP.

Table 4: Summary of costs and savings for the channel instance

	Non-CRP Cost	Distribution of SC cost	CRP Cost	Distribution of SC cost	CRP savings ($)	CRP savings (%)	Distribution of savings
Distributor	$ 2,256	16.32%	$ 1,482	16.49%	$ 774	34.30%	16.01%
Manufacturer	$ 11,569	83.68%	$ 7,509	83.51%	$ 4,060	35.09%	83.99%
Supply Chain	$ 13,825	100.00%	$ 8,991	100.00%	$ 4,834	34.96%	100.00%

2.7.2 Case Study Results and Discussion

This section presents the expected cost savings of CRP across the entire network. Such information is critical for both partners from different perspectives. First, it clarifies the impact of CRP not only on their organizations and but on the entire supply chain which could justify any initial investment associated with CRP. Second, it helps companies in setting up the partnership contract in a mutually beneficial manner.

The results reveal that the cost of supply chain, which is an accumulation of both manufacturer and distributor's costs across the network, will be reduced by 19.1% in CRP. As Table 5 indicates, the distributor will gain 33% of the total savings while the manufacturer gains the remaining 67%. The larger gain of the manufacturer is due to the fact that transportation is managed and paid by the manufacturer. However, Table 5 shows that the distributor saves more than what they contribute to the total cost of supply chain (i.e. 33% > 22%). This is why the distribution of total cost between two partners shifted towards the manufacturer in CRP.

Table 5: Distribution of CRP savings and supply chain cost between both partners

	Distribution of SC cost in non-CRP	Distribution of CRP Savings	Distribution of SC cost in CRP
Distributor	22%	33%	19%
Manufacturer	78%	67%	81%

The largest portion of the total cost for each partner is in inventory holding cost. This is due to the high sales price of items in the medical device industry (Figure 8**).** The results show that CRP reduces the cost in every cost component for both partners except for transportation where some channels do not expect reductions (Figure 9). This is due to the high transportation efficiency of those channels in non-CRP where no significant improvement is expected in CRP.

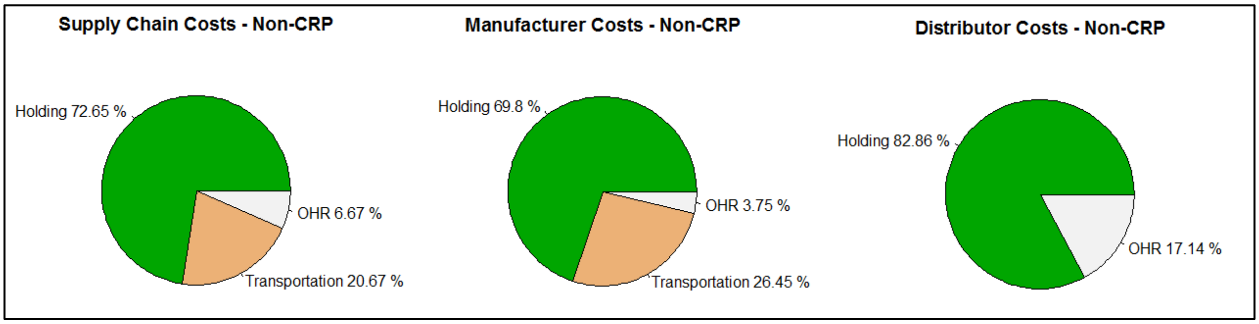

Figure 8: Distribution of cost components in the supply chain and for each partner[1]

Figure 9: Variability of cost reductions in each component across channels: manufacturer vs. distributor

Figure 9 also depicts the variability of cost reductions across the channels in which there is a noticeable pattern. The results indicate that manufacturer's benefits are greater but retailers have higher savings percentage rates. This is a very important observation that tremendously

[1] OHR stands for ordering/handling/receiving cost component.

helped both partners in the negotiation process to reach a sustainable agreement for starting a CRP. Mangiaracina et al. (2012) observed the same savings pattern in their case study. The distributor saves more in OHR cost because of first, zero ordering cost in CRP (OC_{ij}^{d}) and second more improvement room in the handling cost (RC_{ij}^{d}). The distributor also benefits more from holding cost savings because the impact of lead time, and variance of lead time ($E[L_c]$& $Var[L_c]$) is significant on decreasing the reorder point (r_c) that is needed for meeting the service level.

A sensitivity analysis on channel savings indicates that the general perception that higher savings belong to larger channels is not necessarily correct. The graph on the left of Figure 10 illustrates that although an increasing trend in monetary savings is visible as channels become larger, the slope is different for each product family. Therefore, the combination of channel size and product mix is a better indicator of monetary savings magnitude in CRP. On the other hand, percentage of savings in CRP (i.e., [CRP savings/cost in non-CRP] ×100) does not show the same pattern (graph on the right) as channel size increases. Percentage savings is an indicator of improvement potential of channels in CRP. Overall, these two graphs suggest that monetary savings might increase as channel size increases but improvement potential does not increase in the same way.

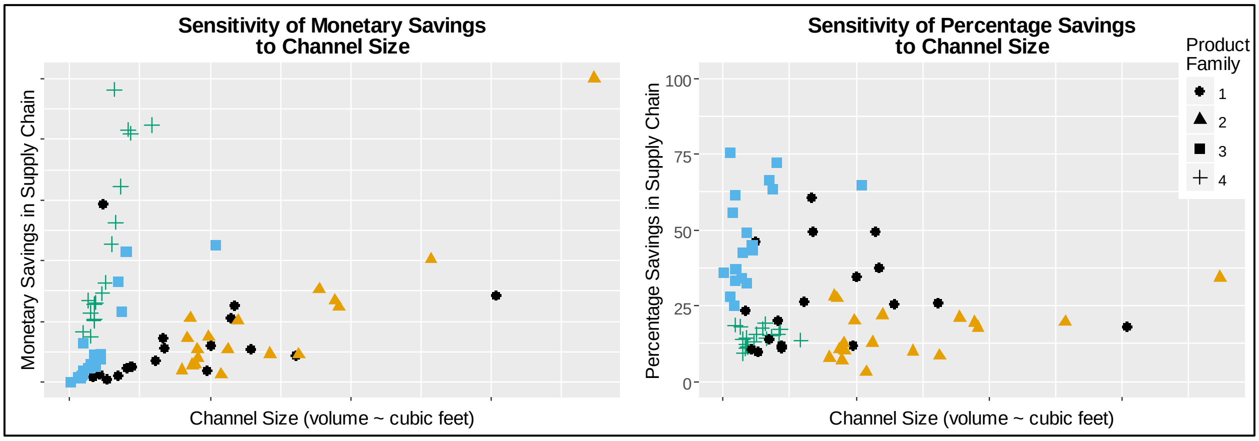

Figure 10: Sensitivity of savings to channel size

2.8 Conclusion and Future Work

This paper contributes to the literature by developing a data-driven model that captures the cost savings of CRP in different cost components and for both partners. In addition, the model does not impose assumptions that normally do not hold in practice. As discussed in the literature review section, most of the analytical and simulation models that have been developed either do not consider the impact of CRP on all the cost components or contain certain assumptions that limit the applicability of the models.

Another advantage of this model is that costs are estimated at the channel level. This allows the model to capture the dynamics of a business relationship between two potential CRP partners thoroughly. The model is used in a case study to help a healthcare manufacturer in analyzing a potential distributor for CRP. The results reveal that savings significantly vary across the channels depending on product mix, demand characteristics, handling and transportation requirements, etc. In addition, manufacturer and distributor locations experience different levels of expected savings. Results showed that the distributor generally gains more savings than the manufacturer in the shared cost components. The model substantially helped the CRP partners to have a clear understanding of the financial benefits of CRP, which is crucially important for selecting the partners effectively and setting fair contractual terms. One of the future works of this study is performing a comprehensive sensitivity analysis on various input parameters of the model. This can provide a great insight to the process of evaluating and selecting partners.

Although the model provides helpful insights about the benefits of CRP, it could be improved in different ways. First, the standardized item set is assumed to be a representative of product mix on each channel and the average characteristics of items demanded on that channel. A good area of improvement is proposing a methodology that better captures the variability across the product mix and its impact on the cost components. Another natural improvement is extending the model to the upstream where the manufacturer's DCs interact with the manufacturing plants. The impact of CRP on the upstream is not considered in this paper but is a valuable extension to the model.

One of the immediate future directions for this research is motivated by the fact that organizations do not enter into a strategic relationship such as CRP just because of cost savings (Parsa et al., 2016). They consider other factors that sometimes outweigh the cost savings of a CRP relationship. Those factors are generally qualitative factors and may have considerable influence during the decision-making process. Factors such as trust, team attitude, cooperation, power shift, implementation capability and shared business philosophy should be considered. A multi-objective decision model that can integrate the quantitative and qualitative decision factors would be a great contribution.

Appendix

Table 6: Basic inputs for the model

Parameter	Value	Parameter	Value
ω_{ij}	919 miles	τ_w	98%
R_{ij}^{FTL}	\$1.71 per mile	i_c & i_w	20% \$/\$/year
R_{ij}^{LTL}	\$15.59 cwt[2]	$E[T_c]$	7 days
R_{ij}^{Parcel}	\$21.20 cwt	$Var[T_c]$	4 hours
V_{min}^{FTL}	960 ft^3	k_1	\$5.50 /order
V_{max}^{FTL}	2000 ft^3	z_1	\$0.89 /order line
W_{max}^{FTL}	45000 lbs.	k_2	\$3.73 /order
τ_c	99%	z_2	\$0.20 /PO line

[2] "cwt" denotes cost of transportation (\$) per 100 lbs. This is a common unit of measure in transportation

Table 7: Inventory holding cost calculation for a channel (CRP vs. Non-CRP)

	Parameter	Non-CRP	CRP
Manufacturer	$E[D_t]$	1,121	1,121
	$Var[D_t]$	$(590)^2$	$(295)^2$
	r_w	7,109	6,753
	q_w	211	211
	$\bar{I}_w$	4,370	4,002
	$\bar{B}_w$	70	58
	$\overline{RR}_w$	0.96	0.96
	HC_w	\$2,504 per week	\$2,293 per week
In transit	$\bar{I}_T$	2,914	2,914
	HC_T	\$1,670 per week	\$1,670 per week
Distributor	$E[T_c]$	7 days	7 days
	$Var[T_c]$	16.8 $hours^2$	8.4 $hours^2$
	r_c	5,423	4,787
	q_c	1,121	1,121
	$\bar{I}_c$	2,713	2,084
	$\bar{B}_c$	4	1.5
	$\overline{RR}_c$	0.992	0.995
	HC_c	\$1,555 per week	\$1,194 per week

Table 8: Transportation cost calculation for a channel (CRP vs. Non-CRP)

	Parameter	Non-CRP Estimation	CRP Estimation
Manufacturer	$\hat{S}_{ij}^{w}$	9,528	9,528
	$\hat{S}_{ij}^{v}$	1,435	1,435
	N_{FTL}^{W}	0	0
	N_{FTL}^{V}	1	1
	N_{ij}^{FTL}	1	1
	$\hat{s}_{ij}^{w}$	0	0
	TC_{ij}^{FTL}	$1,571	$1,571
	TC_{ij}^{LTL}	0	0
	TC_{ij}^{Parcel}	0	0
	TC_{ij}	$1,571	$1,571
	C_{ij}	40.29	40.29
	C_{ij}^{max}	67.53	–
	ρ_{ij}	1.676	–
	C_{ij}^{CRP}	–	85.06
	ρ_{ij}^{CRP}	–	0.474
	ATC_{ij}	$6,845	$3,242

Table 9: Ordering and handling cost calculation for a channel (CRP vs. Non-CRP)

		Parameter	Non-CRP	CRP
Manufacturer	Order Processing	k_2	\$3.73	\$4.29
		$\overline{OF}_D^{ij}$	2.6	2.6
		OPC_{ij}^{m}	\$9.70	\$11.15
	Handling	z_2	\$0.70	\$0.70
		$\bar{u}_{ij}$	772	418
		PC_{ij}^{m}	\$540.40	\$292.60
Distributor	Ordering	k_1	\$5.50	\$0.00
		$\overline{OF}_D^{ij}$	2.6	2.6
		OC_{ij}^{d}	\$14.30	\$0.00
	Handling	z_1	\$0.89	\$0.69
		$\bar{u}_{ij}$	772	418
		RC_{ij}^{d}	\$687.08	\$288.42

References

Achabal, D.D., McIntyre, S.H., Smith, S.A., Kalyanam, K., 2000. A decision support system for vendor managed inventory. Journal of Retailing, 76(4),430-54.

Agrawal, V., Seshadri, S., 2000. Distribution free bounds for service constrained (Q, r) inventory systems. Naval Research Logistics, 47(8),635-56.

Angulo, A., Nachtmann, H., Waller, M.A., 2004. Supply chain information sharing in a vendor managed inventory partnership. Journal of business logistics, 25(1),101-20.

Anthony, S.W., Censlive, M., 2006. Observations on modelling strategies for vendor-managed inventory. Jnl of Manu Tech Mnagmnt, 17(4),496-512.

Aviv, Y., 2001. The effect of collaborative forecasting on supply chain performance. Management science, 47(10),1326-43.

Bhakoo, V., Singh, P., Sohal, A., 2012. Collaborative management of inventory in Australian hospital supply chains: practices and issues. Supply Chain Management: An International Journal, 17(2),217-30.

Blatherwick, A., 1998. Vendor-managed inventory: fashion fad or important supply chain strategy? Supply Chain Management: An International Journal, 3(1),10-1.

Bookbinder, J.H., Gümüs, M., Jewkes, E.M., 2010. Calculating the benefits of vendor managed inventory in a manufacturer-retailer system. International Journal of Production Research, 48(19),5549-71.

Brynjolfsson, E., Hitt, L.M., Kim, H.H., 2011. Strength in numbers: How does data-driven decisionmaking affect firm performance? Available at SSRN 1819486.

Cachon, G.P., Fisher, M., 2000. Supply chain inventory management and the value of shared information. Management science, 46(8),1032-48.

Cachon, G.P., Lariviere, M.A., 2005. Supply chain coordination with revenue-sharing contracts: strengths and limitations. Management science, 51(1),30-44.

Cachon, G.P., Zipkin, P.H., 1999. Competitive and cooperative inventory policies in a two-stage supply chain. Management science, 45(7),936-53.

Cachon, G., Fisher, M., 1997. Campbell Soup's continuous replenishment program: evaluation and enhanced inventory decision rules. Production and Operations Management, 6(3),266-76.

Cetinkaya, S., Tekin, E., Lee, C., 2008. A stochastic model for joint inventory and outbound shipment decisions. IIE transactions, 40(3),324-40.

Chen, D.Q., Preston, D.S., Xia, W., 2013. Enhancing hospital supply chain performance: A relational view and empirical test. Journal of Operations Management, 31(6),391-408.

Cheung, K.L., Lee, H.L., 2002. The inventory benefit of shipment coordination and stock rebalancing in a supply chain. Management Science, 48(2),300-6.

Choudhary, D., Shankar, R., 2015. The value of VMI beyond information sharing under time-varying stochastic demand. International Journal of Production Research, 53(5),1472-86.

Choudhary, D., Shankar, R., 2015. The value of VMI beyond information sharing in a single supplier multiple retailers supply chain under a non-stationary (R n, S n) policy. Omega, 51,59-70.

Choudhary, D., Shankar, R., Dey, P.K., Chaudhary, H., Thakur, L.S., 2014. Benefits of retailer–supplier partnership initiatives under time-varying demand: a comparative analytical study. International Journal of Production Research, 52(14),4279-98.

Clark, T.H., Croson, D., McKenney, J., Nolan, R., 1994. HE Butt Grocery Company: a leader in ECR implementation. HBS Case,9-195.

Darwish, M.A., Odah, O., 2010. Vendor managed inventory model for single-vendor multi-retailer supply chains. European Journal of Operational Research, 204(3),473-84.

Daugherty, P.J., Myers, M.B., Autry, C.W., 1999. Automatic replenishment programs: an empirical examination. Journal of Business Logistics, 20(2),63.

Derrouiche, R., Neubert, G., Bouras, A., 2008. Supply chain management: a framework to characterize the collaborative strategies. International Journal of Computer Integrated Manufacturing, 21(4),426-39.

Disney, S.M., Potter, A.T., Gardner, B.M., 2003. The impact of vendor managed inventory on transport operations. Transportation research part E: logistics and transportation review, 39(5),363-80.

Dong, Y., Dresner, M., Yao, Y., 2014. Beyond Information Sharing: An Empirical Analysis of Vendor-Managed Inventory. Production and Operations Management, 23(5),817-28.

Dong, Y., Xu, K., 2002. A supply chain model of vendor managed inventory. Transportation research part E: logistics and transportation review, 38(2),75-95.

Gavirneni, S., Kapuscinski, R., Tayur, S., 1999. Value of information in capacitated supply chains. Management science, 45(1),16-24.

Giannoccaro, I., Pontrandolfo, P., 2004. Supply chain coordination by revenue sharing contracts. International Journal of Production Economics, 89(2),131-9.

Govindan, K., 2013. Vendor-managed inventory: a review based on dimensions. International Journal of Production Research, 51(13),3808-35.

Gümüs, M., Jewkes, E.M., Bookbinder, J.H., 2008. Impact of consignment inventory and vendor-managed inventory for a two-party supply chain. International Journal of Production Economics, 113(2),502-17.

Holweg, M., Disney, S., Holmström, J., Småros, J., 2005. Supply chain collaboration:: Making sense of the strategy continuum. European management journal, 23(2),170-81.

Hopp, W.J., Spearman, M.L., 2011. Factory physics. : Waveland Press.

Iyer, A.V., Bergen, M.E., 1997. Quick response in manufacturer-retailer channels. Management Science, 43(4),559-70.

Kamalapur, R., Lyth, D., Houshyar, A., 2013. Benefits of CPFR and VMI collaboration strategies: A simulation study. Journal of Operations and Supply Chain Management, 6(2),59-73.

Kannan, G., Grigore, M.C., Devika, K., Senthilkumar, A., 2013. An analysis of the general benefits of a centralised VMI system based on the EOQ model. International Journal of Production Research, 51(1),172-88.

Keh, H.T., Park, S.Y., 1997. To market, to market: the changing face of grocery retailing. Long range planning, 30(6),836-46.

Kiesmüller, G.P., de Kok, T.G., Smits, S.R., van Laarhoven, P.J., 2004. Evaluation of divergent N-echelon (s, nQ)-policies under compound renewal demand. OR Spectrum, 26(4),547-77.

Krichanchai, S., MacCarthy, B.L., 2016. The Adoption of Vendor Managed Inventory for Hospital Pharmaceutical Supply. The International Journal of Logistics Management.

Kuk, G., 2004. Effectiveness of vendor-managed inventory in the electronics industry: determinants and outcomes. Information & management, 41(5),645-54.

Lee, H.L., Padmanabhan, V., Whang, S., 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. Management Science, 43(4),546-58.

Lee, H.L., So, K.C., Tang, C.S., 2000. The value of information sharing in a two-level supply chain. Management science, 46(5),626-43.

Lee, J., Cho, R.K., 2014. Contracting for vendor-managed inventory with consignment stock and stockout-cost sharing. International Journal of Production Economics, 151,158-73.

Lee, J., Ren, L., 2011. Vendor-managed inventory in a global environment with exchange rate uncertainty. International Journal of Production Economics, 130(2),169-74.

Li, S., Yu, Z., Dong, M., 2015. Construct the stable vendor managed inventory partnership through a profit-sharing approach. International Journal of Systems Science, 46(2),271-83.

Lyu, J., Ding, J., Chen, P., 2010. Coordinating replenishment mechanisms in supply chain: From the collaborative supplier and store-level retailer perspective. International Journal of Production Economics, 123(1),221-34.

Mangiaracina, R., Melacini, M., Perego, A., 2012. A critical analysis of vendor managed inventory in the grocery supply chain. International Journal of Integrated Supply Management, 7(1-3),138-66.

Mateen, A., Chatterjee, A.K., 2015. Vendor managed inventory for single-vendor multi-retailer supply chains. Decision Support Systems, 70,31-41.

Mateen, A., Chatterjee, A.K., Mitra, S., 2015. VMI for single-vendor multi-retailer supply chains under stochastic demand. Computers & Industrial Engineering, 79,95-102.

Mikael Ståhl Elvander, Sarpola, S., Stig-Arne Mattsson, 2007. Framework for characterizing the design of VMI systems. Int Jnl Phys Dist & Log Manage, 37(10),782-98.

Nagarajan, M., Rajagopalan, S., 2008. Contracting under vendor managed inventory systems using holding cost subsidies. Production and Operations Management, 17(2),200-10.

Niranjan, T.T., Wagner, S.M., Nguyen, S.M., 2012. Prerequisites to vendor-managed inventory. International Journal of Production Research, 50(4),939-51.

Ofuoku, M., 2009. Stochastic vendor managed inventory: impact on non-identical retailers, New Mexico State University.

Parsa, P., 2017 (Unpublished). Modeling and Monitoring the Inventory and Transportation Benefits of Supply Chain Collaboration PhD Dissertation, University of Arkasas.

Parsa P., Shbool M., Rossetti M., Zhang S., and Pohl E., 2016, “A Multi-Objective Decision Model for Supply Chain Collaboration Programs,” Proceedings of the Annual Decision Science Institute Conference

Pasandideh, S.H.R., Niaki, S.T.A., Nia, A.R., 2011. A genetic algorithm for vendor managed inventory control system of multi-product multi-constraint economic order quantity model. Expert Systems with Applications, 38(3),2708-16.

Ramanathan, U., 2014. Performance of supply chain collaboration–A simulation study. Expert Systems with Applications, 41(1),210-20.

Razmi, J., Rad, R.H., Sangari, M.S., 2010. Developing a two-echelon mathematical model for a vendor-managed inventory (VMI) system. The International Journal of Advanced Manufacturing Technology, 48(5-8),773-83.

Rossetti, M.D., Ünlü, Y., 2011. Evaluating the robustness of lead time demand models. International Journal of Production Economics, 134(1),159-76.

Sabath, R.E., Fontanella, J., 2002. The Unfulfilled Promise of Supply Chain Collaboration. Supply Chain Management Review, V.6, No.4 (July/Aug.2002), P.24-29

Sahin, F., Robinson, E.P., 2002. Flow coordination and information sharing in supply chains: review, implications, and directions for future research. Decision sciences, 33(4),505-36.

Sari, K., 2008. On the benefits of CPFR and VMI: A comparative simulation study. International Journal of Production Economics, 113(2),575-86.

Savaşaneril, S., Erkip, N., 2010. An analysis of manufacturer benefits under vendor-managed systems. IIE Transactions, 42(7),455-77.

Schiano, W., McKenney, J., 1996. Spartan stores incorporated: Reengineering for efficient consumer response. HBS Case,9-396.

Song, D., Dinwoodie, J., 2008. Quantifying the effectiveness of VMI and integrated inventory management in a supply chain with uncertain lead-times and uncertain demands. Production Planning and Control, 19(6),590-600.

Southard, P.B., Swenseth, S.R., 2008. Evaluating vendor-managed inventory (VMI) in non-traditional environments using simulation. International Journal of Production Economics, 116(2),275-87.

Stalk, G., Evans, P., Sgulman, L.E., 1992. Competing on capabilities: the new rules of corporate strategy. : Harvard Business Review.

Toptal, A., Çetinkaya, S., 2006. Contractual agreements for coordination and vendor-managed delivery under explicit transportation considerations. Naval Research Logistics (NRL), 53(5),397-417.

Torres, F., Ballesteros, F., Villa, M., 2014. Modeling a Coordinated Manufacturer–Buyer Single-Item System Under Vendor-Managed Inventory. In: Anonymous Handbook of EOQ Inventory Problems. : Springer, 247-278.

Van der Vlist, P., Kuik, R., Verheijen, B., 2007. Note on supply chain integration in vendor-managed inventory. Decision Support Systems, 44(1),360-5.

Waller, M., Johnson, M.E., Davis, T., 1999. Vendor-managed inventory in the retail supply chain. Journal of business logistics, 20,183-204.

Wang, H., Guo, M., Efstathiou, J., 2004. A game-theoretical cooperative mechanism design for a two-echelon decentralized supply chain. European Journal of Operational Research, 157(2),372-88.

Wong, W., Qi, J., Leung, S., 2009. Coordinating supply chains with sales rebate contracts and vendor-managed inventory. International Journal of Production Economics, 120(1),151-61.

Xu, K., Leung, M.T., 2009. Stocking policy in a two-party vendor managed channel with space restrictions. International Journal of Production Economics, 117(2),271-85.

Yang, K., Ruben, R.A., Webster, S., 2003. Managing vendor inventory in a dual level distribution system. Journal of Business Logistics, 24(2),91-108.

Yang, L., Ng, C., Cheng, T.E., 2010. Evaluating the effects of distribution centres on the performance of vendor-managed inventory systems. European Journal of Operational Research, 201(1),112-22.

Yao, Y., Dong, Y., Dresner, M., 2010. Managing supply chain backorders under vendor managed inventory: An incentive approach and empirical analysis. European Journal of Operational Research, 203(2),350-9.

Yao, Y., Dresner, M., 2008. The inventory value of information sharing, continuous replenishment, and vendor-managed inventory. Transportation Research Part E: Logistics and Transportation Review, 44(3),361-78.

Yao, Y., Evers, P.T., Dresner, M.E., 2007. Supply chain integration in vendor-managed inventory. Decision Support Systems, 43(2),663-74.

Zanoni, S., Jaber, M.Y., Zavanella, L.E., 2012. Vendor managed inventory (VMI) with consignment considering learning and forgetting effects. International Journal of Production Economics, 140(2),721-30.

Zavanella, L., Zanoni, S., 2009. A one-vendor multi-buyer integrated production-inventory model: The ‘Consignment Stock’case. International Journal of Production Economics, 118(1),225-32.

Zhang, T., Liang, L., Yu, Y., Yu, Y., 2007. An integrated vendor-managed inventory model for a two-echelon system with order cost reduction. International Journal of Production Economics, 109(1),241-53.

Zipkin, P.H., 2000. Foundations of inventory management. : McGraw-Hill New York.

3 CHAPTER 2

3.1 Introduction

Monitoring the performance of a supply chain is essential for the success of all the involved organizations as well as in communicating the necessary information for decision making. Performance measurement can reveal the areas that need improvement in order to meet customer expectations and strategic objectives (Chan, 2003). Traditionally, performance measurement is defined as the process of quantifying the effectiveness and efficiency of action (Neely et al., 1995). As Gleason and Barnum (1982) defined, effectiveness as “the extent to which an objective has been achieved” and efficiency as “the degree to which resources have been used economically.” Chan (2003) identified seven attributes as the important metrics for supply chain performance measurement, where two of them, *cost* and *resource utilization*, are quantitative. The profit of an enterprise is directly depended on the cost of its operations. In the distribution sector, cost is mostly a function of transportation, inventory and order processing activities; therefore, efficiency of logistics operations plays a key role in profitability.

This research is motivated by a collaborative project, with major healthcare manufacturers and distributors, about proper execution of a vendor managed inventory (VMI) program in the healthcare sector. The key incentive behind VMI is reducing inventory levels, while it enables gaining cost efficiency in transportation and order processing. Different industries have different priorities in performance measurement depending on their primary function. The healthcare sector, like many other sectors, substantially suffers from high inventory levels which is a key contributor to the excessive cost of logistics within the sector. Therefore, lowering inventory levels is the main objective of their VMI programs. However, both literature and industry practices have shown that significant transportation cost savings can be achieved in VMI through shipment consolidation and timely replenishment (Çetinkaya et al., 2008; Parsa et al., 2017). Utilizing the full potential of a VMI program is a collaborative endeavor between partners. Our experience shows that this can only be achieved if the objectives of all involved parties are considered. Thus, a verifiable performance measurement system that can monitor inventory, transportation and order processing efficiency over time is necessary to ensure the benefits of VMI for all partners.

As Gunasekaran and Kobu (2007) identified through a multi-faceted literature review on supply chain and logistics performance metrics, there are numerous overlapping metrics with 85% of them being quantitative, mostly concerned with financial performance, and focused on a single function of logistics operations (i.e. inventory, transportation, etc.). There has not been a significant work to design metrics that explore the relationship between functions or propose a statistical screening framework to monitor them over time. In this paper, we present multi-attribute efficiency metrics that can show the trade-off in gaining efficiency between multiple functions of logistics.

One of the challenges for downstream partners in VMI programs is ensuring a desired service level for end customers, which requires holding enough inventory. On the other hand, gaining efficiency in transportation requires shipment consolidation which can be harmful from the perspective of inventory efficiency. It is essentially a tradeoff between inventory efficiency and transportation efficiency. A similar trade-off exists between inventory efficiency and order processing efficiency. In this paper, we develop metrics that can illustrate the status of a system with respect to such tradeoffs over time. In addition, we determine optimal trade-off levels for each metric as well as develop a statistical process control (SPC) system to monitor them over time. The SPC system suggests whether the system is acting normal or if a significant shift has happened. We will discuss how to use appropriate statistical methods for various time-series behaviors of the metrics. The discussion will be coupled with examples on the application of the metrics in a healthcare supply chain, using datasets obtained from a group of major healthcare partners in the U.S.

3.2 Literature Review

Research in the area of supply chain performance measurement has been active since early 1990's. The research contributions can be categorized into two groups. One group is focused on proposing individual metrics to measure performance, while the other group proposes appropriate measurement systems. The definition of performance can be different for each organization and it depends on the goals of the organization. There is not a consensus in the supply chain literature about the definition of performance. In the literature, performance has been defined as a combination of other measures or so-called *performance dimensions*. However, from the most general and holistic view, many categorized these dimensions into two general groups of efficiency and effectiveness (Mentzer and Konrad, 1991). In other words, performance has been defined as a function of effectiveness and efficiency.

As Gleason and Barnum (1982) defined, effectiveness is "the extent to which an objective has been achieved" and efficiency is "the degree to which resources have been used economically." Therefore a well-balanced and interconnected group of metrics from both categories of efficiency and effectiveness forms a good performance measurement system. Other researchers have introduced additional dimensions for performance in order to propose a more comprehensive concept with greater level of details. For example, Beamon (1999) introduced flexibility as the third dimension of performance. Flexibility in a supply chain represents the ability of responding to a changing environment. Likewise, Fugate et al. (2010) and Langley Jr and Holcomb (1992) introduced differentiation as the third dimension to performance. Differentiation refers to logistics superiority when compared to competitors.

From another standpoint, individual metrics can be categorized into “hard” and “soft” metrics. Hard metrics are strictly quantitative such as net income or days of inventory on hand while soft metrics, such as customer satisfaction ratings, are more qualitative and are subject to personal judgments. Since some dimensions of performance cannot be measured quantitatively (e.g. customer service satisfaction), hard measures should be supplemented with soft ones in a well-representative measurement system (Chow et al., 1994).

The primary goal of physical distribution is to move goods from the supplier all the way to final the selling points. In this mission, the two most important criteria that determine the execution performance are cost and customer service (Mentzer and Konrad, 1991). We know that companies seek to reach a point that serves their desired balance between cost and customer service, or in other words efficiency and effectiveness. Therefore, a good performance measurement system in supply chain management is usually a mix of soft, hard, efficiency and effectiveness metrics.

Caplice and Sheffi (1994) studied research that proposed several criteria to consider when selecting individual performance metrics for monitoring logistics operations. Individual metrics are the building blocks of a measurement system and their “goodness” is essential to performance measurement. The paper summarizes the evaluation criteria existing in the literature to eight different criteria: *validity*, *robustness*, *usefulness*, *integration*, *economy*, *compatibility*, *level of detail*, and *behavioral soundness*. It is practically impossible to develop metrics that perform excellent in each of the eight criteria. The paper investigates the critical tradeoffs that exist between the criteria. The same authors also proposed a useful set of evaluation criteria for selecting the right combination of metrics to design a balanced and meaningful logistics measurement system (Caplice and Sheffi, 1995).

One way to perform an overall evaluation or to monitor performance as a whole is to use economic theory called *utility*. Utility is the final performance measure of a system when multiple active performance metrics are considered. Assume there are n metrics and each has a value $x_1, x_2, \dots, x_n$. In utility theory, there is a function called the utility function, which maps these attributes into a single cardinal utility $u = f(x_1, x_2, \dots, x_n)$. The form of utility function is usually approximated through a simplified function such as a linear additive model. There are similar scoring methods such as a Kiviat graph, where good and bad attributes alternate so a good performer results in a star-shaped graph (Figure). Spider graph is another scoring method in which the total performance is expressed as the percentage of the surface covered by the diagram created by connecting actual scores (Figure). In other words, performance is summarized as a single number (Kleijnen and Smits, 2003).

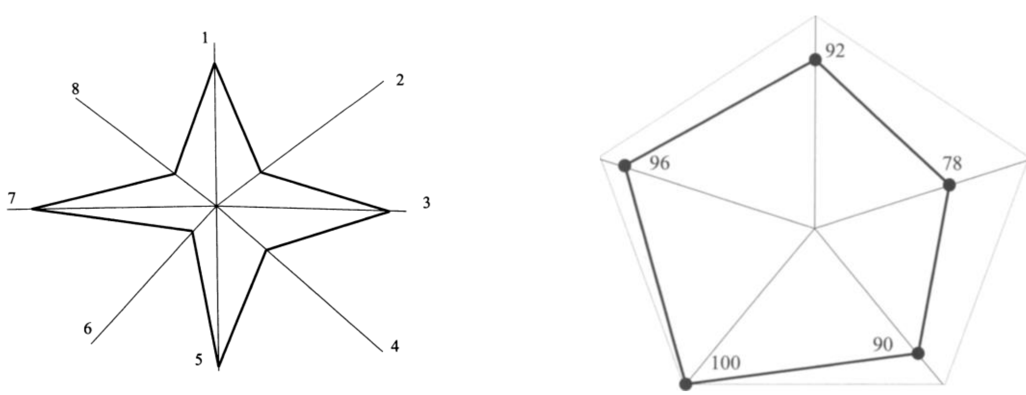

Figure 1: Kiviat graph (left) and Spider graph (right)

Perhaps the initial effort toward developing a comprehensive and balanced performance measurement system is made by Kaplan and Norton (2005) with the balanced scorecard methodology. Supply chain performance used to be monitored by only financial accounting measures such as sales figures, cash flow and operating income. These measures were criticized by many due to their backward-looking focus and their inability to cope with the new terms of competition and the idea that traditional financial metrics do not enhance success factors such as customer satisfaction, quality, employee motivation and etc. In other words, the traditional performance measurement had not looked at the entire picture. The balanced scorecard approach complements the financial measures by adding operational measures that are mostly focused on customer satisfaction, innovation, and improvement activities, which drive future financial performance. The scorecard methodology provides four critical perspectives to managers, especially senior executives who want to view the most complete picture of company's status. The four perspectives are: *financial perspective, customer perspective, internal business perspective,* and *innovation and learning perspective*. Each perspective is divided to a number of goals and each goal associates with a metric that reflects the performance with respect to the corresponding goal and perspective. To summarize, balanced scorecard provides an integrated dashboard of metrics that should be monitored to have a comprehensive picture of business performance. The importance of integrated metrics and their implementation process are investigated by other prominent research works such as Bullinger et al. (2002) and Lambert and Pohlen (2001) where they extend the use of balanced scorecard within the supply chain operations reference model (SCOR).

Performance measurement in supply chain collaboration programs, such as a continuous replenishment program (CRP) or vendor managed inventory (VMI), is very important. Several benefits of such programs have been achieved by suppliers, retailers, manufacturers and customers. These benefits are also well documented in the literature. The benefits include cost reductions throughout the supply chain, bullwhip effect reduction, improved service, sales increase, improved product availability, shorter lead times, etc. (Disney et al., 2003; Lee et al., 1997). On the other hand, implementing and maintaining supply chain collaboration programs impose extra costs to partners such as technological investments (e.g. advanced warehouse management and electronic data interchange systems), and higher skilled planners (Barratt, 2004). Many of the collaboration efforts failed in the past mostly due to lack of trust, fear of failure and operational complexity (Kohli and Jensen, 2010). Therefore, monitoring the performance of a supply chain collaboration program, once it is initiated, is crucial for partners and eventually for the success of the program.

As mentioned before, correct metrics that represent both efficiency and effectiveness need to be selected for performance measurement. Collaboration programs are primarily implemented to reduce costs and improve the service level. Thus, a mix of metrics from both operational and financial stand points that cover both dimensions of performance should be used for a meaningful assessment. The metrics that have been used are inventory levels, cycle time, fill rate, transportation cost, sales, market competitiveness, and etc. (Kohli and Jensen, 2010).

3.3 Metrics

3.3.1 Transportation and Inventory Efficiency (TIE) Metric

Transportation and inventory management can potentially be contradictive. Transportation efficiency can be increased by achieving a higher level of consolidation, which is potentially against inventory efficiency and possibly effectiveness because not only it can increase inventory levels but may increase the chance of stock out. A metric that can illustrate the trade-off between transportation and inventory efficiencies and be used to monitor efficiency over time is valuable. In addition, it can be coupled with a similar hybrid effectiveness metric to monitor performance over time.

The idea is to guarantee a certain fill rate level for incoming demands and computing the required replenishment quantity and cycle time that can satisfy that fill rate requirement. Our assumption is stochastic approximation of demand over time and different fill rate requirements based on an ABC classification. The classification helps to treat items appropriately without setting a single fill rate requirement for all of them. The time period that the demand is supposed to represent can vary but here we assume weekly demand because the replenishment decisions are made on a weekly basis in the data available. This will generally cause weekly performance metrics to be tabulated. However, this decision should be made wisely and with more analytical considerations.

The rest of this section is a mathematical illustration of the metric followed by a case study that illustrates the application of the metric on a healthcare supply chain demand dataset.

$\lambda_{j,t}$	the aggregate demand of class *j* items in week *t* (Qty/week)
k	the cost of ordering ($/order)
i	the holding charge ($/$/week)
$\overline{c_J}$	the average unit cost of class j items ($/item)
τ_j	the fill rate requirement for class j items
$\widehat{q_{j,t}^*}$	an estimate of optimal aggregate order quantity for class j items in week *t* (Qty)
$\widehat{OF_{J,t}^*}$	an estimate of optimal aggregate order frequency for class j items in week *t*
$\widehat{OF_t^*}$	an estimate of optimal aggregate order frequency for all items in week *t*
$\widehat{Q_t^*}$	an estimate of optimal aggregate order quantity for all items in week *t* (Qty)
$\widehat{Q_{v,t}^*}$	volume estimate of optimal aggregate order quantity for all items in week *t* (ft^3)
$\widehat{Q_{w,t}^*}$	weight estimate of optimal aggregate order quantity for all items in week *t* (lbs.)

$\overline{v_J}$ the average unit volume of class j items

$\overline{w_J}$ the average unit weight of class j items

At this stage, we look at each lane in the supply network separately because transportation decisions are normally made at the lane level. A lane is a supply connection between a supplier (manufacturer) and their customer (distributor). Therefore, all the above parameters should be computed for each lane separately which makes the metric monitor each lane separately. First, we need to compute the order quantity and frequency that can satisfy the fill rate requirement of each item class. Optimal order quantity subject to a fill rate constraint (τ_j) can be approximated using a lower bound that is obtained in Agrawal and Seshadri (2000) and a heuristic approach discussed on page 226 of Zipkin (2000).

$$\widehat{q_{j,t}^*} = \sqrt{\frac{2k\lambda_{j,t}}{i\overline{c_j}\tau_j}} \quad , \quad \widehat{OF_{j,t}^*} = \frac{\lambda_{j,t}}{\widehat{q_{j,t}^*}} \qquad j \in \{A, B, C\}$$

There are likely to obtain different optimal order frequency values for different item classes (i.e. $\widehat{OF_{J,t}^*}$ varies across item classes in each week). In order to compute $\widehat{Q_t^*}$, which is an estimate of optimal aggregate order quantity for all item classes in week *t*, we need to have a common order frequency across the item classes. In order to avoid increasing inventory levels, we choose the largest order frequency among the order frequency of three item classes:

$$\widehat{OF_t^*} = max\ \left\{\widehat{OF_{J,t}^*}\right\}\ over\ j$$

Hence, an estimate of optimal aggregate order quantity for all items in week *t* (i.e. $\widehat{Q_t^*}$) can be computed:

$$\widehat{Q_t^*} = \frac{\sum_j \lambda_{j,t}}{\widehat{OF_t^*}} \quad , \quad \widehat{Q_{v,t}^*} = \widehat{Q_t^*} \times \frac{\sum_j \lambda_{j,t}\,\overline{v_J}}{\sum_j \lambda_{j,t}} \quad , \quad \widehat{Q_{w,t}^*} = \widehat{Q_t^*} \times \frac{\sum_j \lambda_{j,t}\,\overline{w_J}}{\sum_j \lambda_{j,t}}$$

Now, in order to understand how well actual shipment quantities match the optimal weekly order quantities, they will be compared with each other. The metric will illustrate this comparison over weeks. In other words, the metric will show over time, how close weekly average shipment size is to the optimal shipment size. If historical weekly average shipment size is larger than the optimal size (i.e. $\widehat{Q_{v,t}^*}$), it means excessive shipment consolidation is applied to make transportation too efficient. This typically causes higher inventory levels at the destination and could lead to a high chance of stock out due to infrequent replenishment. On the other hand, if the average shipment size is smaller than the optimal size, it means transportation is inefficient due to the insufficient shipment consolidation which in turn makes inventory holding to be excessively efficient. This metric will capture the trade-off between inventory and transportation efficiency and will show how a system performs over time with respect to this trade-off.

In order to make the comparison between optimal shipment size and average weekly shipment size correctly, we need to use the most representative Q_t^* estimate ($\widehat{Q_t^*}$ or $\widehat{Q_{v,t}^*}$ or $\widehat{Q_{w,t}^*}$). Transportation cost and decisions are driven by size properties of shipments which are normally cube and weight. Shippers use them to determine a suitable transportation mode for their shipments. Whether to use cube or weight depends on whether products cause a shipping truck to exceed its cube limit first or weight limit first. Once this rule is established, each week's demand size determines the best transportation mode to use. If the demand is less than the cutoff limit between FTL and LTL, LTL is the best mode to use and the metric value will be computed using $\widehat{Q_{w,t}^*}$ because LTL cost structure uses shipment weight as the cost driver. In contrast, if demand is within the FTL limits, FTL should be selected and the metric for will be using $\widehat{Q_v^*}$ and this is due to our assumption that the cube limits of shipping truck is exceeded first.

$\bar{s}_{w,t}$ the average shipment weight (lbs.) on a lane in week $\underline{t}$

$\bar{s}_{v,t}$ the average shipment volume (ft^3) on a lane in week t

M_t the value of the metric in week t

The value of metric in each week is the following ratio:

$$M_t = \begin{cases} \left(\bar{s}_{w,t} \big/ \widehat{Q^*_{w,t}} \right) \times 100 & \text{if LTL is the mode} \\ \left(\bar{s}_{v,t} \big/ \widehat{Q^*_{v,t}} \right) \times 100 & \text{if FTL is the mode} \end{cases}$$

M_i values vary from zero to virtually infinity because average shipment size (i.e. $\bar{s}_{w,t}$ or $\bar{s}_{v,t}$) can become much larger than optimal order quantity (i.e. $\widehat{Q^*_{w,t}}$ or $\widehat{Q^*_{v,t}}$). However, 100 is the target level for this metric, where average weekly shipment size is equal to the estimate of optimal order quantity. Above the target level, as metric values become larger, the system becomes excessively efficient in transportation and inefficient in inventory management. On the other hand, below the target level as metric values decrease, we see small size/frequent shipments, which indicate transportation inefficiency and excessively efficient inventory management (Figure 2)

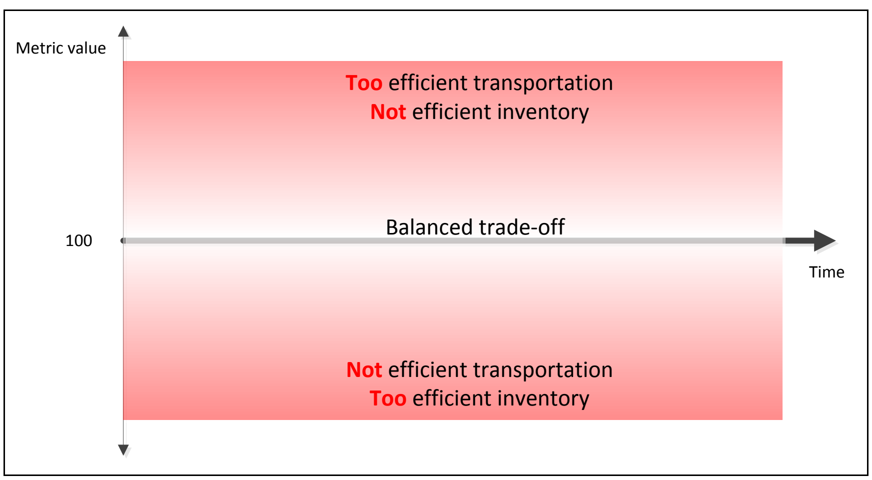

Figure 2: a schematic graph for the TIE metric

We computed this metric to monitor the efficiency of transportation and inventory operations for a healthcare supply chain. A data set containing two years of demand and shipping data is used. The shipped items are classified into three groups of A, B, and C with 0.98, 0.9, and 0.85 fill rate requirements (i.e. τ). Inventory carrying charge (i.e. i) is assumed to be 20% annually while ordering cost (i.e. k) is assumed to be $50 per order, where an order represents a purchase order that can contain multiple items. Figure 3 depicts the developed metric over two years on a weekly basis for an example lane from Atlanta, GA to Dallas, TX.

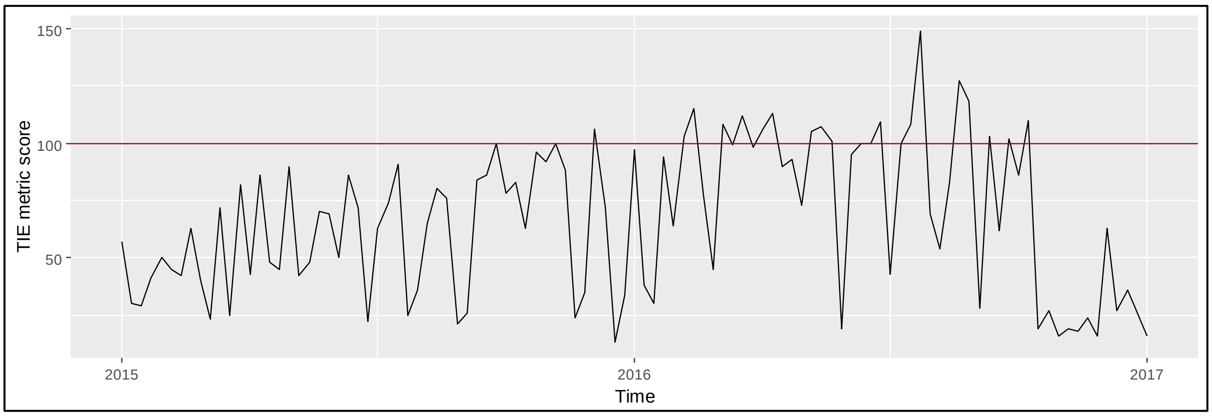

Figure 3: TIE metric scores over time for an example lane

As Figure 3 shows, the metric scores are mostly under the target level which clearly implies inventory efficiency is preferred to transportation efficiency. There is also a visible upward trend to achieve better transportation efficiency. The target line represents optimal trade-off between inventory and transportation efficiency considering the associated costs and fill rate requirements. Points below the target line indicate frequent and small size shipments which suggest poor shipment consolidation. On the other hand, inventory had been managed efficiently since small size/frequent shipments are used to satisfy the demand over time.

To summarize, this metric shows the trade-off between transportation and inventory efficiency over time. In addition, scores of a lane over time can be aggregated into a single score and be used as an aggregate efficiency score to compare efficiencies of multiples lanes with each other. It is noteworthy that this metric does not provide decisive insights into effectiveness of either transportation or inventory management because if we define effectiveness as satisfying the demand on time, any value of this metric does not help in drawing a conclusion.

3.3.2 Order Processing and Inventory Efficiency (OIE) Metric

Order processing involves a set of activities that starts with a customer generating a purchase order (PO) and finishes with the same customer putting away the received purchase order in warehouse. Figure 4 graphically list the activities involve in order processing.

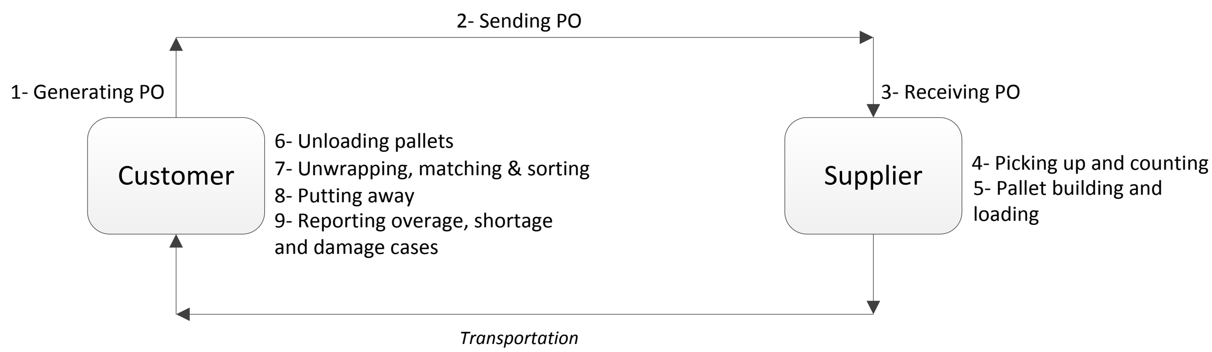

Figure 4: Order processing activities

There are several factors that increase the efficiency of order processing such as material handling automation, electronic PO processing, etc. However, our cost analysis revealed that order frequency and size of orders (in particular, quantity ordered for each item) are the main drivers for efficiency of order processing activities. The entire process of order processing becomes more efficient, and possibly effective, if order quantities round up to multiple numbers of tiers or pallets. There are three reasons behind this claim. First of all, when order quantities round up to the nearest multiple of tier or pallet, order frequency decreases. In other words, the orders become larger and more consolidated which reduces the work load and travel time in DC. Second, activities 4, 7, and 8 in Figure 4 are the most time consuming and costly activities of order processing (i.e. our studies show that they make up more than 75% of the total time and cost) and they are negatively impacted when the order quantities are not multiples of tier or pallet. Picking up, counting, sorting, and putting away processes become much faster and efficient when tiers or pallets need to be processed instead of individual piece picking. Third, ordering tier and pallet quantities make pallet building and loading easier and also reduces the number of overages, shortages and damages (OSD) during the transportation phase.

However, one important point to consider is that rounding up to multiple numbers of tiers or pallets might increase the inventory levels and cost. Therefore, in this metric we focus on the trade-off between inventory holding cost and order processing cost. Just like the previous metric, we consider three classes of items and show the efficiency trade-off between order processing and inventory holding. The metric that we introduce in this section shows the efficiency of historical order processing in comparison with the best-case scenario, where optimal quantities are rounded up to the closest tier or pallet sizes. By choosing to round up (and not round down), we make sure that we do not increase the chance of stock out.

The metric will be developed to monitor the efficiency of order processing at both individual item level and item class level (i.e., three item classes of A, B and C). The metric will be used over time and for each lane separately because the class of items could differ among the lanes. Similar to the previous section, the time increment is assumed to be weekly but it could be assumed differently. The rest of this section is a mathematical illustration of the metric followed by a case study that illustrates the application of the metric on a healthcare supply chain demand data set.

$\lambda_{j,t}$ the demand for item j in week t (Qty/week)

k the cost of ordering per item ($/item)

i the holding charge ($/$/week)

c_j the unit cost of item j ($/item)

τ_j the fill rate requirement for item j

$q_{j,t}^{*}$ the optimal weekly order quantity for item j in week *t* (Qty)

$OF_{j,t}^{*}$ the optimal weekly order frequency for item j in week *t*

t_j the number of cases in a tier for item j (i.e. tier quantity)

p_j the number of cases in a pallet for item j (i.e. pallet quantity)

$OF_{j,t}$ weekly order frequency of item j in week *t*

First, we need to compute the optimal order quantity and frequency that can satisfy the fill rate requirement for each item:

$$q_{j,t}^{*} \cong \left\lceil \sqrt{\frac{2k\lambda_{j,t}}{ic_j\tau_j}} \right\rceil \quad , \quad OF_{j,t}^{*} = \frac{\lambda_{j,t}}{q_{j,t}^{*}} \tag{1}$$

In order to achieve efficiency in order processing and not increasing the chance of stock out, we round up $q_{j,t}^{*}$ to the nearest multiple of corresponding tier quantities (t_j). $q_{j,t}^{con}$ indicates the rounded-up order quantity and $OF_{j,t}^{con}$ is the corresponding frequency:

$$q_{j,t}^{con} = \begin{cases} \left\lceil \dfrac{q_{j,t}^{*}}{t_j} \right\rceil t_j & if\ q_j^{*} \neq 0 \\ 0 & if\ q_j^{*} = 0 \end{cases} \quad , \quad OF_{j,t}^{con} = \frac{\lambda_j}{q_{j,t}^{con}} \tag{2}$$

As indicated earlier in this section, order frequency is one of the main efficiency drivers of order processing activities. The metric compares actual weekly order frequency $OF_{j,t}$ with $OF_{j,t}^{con}$ to indicate how efficient the order processing has been with respect to the trade-off with inventory holding. $OF_{j,t}^{con}$ represents the “optimal” trade-off between order processing efficiency and inventory holding efficiency because it is the smallest order frequency that generates an order quantity (i.e. $q_{j,t}^{con}$) with the size of multiple tiers. In other words, $q_{j,t}^{con}$ represents an order quantity that is rounded up just about enough to achieve the closest order processing efficiency.

Hence, in our metric we assign score of 100 to $OF_{j,t}^{con}$ and compute a relative score for $OF_{j,t}$ which represents the historical ordering.

$\rho_{j,t}$ the metric score for historical ordering $(q_{j,t}, OF_{j,t})$ for item j in week t

$$\rho_{j,t} = \left(\frac{OF_{j,t}^{con}}{OF_{j,t}}\right)\times 100 \tag{3}$$

This metric will be compared against 100, which is the target and the score for ordering tier quantity. As illustrated in Figure 5, above the target level, as metric value becomes larger, system becomes excessively efficient in order processing and inefficient in inventory management. Below the target level as metric value decreases, we see small size/frequent shipments, which indicate order processing inefficiency and excessively efficient inventory management.

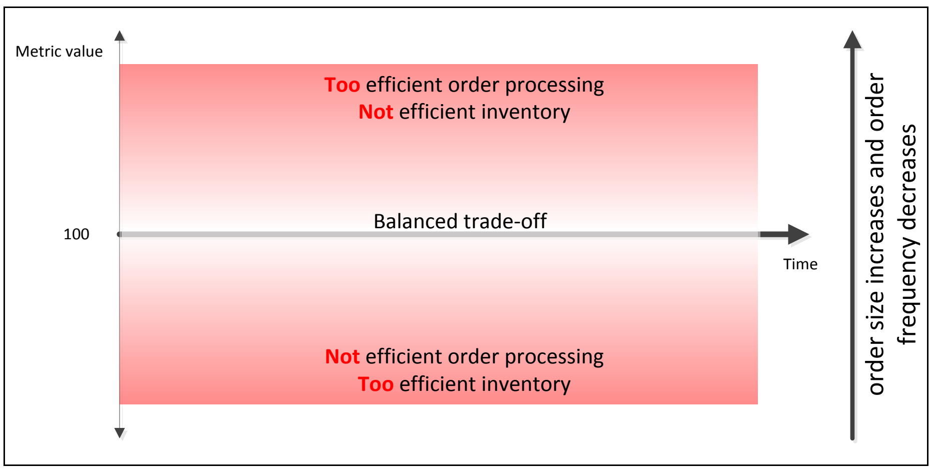

Figure 5: a schematic graph for order processing/inventory efficiency metric

We computed this metric for the same instance lane to monitor the efficiency of order processing and inventory and the trade-off between them over time. A data set containing two years of demand data is used in which the shipped items classified into three groups of A, B, and C with 0.98, 0.9, and 0.85 fill rate requirements (i.e. τ). Inventory carrying charge (i.e. *i*) is assumed to be 20% annually while ordering cost (i.e. k) is assumed to be \$1.55 per individual item. Figure 6 depicts the metric over time for an individual A-class item with high demand.

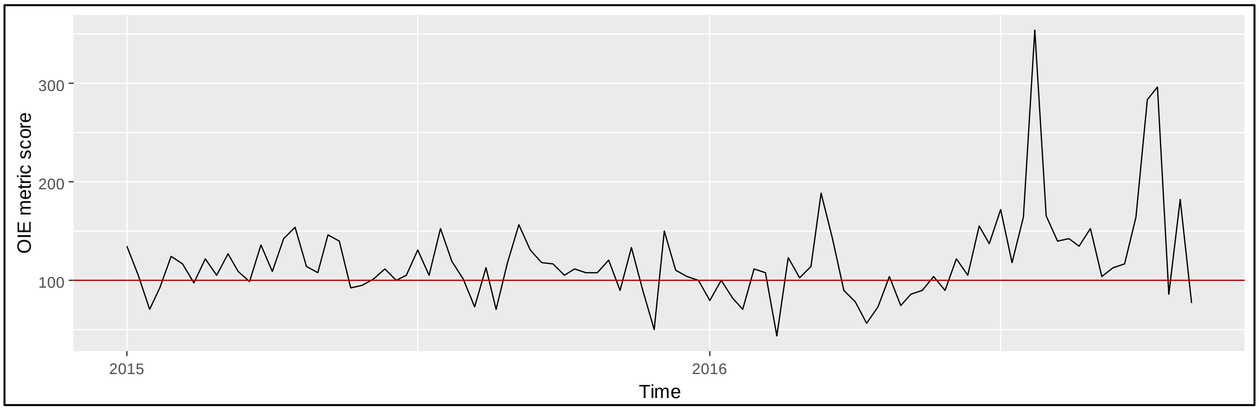

Figure 6: OIE metric scores for an individual A-class item on an example lane over time

The metric values are fluctuating around the target line but sometimes they are significantly distant from the balanced trade-off level. For example, toward the end of 2016, the metric score is at 350 (i.e. $\rho_{j,t} = 350$) which indicates that order processing is very efficient due to an order quantity that is substantially larger than the optimal level. Large order sizes cause reduction in order frequency which increases order processing efficiency. However, this is achieved at the expense of losing inventory efficiency because the order size in that week can excessively increase the inventory level.

Monitoring the efficiency of order processing is not practical when it is at the individual item level. Companies normally do not treat items individually; they instead classify them into three classes of A, B and C and manage each class differently. It is much easier and practical to monitor efficiency for each class of items. Therefore, we will modify the metric and tailor it for the item class level instead of individual item level. Let

$\beta_{n,t}$ the metric score of historical ordering for item class n in week t

The metric can be computed for each item class (three classes of A, B and C) by taking the weighted average of the individual item scores within the same class. The weights are the demand magnitudes of individual items (i.e. $\lambda_{j,t}$):

$$\beta_{n,t} = \frac{\sum_{j \in n} \rho_{j,t} \times \lambda_{j,t}}{\sum_{j \in n} \lambda_{j,t}} \tag{4}$$

$\beta_{n,t}$ is the equivalent of $\rho_{j,t}$ at the item class level and can be similarly monitored over time to ensure the efficiency of order processing and its trade off with inventory efficiency. Monitoring three charts per lane is much more reasonable and practical than one chart per individual item per lane. Using the same data set, we computed $\beta_{n,t}$ for three item classes on the same lane from a DC in Atlanta to a customer location in Dallas (Figure 7).

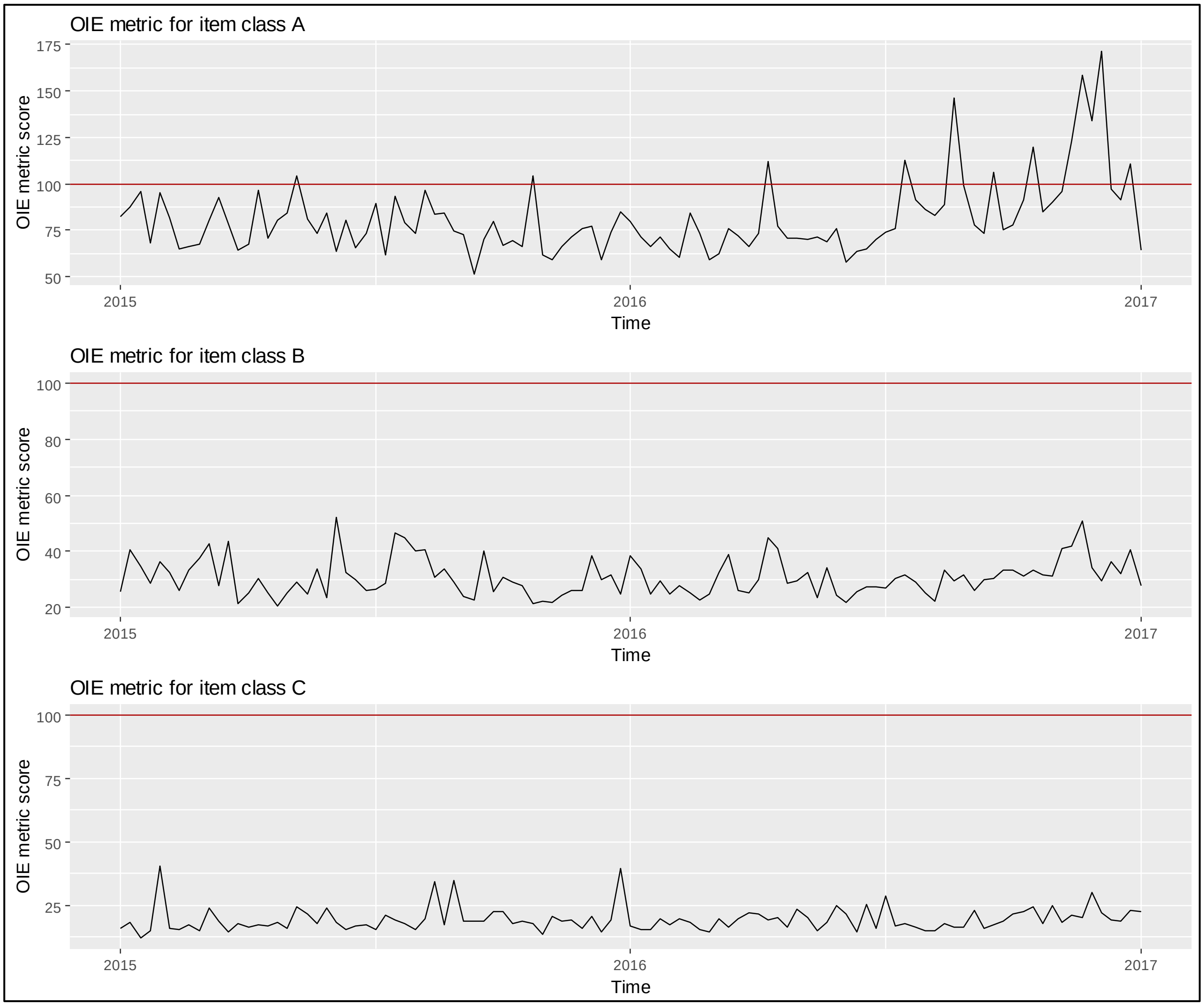

Figure 7: Order processing metric scores for three item classes on an example lane over time

The striking trend that draws attention in the first glance is that the metric ($\beta_{n,t}$) is almost consistently below the target line. As discussed earlier, metric values below the target level indicate small size/frequent shipments, which cause order processing inefficiency and excessively efficient inventory efficiency. The consistency of this pattern over time means that both partners in this VMI relationship have been over-cautious about inventory efficiency and did not value the efficiency that they could have gained from order processing. However, there is a significant upward trend in the second half of 2016 for A-class items which suggests a change in ordering patterns at the end of the year.

Another important trend to observe is that as item class changes from C to B and to A, the difference between the black line and the target level shrinks. This trend suggests that as the importance level of items increases, the replenishment pattern for the class becomes closer to the balanced trade-off level. Simply put, the most important reason behind this trend is that the demand for B-class and C-class items do not support tier quantity replenishment. Let's reiterate that the target line at 100 represents replenishments at the closest tier quantity (t_j) larger than the optimal order quantity ($q_{j,t}^*$). Oftentimes for B-class and C-class items $q_{j,t}^* \ll t_j$, thus when $q_{j,t}^*$ values get rounded up to $q_{j,t}^{con}$ for attaining order processing efficiency (i.e. $q_{j,t}^{con} = \left\lceil \frac{q_{j,t}^*}{t_j} \right\rceil t_j$), $q_{j,t}^{con}$ values become far larger than the optimal level (i.e. $q_{j,t}^{con} \gg q_{j,t}^*$) which is undesirable due to inventory management considerations such as expiration and etc.

To illustrate the substantial gap between $q_{j,t}^{con}$ and $q_{j,t}^*$ values for B-class and C-class items, we compute and plot the hypothetical OIE scores if we ordered optimal quantities for every item in every week. This can be performed by replacing $OF_{j,t}^{con}$ with $OF_{j,t}^*$ in Equation 3 and re-computing $\beta_{n,t}$ in Equation 4. As shown in Figure 8, the gap between the yellow lines and 100 is substantial for B-class and C-class items. The yellow line represents the optimal ordering pattern in a hypothetical situation. This suggests that on this particular lane, A-class items, which form 70% of the monetary transactions, are the most capable items for gaining order processing efficiency through replenishment adjustment. For B-class and C-class items, the gap between the yellow line and 100 is too much to compromise for gaining order processing efficiency by tier ordering.

Figure 8: Order processing metric scores for three item classes on an example lane over time. Black line represents the historical ordering pattern and yellow line represents the optimal ordering pattern is a hypothetical situation.

In order to go beyond the lane perspective for this metric and be able to monitor the inventory and order processing efficiencies at the supply chain network level, we propose a modified version of OIE metric. This metric is not to be observed over time and instead offers a picture over an extended period of time such as a quarter or longer. It has two components that represent the efficiency of operations with respect to inventory holding and order processing. The metric will be monitored on two axes, each representing a component, which form a quadrant plot. Starting with the inventory component, we define:

$$E_{j,t} = 100\left(1 - \frac{OF_{j,t}}{OF_{j,t}^{*}}\right), \qquad S_l = \frac{\sum_t \sum_{j\in n} \lambda_{j,t}\, E_{j,t}}{\sum_t \sum_{j\in n} \lambda_{j,t}} \tag{5}$$

where S_l denotes the inventory score for a lane over a period of time (x axis in Figure 9). Likewise, the order processing score (S_l') can be computed in a similar fashion (y axis in Figure 9):

$$E'_{j,t} = 100\left(1 - \frac{OF_{j,t}}{OF_{j,t}^{con}}\right), \qquad S_l' = \frac{\sum_t \sum_{j\in n} \lambda_{j,t}\, E'_{j,t}}{\sum_t \sum_{j\in n} \lambda_{j,t}} \tag{6}$$

A quadrant plot can visualize the metric and be useful to evaluate lanes based on their (S_l, S_l') scores. Figure 9 describes what each quadrant means but the key point is that the center of the graph ($S_l = 0, S_l' = 0$) represents the best scenario where $OF_{j,t} = OF_{j,t}^{*} = OF_{j,t}^{con}$. The center obviously is the ideal which can only happen hypothetically because it requires the optimal order quantity be equal to tier quantity for all the items in a lane. However, it is a valid reference to compare the relative position of other lanes to it. Moving away from center indicate worsening the conditions in the corresponding quadrant (Figure 9). The bounds on the graph are worth discussing because axes are bounded on the positive sides but not on the negative sides. The reason is that order frequency ($OF_{j,t}$) has a natural lower bound of zero but does not have any upper bound. Technically and empirically, $OF_{j,t}$ can be much larger than $OF_{j,t}^{*}$ or $OF_{j,t}^{con}$ which makes the axes on the graph boundless on the negative sides.

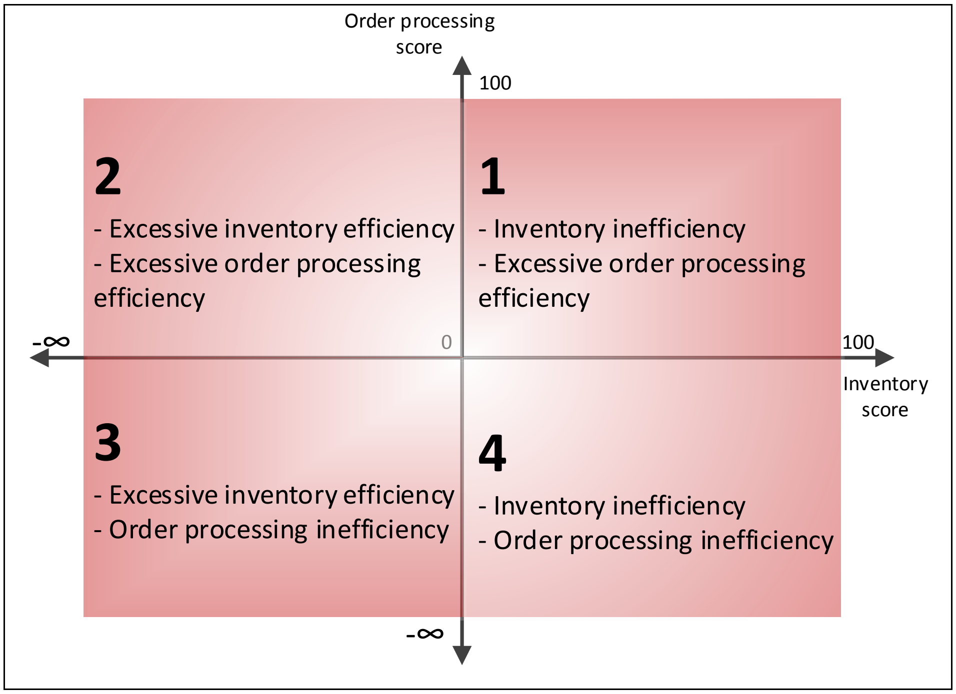

Figure 9: a schematic quadrant plot for the network-level OIE metric

Figure 10 illustrates the application of this metric on set of 20 lanes that support a VMI relationship between two partners in a healthcare supply chain network in 2016. It depicts the clear distinction in metric scores for the three item classes. This figure indicates that the pattern that we observed for a particular lane across its item classes (Figure 7) actually exists across the network. Another important trend to observe is the inclination of points toward the third quadrant along the y-axis. This indicate the overemphasis on inventory efficiency and lack of efficiency in order processing. The positive take from the figure is that A-class items, which form 70% of the monetary transactions on the network and 66.5% of the total volume (ft^3) transported, are spread around the center. Given the pattern present in Figure 10, solely looking at A-class items would also be insightful. Figure 11 illustrates that the demand size of lanes does not necessarily associate with better efficiency levels.

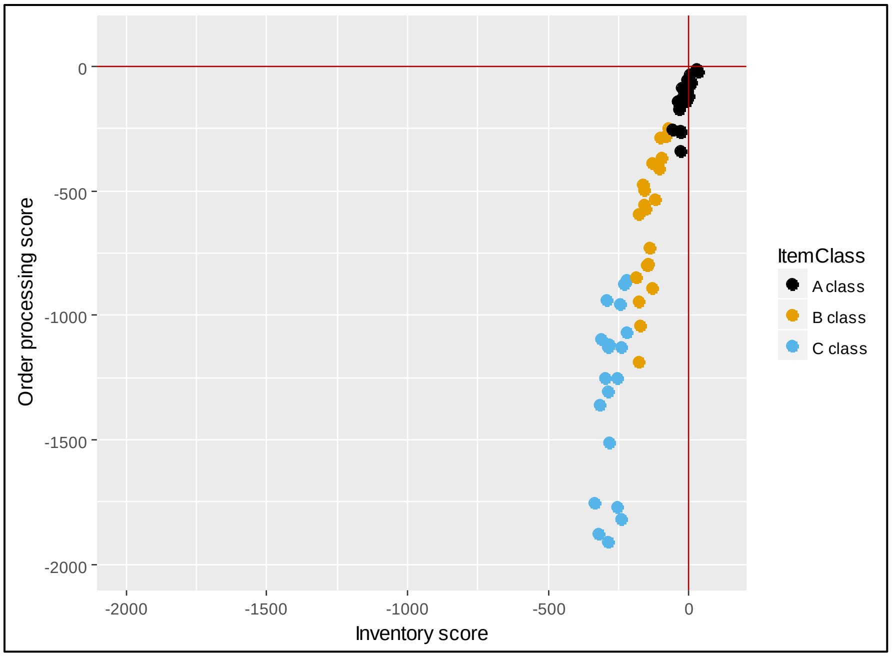

Figure 10: a quadrant plot of the network-level OIE metric for 20 major lanes

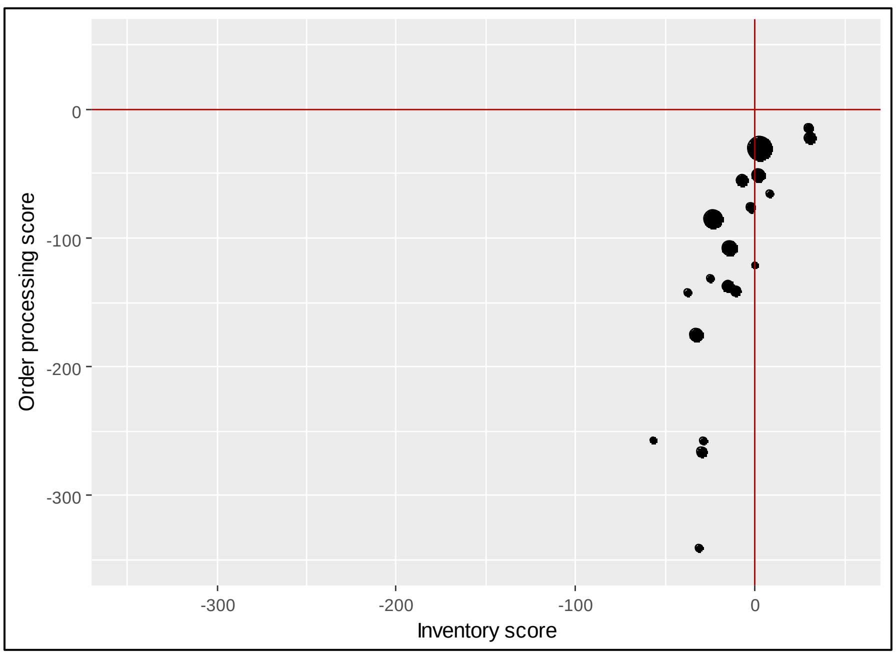

Figure 11: a quadrant plot of the network-level OIE metric for 20 major lanes (only A-class items). Size of the dots corresponds to the demand size of the lanes (Qty)

3.3.3 Transportation Cost Efficiency Metric

Transportation cost and specifically shipping cost depends on various factors. Companies that ship frequently, normally have year-long contracts with carriers that offer negotiated rates.

FTL carriers charge the shippers on a per mile basis while LTL and parcel carriers charge on a per load basis with rates that are primarily based on weight and distance. Although parcel shipments can only be 150 lbs. or lighter, FTL and LTL carriers accept shipments of any size. However, it is not economical to ship shipments that are larger than a size threshold with LTL or smaller than that threshold with FTL. Shippers normally determine the best threshold for mode selection based on various factors such as cost, product dimensions, and stackability. It is a well-known fact that FTL is a more cost-effective mode than LTL and LTL is a more cost-effective mode than parcel. In fact, the motivation behind initiatives such as shipment consolidation or multi-stop trucking is to benefit from better shipping rates of FTL shipments. The metric that we propose here focuses on the cost of shipping by incorporating the cost difference between shipping modes and other factors that affect shipping cost such as the space utilization of FTL trucks. This metric can be used to monitor cost efficiency over time; however, it is more meaningful to use at an aggregate level for evaluating the efficiency of transportation across the network. This would enable a comparison of lanes with each other over a long period of time.

The essence of this metric is the existing difference between the rates of FTL, LTL, and parcel. The main driver of metric score is proportion of shipped volume by each mode. Therefore, lanes with higher usage of FTL would have higher scores than lanes with predominantly LTL shipments. Another driver is space utilization of FTL trucks because FTL shipments become more economical as the utilization of space within the trucks increases. In fact, FTL would not be an economical mode if the available space in trucks remains substantially unutilized. Therefore, lanes with higher average FTL space utilization would have higher metric scores.

As mentioned above, shippers are charged by different rules depending on the transportation mode. FTL is per mile basis, LTL is per hundred pound (cwt) on a lane by lane basis, and parcel is per pound and distance. To make a comparison between modes, the pricing rules should be converted into a common one. Thus, we convert LTL and parcel rates to $/mile, which is the pricing rule of FTL. The conversion process uses lane distance and typical truck weight capacity (45000 lbs.) to change LTL and parcel rates to $/mile. Table shows this process.

d the distance (mile) of a lane

C_w the weight capacity of trucks on a lane (it is normally 45000 lbs.)

Table 1: Conversion of LTL and parcel costing rules to FTL cost rule

	FTL	LTL	Parcel
Pricing unit of measure	$/mile	cwt ($/100 lbs.)	$/lb/mile
Conversion - Step 1	$/mile	$\text{cwt}/(100\times d) = \$/\text{lb}/\text{mile}$	$/lb/mile
Conversion - Step 2	$/mile	$(\$/\text{lb}/\text{mile})\times C_w = \$/\text{mile}$	$(\$/\text{lb}/\text{mile})\times C_w = \$/\text{mile}$

Since the shipping rates are different from lane to lane, the savings impact of utilizing FTL shipments varies among the lanes. Therefore, the metric should capture this difference on a lane by lane basis by performing the rate conversion process for every lane separately. For the sake of simplicity and easy interpretation, this metric is designed to be a unit less metric that can vary from 0 to 100, where zero indicates the lowest efficiency and 100 indicates the maximum efficiency. The lowest efficiency is a situation where the entire demand is shipped via the most expensive shipping mode, which is normally the parcel mode. In contrast, the highest efficiency is a situation where the entire demand is shipped via the most economical shipping mode, which is 100% space utilized FTL's.

Thus, we need to rescale the converted shipping rates (i.e. $/mile rates in Table) to unit less numbers between 0 and 100 that represent the efficiency level of each shipping mode. These numbers are called *efficiency weights*. The rescaling process is performed for each lane separately and using a linear decreasing function shown in Figure 12.

r_i the \$/mile rate of transportation mode i on a lane. $i \in \{FTL, LTL, Parcel\}$

W_i the efficiency weight of mode i on a lane. $i \in \{FTL, LTL, Parcel\}$

$P_{i,t}$ the percentage of shipped volume on a lane via mode i in week t.

$\bar{u}_t$ the average space utilization of FTL trucks in week t.

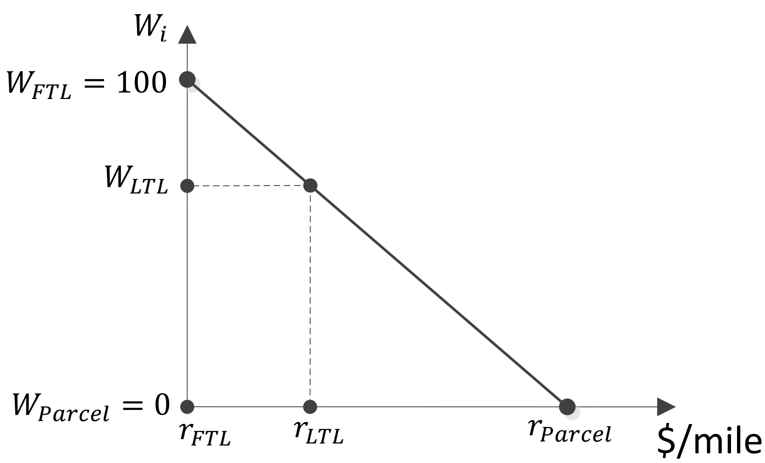

Figure 12: Rescaling linear function to transform \$/mile shipping rates (r_i) to efficiency weights (W_i)

As discussed earlier, space utilization of truck is another efficiency driver of transportation operations. Truck space is an available capacity for shippers and if not used, it is still paid for. FTL carriers charge shippers on dollar per mile basis thus any empty space in a truck is a lost opportunity for shippers. This is only applicable for FTL shipments since LTL and parcel cost structure is on a dollar per pound basis. In order to reflect this in the metric, efficiency weight of FTL should be reduced by the lost space in the shipped trucks. Therefore, average space utilization of FTL shipments over time period *t* (i.e. $\bar{u}_t$) will be multiplied to efficiency weight of FTL, which is originally set at 100. Once the efficiency weights (W_i) are obtained for a lane, the metric score can be computed over time (e.g. weekly) by multiplying the percentage usage of modes over time period t ($P_{i,t}$) to the corresponding efficiency weights (W_i). Since shipping rates do not change frequently (i.e. they usually change on a yearly basis due to contact renewals), r_i and W_i do not have time subscripts.

ϕ_t the value of transportation cost efficiency metric in week t.

$$\phi_t = \bar{u}_t W_{FTL} P_{FTL,t} + W_{LTL} P_{LTL,t} + W_{Parcel} P_{Parcel,t}$$

For monitoring the cost efficiency of transportation on a single lane, ϕ_t can be tracked over time. One difference that this metric has in comparison with the previous ones, is that there is no common target level for all the lanes. ϕ_t values of low volume lanes are expected to be less than high volume lanes because their demand does not support FTL shipments, therefore they are not able to achieve high scores in this metric. Thus, we expect to see a correlation between demand and ϕ_t values across lanes within a supply network (Figure 13).

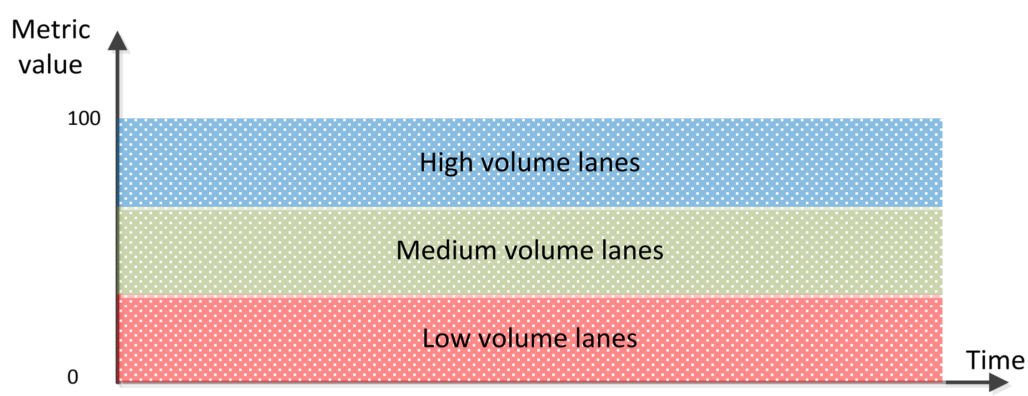

Figure 13: Expected values of ϕ_t over time for different size lanes

Lastly, if *t* is assumed to be a long period of time (e.g. 6 months or one year) then an overall efficiency level for lanes can be computed, analyzed and plotted on a single graph to evaluate the efficiency of the entire network over a long period of time. Figure 14 illustrates the transportation cost efficiency scores of 143 lanes over a period of 6 months. Each point on the graph represents a lane in a healthcare supply network that support major distributors across the

United States. The most noticeable pattern is an increase in scores (y axis) as demand increases (x axis). More importantly, the graph shows that lanes that are managed in a continuous replenishment program (CRP) performed significantly better over the 6 months period. CRP gives the replenishment responsibility to the supplier which has resulted in a better consolidation achievement. This metric is a good complement to the other two metrics that are discussed earlier because first it only focuses on cost and second it provides a big picture on the transportation efficiency.

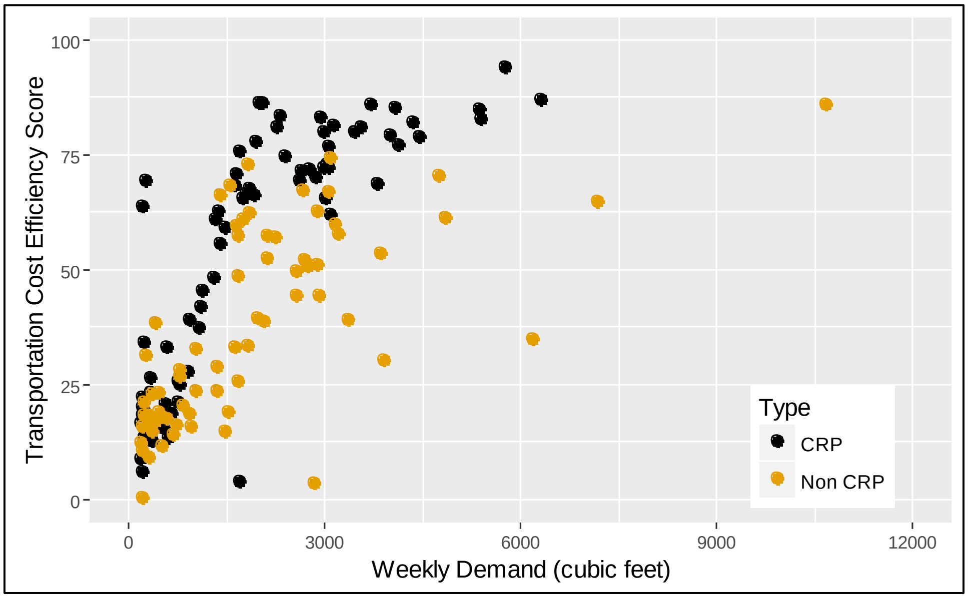

Figure 14: Transportation cost efficiency scores of 143 different lanes over a period of 6 months

3.4 Statistical Process Control System

The purpose of this section is to develop statistical process control (SPC) systems to monitor the metrics over time. An SPC system ensure whether a time-series maintains a state of statistical control or any departure from statistical control has occurred. A state of statistical control is known as a process that generates independent and identically distributed (iid) random variables. Departures from statistical control are discovered by plotting data on a variety of control charts. Departures are signals to search for special causes that might cause out of control situations. We present a framework for appropriate SPC approaches for different time-series behaviors of the developed metrics. We will discuss time-series models for SPC under different scenarios including situations where independent and identically distributed observations do not exist. Oftentimes in practice, including in logistics, the presence of autocorrelations, non-stationarity and other time-series effects limit the applicability of traditional control charts. In this section, we briefly discuss traditional SPC for independent and identically distributed observations. We then discuss developing control charts for dependent, stationary and not necessarily normally distributed observations. Finally, we present a detailed framework for screening autocorrelated and non-stationary time-series which are commonly encountered in practice including in our case.

The metrics developed in sections 3.3.1 and 3.3.2 generate time-series observations that should be monitored over time to detect any significant process mean change to subsequently call for corrective action. The metrics have a target or optimal level (i.e. $\mu_0 = 100$) that should be used to monitor the process mean μ against it. The Shewhart $\bar{X}$ chart is a useful SPC tool to accomplish this objective. This chart plots the sample means, $\bar{X}$'s, of subgroups of the individual observations $\{X_1, X_2, \ldots\}$ and is essentially testing the hypotheses $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ conducted over time, using $\bar{X}$ as the test statistic. Failure to reject the null hypothesis (H_0) indicates in control process, otherwise, it is said to be out of control. This decision mechanism can be graphically displayed on a control chart with an *upper control limit (UCL)* and a *lower control limit (LCL)* extended on a horizontal axis which indicates the time order of the observed test statistic, $\bar{X}$. The *y* axis indicates the metric value and as time elapses, we observe the metric values, compute sample mean ($\bar{X}$), and plot them on the chart. The region between LCL and UCL is the acceptance region of H_0 and whenever $\bar{X}$ falls outside the control limits, it suggests that the process mean (μ) might have significantly shifted from the target (μ_0) due to some assignable cause. In order to construct the control charts, which requires LCL and UCL, we need to estimate the process mean and variance using some historical data taken when the process is considered in control. Let's assume the data consists of *k* samples of size *n* where we denote the *i*th sample mean and variance by $\bar{X}_{\mathrm{i}}$ and S_i^2. Then the estimated process mean ($\bar{\bar{X}}$) and variance (S^2) are:

$$\bar{\bar{X}} = \sum_{i=1}^{k} \bar{X}_i \qquad and \qquad S^2 = \sum_{i=1}^{k} \frac{S_i^2}{k} \tag{7}$$

Assuming $\{X_1, X_2, \ldots\}$ follows the normal distribution, the control limits for the $\bar{X}$ chart are approximately:

$$\begin{aligned} LCL &= \bar{\bar{X}} - z\left(1 - \frac{\alpha}{2}\right) S/\sqrt{n} \\ UCL &= \bar{\bar{X}} + z\left(1 - \frac{\alpha}{2}\right) S/\sqrt{n} \end{aligned} \tag{8}$$

where α is the false alarm rate and z_α indicates the α quantile of the standard normal distribution. The center line for the $\bar{X}$ chart is obviously $\bar{\bar{X}}$ which is the average of all X_i's. For more information see Montgomery (2009). The two key assumptions here are normal distribution for observations and independency between them. If the metric values found to be independent or un-correlated over time and follow a normal distribution then the presented limits can be applied to monitor them on a $\bar{X}$ chart. The only limitation of this method in our case is that it requires sampling. Sample size (*n*) normally needs to be greater than two and less than or equal to five. Since our metrics generate weekly observations, for n = 4 one needs to wait and collect data for four weeks to create one sample observation for the control chart. To overcome this issue, we suggest the *moving range* chart which is designed for monitoring individual measurements. The moving range is defined as $MR_i = |x_i - x_{i-1}|$, which is the absolute value of the first difference of the data. The control limits of the individual observations chart are:

$$LCL = \bar{\bar{X}} - 3\frac{\overline{MR}}{1.128} \qquad , \qquad UCL = \bar{\bar{X}} + 3\frac{\overline{MR}}{1.128}$$

where $\overline{MR}$ is the average of all the moving ranges of two observations. The value of 1.128 is the unbiasing constant (d_2), read from table for sample size of two, to make $\frac{\overline{MR}}{1.128}$ an unbiased estimator of standard deviation (Montgomery, 2009).

Normality and independency of observations are two major assumptions that often do not hold in practice. We often encounter observations such as stock market indices that are statistically dependent or correlated. Later in this section we show that these two assumptions do not hold for the developed metrics in this chapter as well. When autocorrelation exists in a time-series, the true variance of $\bar{X}$ involves covariances between X_i's that are not well-estimated by the pooled sample variance S^2 in Equation 7. When the normality of observations is the only assumption in question, the bootstrap method can be used to compute control limits (Liu and Tang, 1996). If the observations are neither independent nor normal but they are only "weakly dependent", a modification of the original bootstrap method, called moving block bootstrap method can be used to develop a valid $\bar{X}$ control chart. Liu and Tang (1996) discusses how to develop such control charts using the bootstrap method and moving block bootstrap method in great detail. Weakly dependent refers to a situation where the correlation between X_i and X_{i+h} tends towards zero sufficiently quickly as h goes to infinity (i.e. $corr(X_i, X_{i+h}) \rightarrow 0$ as $h \rightarrow \infty$). This essentially requires a stationary time series, which is again does not always appear in practice. Therefore, moving block bootstrap would not be able to develop valid control limits where non-stationary autocorrelated time-series exist. This is a common pattern appearing in practice such as in U.S. electricity consumption data or in oil sales price records (Cryer and Chan, 2008). Later we show that autocorrelated non-stationary is the most common pattern appearing for our metrics in the case study.

When systematic time-series effects, such as autocorrelation or non-stationarity, are present, standard Shewhart control charts lead to a substantial chance of not detecting special causes that truly exist while observing apparent special causes that do not exist. The main reason is presence of common causes that appear in the form of trend or seasonality or generally speaking autocorrelation behavior. Standard Shewhart control charts, such as the $\bar{X}$ chart, are able to signal special causes when the common cause is an iid process. Alwan (1992) studies the effect of autocorrelation on the standard Shewhart charts for individual observations with fixed (i.e. 3-sigma) control limits. They discuss that in a non-iid process, where the mean is constantly changing, using a single chart would require to continually move the control limits centered around the estimated conditional means where the width of limits is calculated from the variation of residuals. However, based on the literature, the most common method for developing control charts in the case of autocorrelated and non-stationary observations is presented in Alwan and Roberts (1988). Their approach suggests modeling systematic non-random behavior by using the autoregressive integrated moving average (ARIMA) models which ultimately leads to two basic charts rather than one:

I) Common Cause Chart (CCC): a chart of fitted values based on ARIMA models. It provides guidance in understanding the process. This chart does not have control limits and instead provides a representation of the current and estimated state of the process.

II) Special Cause Chart (SCC): a standard $\bar{X}$ chart of residuals from fitted ARIMA models. All traditional settings of the process control for iid observations are applicable here.

Although the CC chart does not have limits, Wardell et al. (1992) propose control limits on the CC chart in the case of ARMA (1,1) which is a stationary process.

ARIMA models aim to describe the autocorrelations in the data. A time series $\{Y_t\}$ is said to follow an autoregressive integrated moving average (ARIMA) model if the *d*th difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. In other words, if $\{W_t\}$ follows an ARMA *(p,q)* model, we say that $\{Y_t\}$ is an ARIMA *(p,d,q)* process. For practical purposes, the first or at most the second difference (i.e. *d = 1 or 2*) is sufficient to create a stationary time series (Cryer and Chan, 2008). Let's assume $\{W_t\}$ is the first difference of $\{Y_t\}$ ($W_t = Y_t - Y_{t-1}$) then ARIMA*(p,1,q)* will be:

$$W_t = c + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (9)$$

where p is the order of the autoregressive part, q is the order of the moving average part and $\{e_t\}$ represent an unobserved white noise series, that is, a sequence of identically distributed, zero-mean, independent random variables. The procedure discussed in Alwan and Roberts (1988) proposes fitting such a model (Equation 9) to the time series, using the fitted values as the CC chart, and using the residuals to build an $\bar{X}$ control chart. The residuals are the difference between the observations and the corresponding fitted values:

$$e_t = y_t - \hat{y}_t \quad (10)$$

If an ARIMA model with a good fit is selected, then the residuals are expected to be uncorrelated with mean of zero; otherwise the ARIMA model needs to be improved. Later, we will discuss developing the CCC and SCC charts for a time series of the transportation and inventory efficiency metric (TIE), discussed in section 3.3.1, over two years on a channel instance.

Consider a demand channel from a supplier DC in Atlanta, GA to a customer DC in Montgomery, NY. We have computed the TIE metric for two years on a weekly basis (Figure 15). As the autocorrelation function graph (ACF) and the partial autocorrelation function graph (PACF) show, there is a strong autocorrelation in the time series. They illustrate correlation between lagged values of the time series (r_k where k is the lag). In an iid process, where there is no autocorrelation, both ACF and PACF spikes are supposed to lie within the critical limits $\pm 2/\sqrt{T}$ where *T* is the length of the time series. The ACF also indicates that the time series is non-stationary because the spikes decrease slowly with a large and positive value of r_1. In a stationary process, ACF spikes drop to zero relatively quickly and do not go beyond the significance limits (Hyndman and Athanasopoulos, 2014). In addition to interpreting the ACF plot, we can perform a more formal test, called the Ljung-Box test, which considers a whole set of r_k values as a group, rather than treating each one separately. The test statistic is the following:

$$\omega = T(T+2)\sum_{k=1}^{h}(T-k)^{-1}r_k^2 \tag{11}$$

If there was no significant autocorrelation, then ω would have a χ^2 distribution with (*h* – *K*) degrees of freedom, where *h* is being the maximum lag being considered and *K* is the number of parameters in a time-series model. Since we are applying it on a raw data rather than the residuals from a model, *K* is equal to zero. The value of *h* is suggested to be $T/5$ but not larger than 10 for achieving the best performance. Using the *stats* package in R we performed the test on the TIE time-series which resulted in the p-value of 2.9×10^{-7} which confirms significant autocorrelation in the time series.

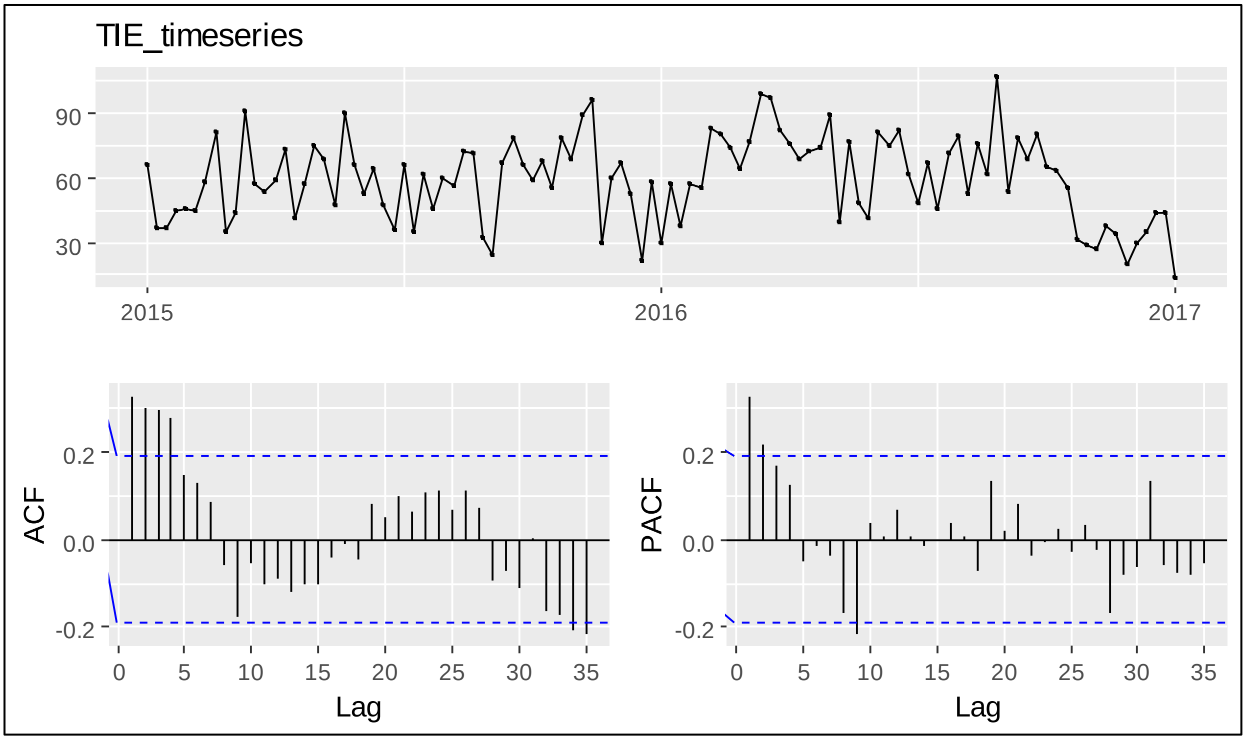

Figure 15: Transportation and Inventory Efficiency (TIE) metric time series, the corresponding autocorrelation function chart (ACF) and partial autocorrelation function chart (PACF)

In order to fit an ARIMA model, the order of the model, which is the appropriate values for *p*, *d*, and *q*, need to be selected. In addition, the parameters $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ need to be estimated (Equation 9). Order selection and parameter estimation of ARIMA models are based on using *maximum likelihood estimation (MLE)* technique and the *Akaike's information criterion (AIC)* as the performance indicator. AIC is helpful in determining the order of an ARIMA model and can be written as:

$$AIC = -2\log(L) + 2(p + q + k + 1) \tag{12}$$

where *L* is the likelihood of the data and the last term in parenthesis is the number of parameters in the model (including the variance of residuals). *k* indicates if the model has a constant *c* or not hence, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$. Statistical software packages obtain the order of ARIMA models by minimizing the *AIC* or *AICc* which is known as the corrected *AIC*, a variation of the original *AIC* presented in Equation (12) (Hyndman and Athanasopoulos, 2014). Once the order of an ARIMA model is identified, the parameters $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ need to be estimated. MLE is the technique that is commonly used to estimate the parameters by maximizing the probability of obtaining the data that has been observed. Since the logarithm of the likelihood function is more convenient to work with than the likelihood itself, for given values of *p*, *d*, and *q*, statistical packages try to maximize the log-likelihood when estimating the parameters (Cryer and Chan, 2008).

There are certain rules that are helpful to determine the order of an ARIMA model. Rules that are based on the patterns of ACF and PACF graphs are fairly subjective and not decisive in determining the best *p*, *d*, and *q* values of an ARIMA model. There have been several attempts to automate ARIMA modeling in the past few decades using different approaches. Some of them have been implemented in commercial software packages; for example, Gómez (1998) proposed an automatic method for multiplicative seasonal ARIMA modeling for TRAMO and SEATS software. (Liu, 1988) developed another automatic method based on a filtering method and heuristic rules for seasonal ARIMA, which then used in the SCA-Expert software. Another approach for univariate ARIMA modeling that allows intervention analysis is developed by Mélard and Pasteels (2000) for software package TSE-AX. Perhaps Forecast Pro (Goodrich, 2000) is the most well-known commercial software for its excellent automatic ARIMA algorithm, however it has not been documented publicly.

We use one of the recent automatic ARIMA modeling algorithm, known as the Hyndman-Khandakar algorithm, which is implemented in the *forecast* package in R. As discussed thoroughly in Hyndman and Khandakar (2008), this algorithm determines the number of differences *d* in an ARIMA model by using the *Kwiatkowski-Phillips-Schmidt-Shin (KPSS)* test. The KPSS test is a hypothesis test of stationarity to determine whether differencing is required to achieve a stationary time-series. This algorithm determines the values of *p* and *q* by minimizing the AICc measure and using a stepwise search to traverse the model space. The use of stepwise search and some approximations is to speed up the search; therefore, it is possible that the model with minimum AICc will not be identified. However, the `auto.arima()` function allows turning off the approximation (`approximation=FALSE`) or the stepwise search (`stepwise=FALSE`). It is also helpful to choose the model based on subjective judgment (using `Arima()` function) and compare its AICc with the result of the `auto.arima()` function. It is important to note that `auto.arima()` or `Arima()` functions perform as expected when a time-series with stable variation is passed. If a time-series shows variation that increases or decreases with the level of the series, then a logarithmic or power transformation should be applied to stabilize the data before developing a model. Section 3.2 of Hyndman and Athanasopoulos (2014) discusses transformations in further detail. Note that in case of seasonality in the data, seasonal ARIMA models that include additional seasonal terms should be developed. Once a model is chosen, it is necessary to plot the ACF of the residuals and ensure that they look like white noise process.

Now let’s consider the TIE time-series illustrated in Figure 15 and develop an ARIMA model. We first look at the output of `auto.arima()` function. Since the time-series does not show any instability in variation over time, it does not need any transformation therefore we can pass the data directly to the `auto.arima()` function. We also use the function when no approximation or stepwise search is used.

```
#1 auto.arima(TIE)

#2 auto.arima(TIE, approximation = FALSE, stepwise = FALSE)
```

They resulted in ARIMA (3,0,2) with AICc of 909.13 and ARIMA (2,0,2) with AICc of 907.93 respectively. As they both indicate, $d = 0$ which means the algorithm recognizes this time-series as stationary with no need for differencing. Although the time-series does not show significantly visible non-stationarity, the ACF graph hints the other way. The first difference of the TIE time-series is illustrated in Figure 16.

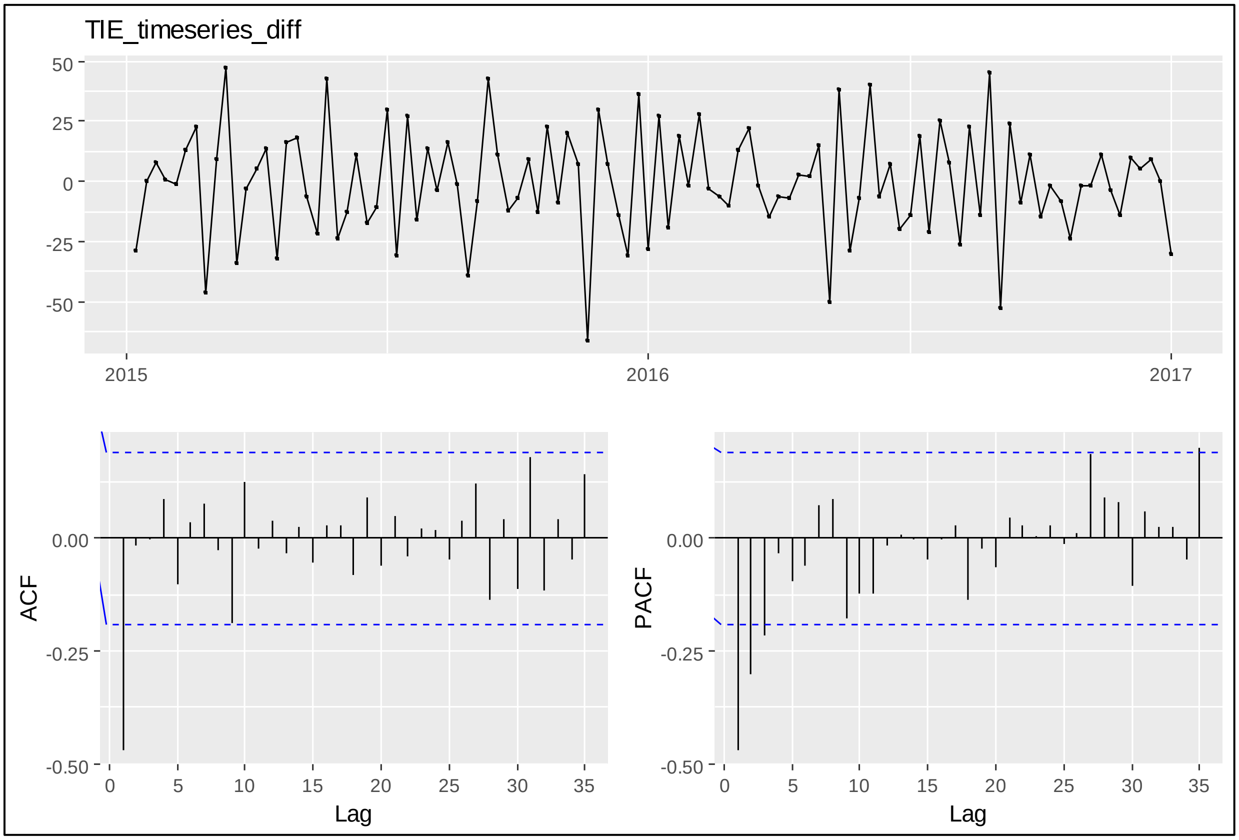

Figure 16: The first difference of the transportation and inventory efficiency (TIE) metric

Both the differenced time-series and its ACF and PACF graphs have much stronger characteristics of a stationary process. Given the first difference is improving, we can use the following visual rules of ARIMA modeling (Hyndman and Athanasopoulos, 2014) to select an appropriate order for the model:

- The data may follow an ARIMA (*p*,*d*,*0*) model if the ACF and PACF plots of the differenced data show the following patterns:

  - the ACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag *p* in the PACF, but none beyond lag *p*.
- The data may follow an ARIMA (*0,d,q*) model if the ACF and PACF plots of the differenced data show the following patterns:
  - the PACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag *q* in the ACF, but none beyond lag *q*.

Based on the above rules, ARIMA (0,1,1) seems an appropriate model. Developing ARIMA (0,1,1) using `Arima()` function results in the AICc of 900.35, which is better than the outputs of the automatic algorithm. Other models such as ARIMA (1,1,0) and ARIMA (1,1,1) are also tested but they all result in higher AICc levels. After all, it is reassuring that Alwan and Roberts (1988) argues that precise model identification may not be essential to effective process control because several alternative ARIMA models may fit the data about equally well. The `Arima()` function sets $c = 0$ by default (Equation 9) but by using argument `include.drift = TRUE` we can include the constant *c*. For the current example, including a constant results in a higher AICc measure which is undesirable. It is also noteworthy that `auto.arima()` automates the inclusion of a constant. Given that the first order integrated moving average, ARIMA (0,1,1), is the superior model and assuming $\{Y_t\}$ is the original series, the model equation is:

$$Y_t = Y_{t-1} + e_t - 0.7112e_{t-1} \tag{13}$$

At this stage, we should ensure that the residuals of the model form an iid process with mean of zero. Figure 17 plots the residuals, their histogram, and the ACF graph. The residuals appear to be stationary and following a normal distribution with mean of zero. However, the ACF graph has one spike on lag 9 that goes beyond the significance limits. Therefore, as discussed earlier, we can use the Ljung-Box test to examine if there is autocorrelation remained in the residuals. The result of the test indicates that there is no significant autocorrelation in the residuals (p-value = 0.3724).

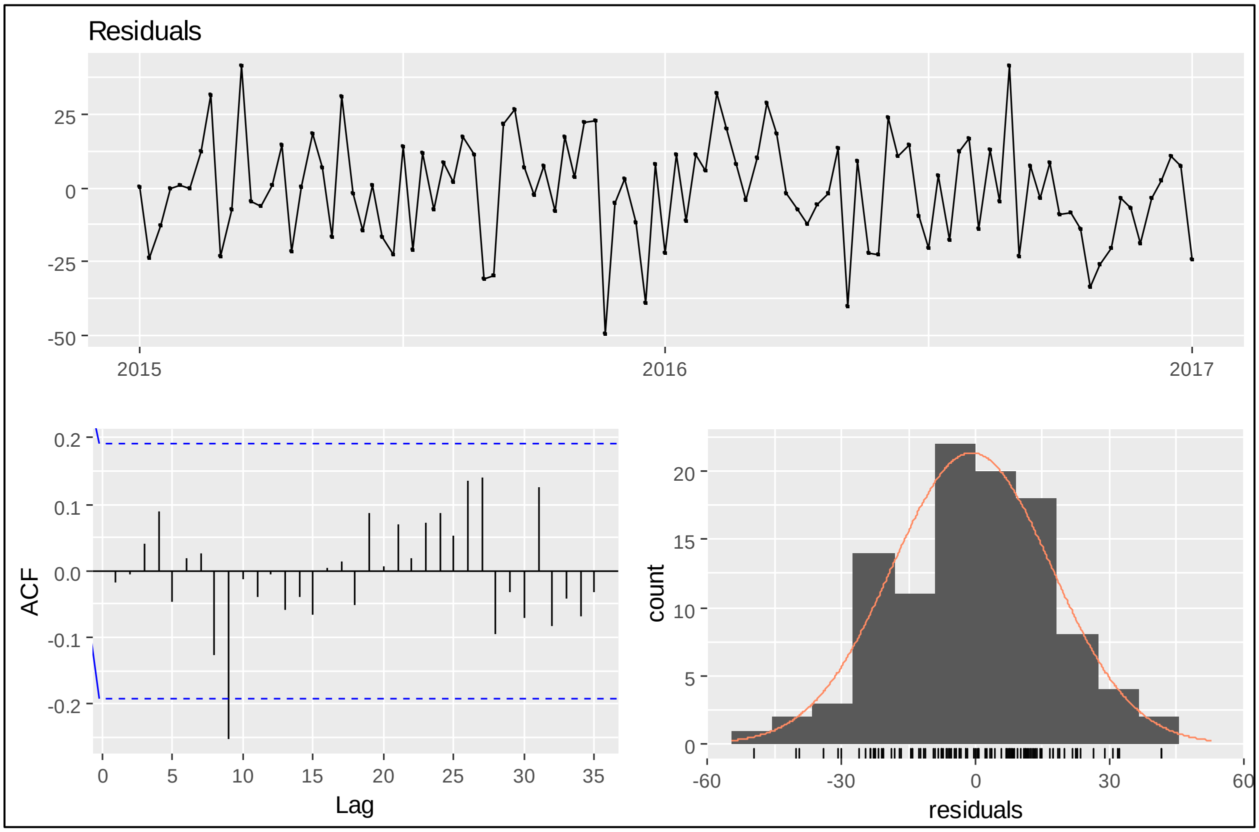

Figure 17: The residuals of the fitted model, ARIMA (0,1,1)

Since an appropriate ARIMA model is identified for the TIE data of this particular lane to Montgomery NY, we can develop the common cause chart (CCC) and the special cause chart (SCC) as discussed before. The CC chart is a chart of fitted values based on ARIMA (0,1,1), which helps in understanding the process by providing a representation of the current and estimated state of the process. The CC chart presents a view of the level of the process and its evolution over time. We know that the most desirable level of the process is 100, which represents the optimal balance between transportation and inventory holding efficiencies. Deviations from that level entail economic loss caused by supply chain management. For this particular lane, the CC chart (Figure 18) clearly shows that inventory efficiency has been consistently favored over transportation efficiency over year 2015 and 2016 (see Figure 2). There has not been a single week that the TIE score is above 100 to indicate an over-consolidation of shipments to gain transportation efficiency.

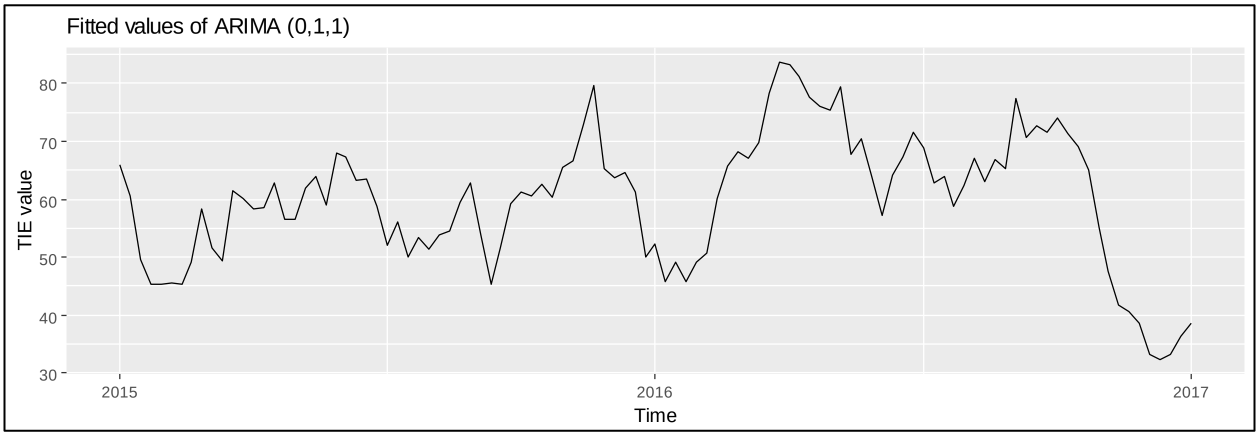

Figure 18: The Common Cause Chart (CCC) for the TIE values of the instance channel

The process shows that small shipments have been sent frequently to keep the inventory levels low which in turn caused losing shipment consolidation opportunities. The two important visible patterns in the CC chart are: 1) the presence of two peaks in the process around months 3 and 9 in each year and 2) a slight increase in the level of the process from year 2015 to 2016. One speculation is whether the demand pattern drives the TIE values, meaning when demand rises, more consolidation opportunities causes TIE values to increase. Looking at Figure 19, and comparing it to the CC chart, there is no visible indication of correlation between them. In fact, the correlation coefficient of 0.034 confirms this fact as well.

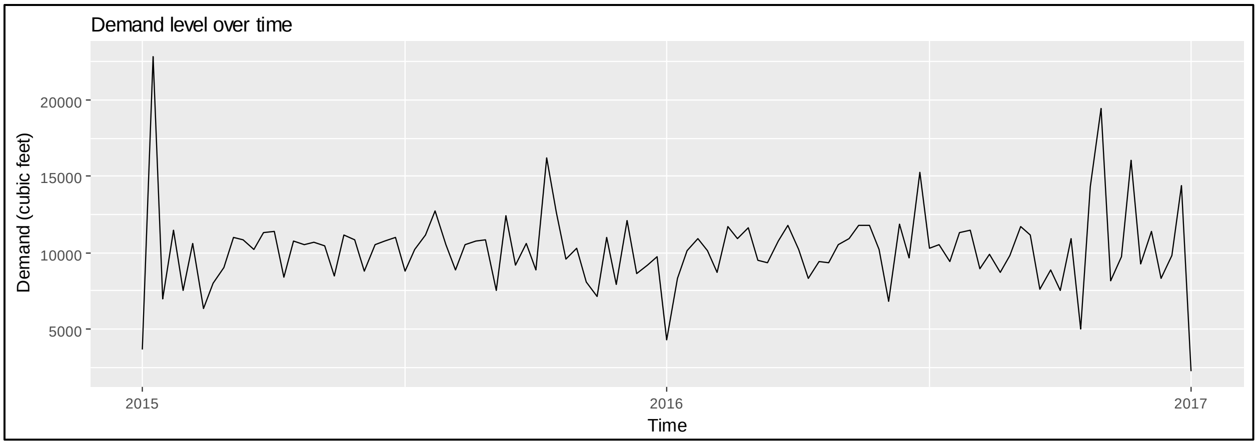

Figure 19: Weekly demand level of the channel to Montgomery NY

The SC chart is a chart of residuals from the fitted ARIMA model. In order to construct the SC chart, the control limits need to be computed. Since the process is iid and we want to monitor individual observations, we should use the moving range control chart as discussed earlier.

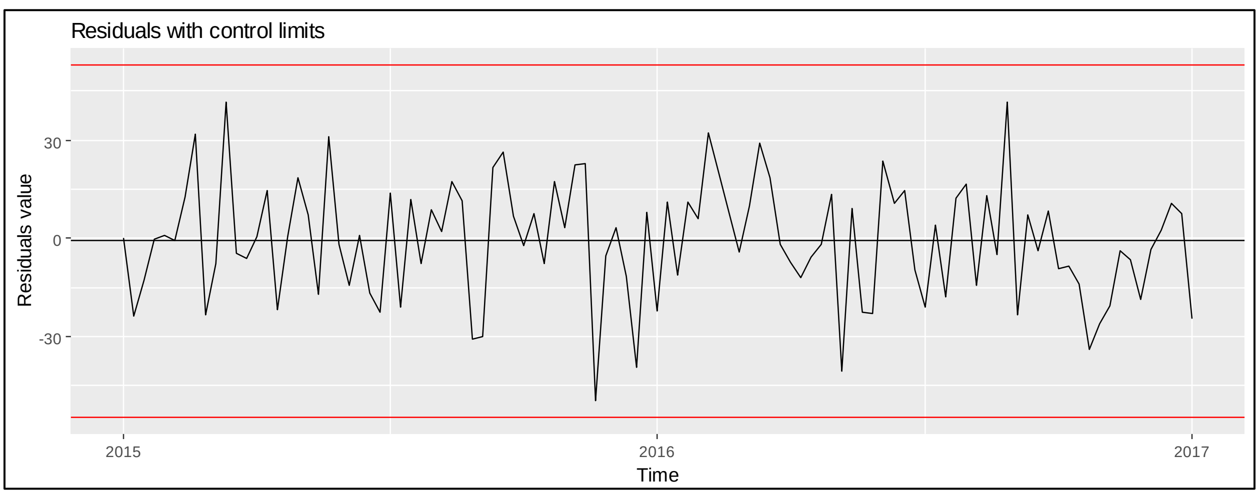

Figure 20: The Special Cause Chart (SCC) for the TIE values of the instance channel

$$LCL = \bar{\bar{X}} - 3\frac{\overline{MR}}{1.128} = -0.8092 - 3\frac{20.1623}{1.128} = -54.53$$

$$CL = \bar{\bar{X}} = -0.8092$$

$$UCL = \bar{\bar{X}} + 3\frac{\overline{MR}}{1.128} = -0.8092 + 3\frac{20.1623}{1.128} = 52.81$$

This chart is essentially a standard control chart for residuals and can be used in traditional ways to detect any special causes, without the risk of confusing special causes with common causes. While the CC chart shows the underlying process, the SC chart detects sudden and substantial shocks due to assignable causes rather than common causes. In addition to considering observations outside the *UCL* and *LCL* as out of control, run count rules should also be used to detect out of control situations (Montgomery, 2009). In the example illustrated in Figure 20, there is no indication of out of control situations. This means that the focus should be on identifying the common causes that drive the pattern existing in the CC chart.

3.5 Summary

As discussed earlier, monitoring efficiency metrics are essential for logistics excellence in a supply chain system. This study has offered new metrics that can capture the trade-off in achieving efficiency between distinct functions of logistics. Logistics functions that are considered are transportation, inventory holding, and order processing, which greatly contribute to the profitability of any distribution operation. In addition, optimal trade-off levels are proposed for each metric as well as a statistical process control system to monitor them over time. We discussed how to utilize appropriate statistical methods for various time-series behaviors that might appear for the metrics. Another key contribution of this study is providing data-driven and traceable metrics that quantify possible tendencies in favoring efficiency in certain functions over others. This is particularly helpful for enhancing communications between partners in supply chain collaboration programs, such as VMI. Such programs attract potential partners for different reasons, thus, each partner might seek a different or sometimes competing objective. The metrics that are proposed here enable decision makers to communicate factually and work collectively toward balanced efficiency levels using statistical control systems.

3.6 Future Work

This research study introduces a new approach to supply chain efficiency screening and performance measurement. It opens new research directions for future to develop a better understanding about the impacts of using such metrics, its applications in other areas besides the supply chain management. The following are the main specific directions for future research that we recommend:

- Conducting sensitivity analysis on the parameters in the metrics in order to understand their impact on the metrics,
- Conducting a correlation analysis for metrics to identify the circumstances under which they provide redundant information,
- Conducting a case study to apply the metrics on datasets from both VMI relationships and non-VMI relationships. This would evaluate the ability of the metrics to capture the difference in efficiency of different supply chain arrangements,
- Developing effectiveness metrics that can provide similar trade-off representation in across different logistics functions,
- Investigating whether companies see the optimal trade-off levels aligned with their strategic objectives and other considerations. What could be the rationale behind favoring the efficiency of one logistics function (e.g. inventory holding) over others?
- Assess the “goodness” of metrics using the evaluation criteria that are discussed in the literature for logistics metrics (Caplice and Sheffi, 1994). Collecting inputs from operations and logistics managers from different industries will be insightful.

References

Agrawal, V., Seshadri, S., 2000. Distribution free bounds for service constrained (Q, r) inventory systems. Nav. Res. Logist. 47, 635–656. doi:10.1002/1520-6750(200012)47:8<635::AID-NAV3>3.0.CO;2-C

Alwan, L.C., 1992. Effects of autocorrelation on control chart performance. Commun. Stat., Theory Methods. doi:10.1080/03610929208830829

Alwan, L.C., Roberts, H. V, 1988. Time-Series Modeling for Statistical Process Control 6, 87–95.

Barratt, M., 2004. Understanding the meaning of collaboration in the supply chain. Supply Chain Manag. An Int. J. 9, 30–42. doi:10.1108/13598540410517566

Beamon, B.M., 1999. Measuring Supply Chain Performance. Ind. Eng. 19, 275–292. doi:http://dx.doi.org/10.1108/MRR-09-2015-0216

Bullinger, H.-J., Kühner, M., Van Hoof, A., 2002. Analysing supply chain performance using a balanced measurement method. Int. J. Prod. Res. 40, 3533–3543. doi:10.1080/00207540210161669

Caplice, C., Sheffi, Y., 1995. A Review and Evaluation of Logistics Performance Measurement Systems. Int. J. Logist. Manag. 6, 61–74. doi:10.1108/09574099510805279

Caplice, C., Sheffi, Y., 1994. 1994 (MIT) A Review and Evaluation of Logistics Metrics.pdf. Int. J. Logist. Manag.

Çetinkaya, S., Tekin, E., Lee, C.-Y., 2008. A stochastic model for joint inventory and outbound shipment decisions. IIE Trans. 40, 324–340. doi:10.1080/07408170701487989

Chan, F.T.S., 2003. Performance Measurement in a Supply Chain. Int. J. Adv. Manuf. Technol. 21, 534–548. doi:10.1007/s001700300063

Chow, G., Heaver, T.D., Henriksson, L.E., 1994. Logistics Performance: Definition and Measurement. Inernational J. Phys. Distrib. Logist. Manag. 24, 17–28.

Cryer, J.D., Chan, K.-S., 2008. Time Series Analysis with Applications in R, 2nd ed. Springer.

Disney, S.M., Potter, A.T., Gardner, B.M., 2003. The impact of vendor managed inventory on transport operations. Transp. Res. Part E Logist. Transp. Rev. 39, 363–380. doi:10.1016/S1366-5545(03)00014-0

Fugate, B., Mentzer, J., Stank, T., 2010. Logistics Performance : Efficiency , Effectiveness , and Differentiation. J. Bus. Logist. 31, 43–62.

Gleason, J., Barnum, D., 1982. Toward Valid Measures of Public Sector Productivity : Performance Measures in Urban Transit. Manage. Sci. 28, 379–386.

Gómez, V., 1998. Automatic model identification in the presence of missing observations and outliers. Ministerio de Econom{’\i}a y Hacienda, Direcci{ó}n General de An{á}lisis y Programaci{ó}n Presupuestaria.

Goodrich, R.L., 2000. The forecast pro methodology.

Gunasekaran, A., Kobu, B., 2007. Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. Int. J. Prod. Res. 45, 2819–2840. doi:10.1080/00207540600806513

Hyndman, R.J., Athanasopoulos, G., 2014. Forecasting: principles and practice. OTexts.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. J. Stat. Softw. 27, C3–C3. doi:10.18637/jss.v027.i03

Kaplan, R.S., Norton, D.P., 2005. The balanced scorecard: Measures That drive performance. Harv. Bus. Rev. doi:00178012

Kleijnen, J.P.C., Smits, M.T., 2003. Performance metrics in supply chain management. J. Oper. Res. Soc. 54, 507. doi:10.1057/palgrave.jors.2601539

Kohli, A.S., Jensen, J.B., 2010. Assessing Effectiveness of Supply Chain Collaboration: An Empirical Study. Supply Chain Forum An Int. J. 11, 2–16.

Lambert, D.M., Pohlen, T.L., 2001. Supply Chain Metrics. Int. J. Logist. Manag. 12, 1–19. doi:http://dx.doi.org/10.1108/09574090110806190

Langley Jr, C.J., Holcomb, M.C., 1992. Creating Logistics Customer Value. J. Bus. Logist. 13, 1.

Lee, H.L., Padmanabhan, V., Whang, S., 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. Manage. Sci. 43, 546–558. doi:10.1287/mnsc.43.4.546

Liu, L.-M., 1988. Identification of seasonal arima models using a filtering method. Commun. Stat. - Theory Methods. doi:10.1080/03610928908830035

Liu, R.Y., Tang, J., 1996. Control Charts for Dependent Measurements Based on Bootstrap Methods. J. Am. Stat. Assoc. 91, 1694–1700.

Mélard, G., Pasteels, J.-M., 2000. Automatic ARIMA modeling including interventions, using time series expert software. Int. J. Forecast. 16, 497–508. doi:10.1016/S0169-2070(00)00067-4

Mentzer, J.T., Konrad, B.P., 1991. An efficiency/effectiveness approach to logistics performance analysis. J. Bus. Logist. 12, 33–62. doi:10.4236/jssm.2015.81002

Montgomery, D.C., 2009. Introduction to statistical quality control, John Wiley & Sons Inc. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C

Neely, A., Gregory, M., Platts, K., 1995. Performance measurement system design: A literature review and research agenda. Int. J. Oper. Manag. Prod. Manag. 15, 80–116.

Parsa, P., Rossetti, M.D., Zhang, S., Pohl, E.A., 2017. Quantifying the benefits of continuous replenishment program for partner evaluation. Int. J. Prod. Econ. 187, 229–245. doi:10.1016/j.ijpe.2017.02.017

Wardell, D.G.., Moskowitz, H., Plante, R.D.., 1992. Control Charts in the Presence of Data Correlation. Manage. Sci. 38, 1084–1105.

Zipkin, P., 2000. Foundations of inventory management, Foundations of inventory management.

4 CHAPTER 3

4.1 Introduction

Multi-Stop Truckload (MSTL) is becoming increasingly popular among shippers (i.e. companies that need their products transported) for shipping less than truckload (LTL). According to a massive dataset from CH Robinson, the business share of MSTL has increased from 6.42% in 2013 to 7.39% in 2015. This is due to its cost savings potential, shorter transit time, reduced damages, more certain delivery time, and positive environmental impacts. At the same time, carriers have become more cautious about accepting multi-stop load tenders because they tend to impose extra travel distance, higher cost of operation, longer detention time for drivers, and cause disruptive effects on the flow balance of the carriers' transportation routes.

The trucking industry primarily consists of two modes of transport: Full Truckload (FTL) and Less-than-Truckload (LTL). MSTL, which mostly targets the LTL market, consists of using one full truckload to deliver to multiple locations on a single trip. While it is mostly recognized as a new transportation option, many still consider it as a variant of FTL. The main reason is that the pricing structure of MSTL is very similar to the pricing structure of FTL within the United States. In the truckload market, there are two pricing alternatives, spot market and annual pricing contracts. Companies that regularly need a large volume of products to be shipped often set annual pricing contracts with carriers to secure lower prices and avoid the volatility and uncertainty of the spot market. Although contracts added more certainty to the freight transportation, uncertainty is not completely removed because carriers are not obligated to accept all the loads offered by the shipper. The primary advantage of contracts over the spot market is the fixed cost per mile for different lanes (lanes are pairs of origin cities and destination cities) while in the spot market freight is put up for bid. The MSTL market uses the same annual contract pricing format with an addition of stop-off charges on a per stop basis and shippers have tried to take advantage of this pricing structure and reduce their transportation cost. A very recent study by Chen and Tsai Yang (2016) shows that carriers have become more selective in accepting multi-stop load tenders and tried to implicitly increase the cost of MSTL shipments. The study applies predictive analytics on a large representative dataset of loads and identifies significant factors that affect both the behavior of carriers in accepting MSTL and the actual price of the load. Factors such as distance, number of stops, proximity of stops, additional distance to travel, lead time, stop-off charge, and origin-destination states, impact both the cost of MSTL and the acceptance rate from carriers. Chen and Tsai Yang (2016) develop regression models to predict the cost and the acceptance chance of multi-stop load tenders. Companies can use this study to plan their transportation network and wisely offer multi-stop routes that maximize their savings and acceptance rate.

This chapter proposes a multi objective decision model to identify the best two-stop routes that maximize the cost savings for the shipper and fulfill the most important load acceptance criteria of the carriers, which are *out-of-route miles* and *proximity of stops*. The model provides a trade-off capability for selecting routes with more appeal to either shipper or carrier. The application of the model is discussed for a healthcare supply network. We use weekly forecast data at the SKU level along with shipping and distance information of the distribution network to compute the potential savings of every possible two-stop route via an exhaustive search. The routes with positive savings will be subject to a multi objective decision model that selects the best routes given the load acceptance criteria of carriers.

4.2 Literature Review

Multi stop trucking is also known as milk run because the practice originates from the dairy industry. It involves a delivery vehicle that visits multiple locations on a single trip to either drop off or pick up orders. According to Chen and Tsai Yang (2016), who used a massive dataset obtained from CH Robinson (contains about 4 million shipment records), 55% of MSTLs were picked up from one location and dropped off at two locations. The next common type of MSTL, which takes 26% of the loads, is picking up from one location and dropping off at three locations. Besides milk run, direct shipping, cross-docking, and tailored networks are other types delivery methods (Du et al., 2007). The automobile industry is one of the pioneers in implementing milk run logistics, primarily in the upstream of the supply chain where manufacturers pick up parts from suppliers to support production processes. Numerous companies including Toyota in both Japan and the United States, Shanghai GM in China, Volkswagen and Jaguar Land Rover in Europe, Turk Tractor in Turkey, and an automotive manufacturer in Indonesia have implemented and benefited from milk run logistics over the years (Brar and Saini, 2011). Although the most accessible benefit is in transportation cost through reduction in traveled miles, there are other substantial benefits in areas such as inventory holding cost, CO_2 emissions, and truck utilization (load factor). Ricoh Express doubled loading efficiency from 30% to 65% by implementing optimized milk run transportation. They significantly shortened travel distances which resulted in 35% reduction in CO_2 emissions (Brar and Saini, 2011). In the retail sector, a study of 750 stores in Japan showed that a milk run delivery system that consolidates vendor shipments to retail stores in Tokyo can reduce the number of truck deliveries by 5.5% (Akiyama and Yano, 2010). Walmart stores is greatly benefiting from multi stop trucking in the U.S. by leveraging their own private fleet. Milk runs can also reduce inventory significantly because deliveries can become more frequent and in smaller sizes. For example, Shanghai GM reported 30% reduction in inventory in addition to 20% reduction in transportation trips and 30% integrated logistics cost reduction (Xu, 2003).

The milk run problem can be categorized into three types: supplier milk run, customer milk run, and mixed milk run. They are respectively focused on inbound transportation planning, outbound transportation planning, and hybrid delivery, which considers bypassing the distribution center (DC) as plausible (i.e. direct shipments from suppliers to customers) (Sadjadi et al., 2009). The focus of the present paper is on customer milk run category, which concerns order drop-off to customers. The milk run problem can be defined as a special kind of vehicle routing problem (VRP) with time windows and a limited number of vehicles (Eroglu et al., 2014). Many operations research studies have developed mathematical models and heuristics to optimize multi stop truckload planning under various assumptions. Yildiz et al. (2010) developed a mixed integer program for Robert Bosch LLC, a leading automotive parts manufacturer, to combine shipments on the same route but opposite flows into returning empty containers from a milk run trip. They also investigated the impact of crossdocking in the study. You and Jiao (2014) developed a mathematical model and used the Clark-Wright savings algorithm to solve large scale milk run problems for a courier company. They assume single type of delivery vehicle and fixed cost structures for transportation. Hosseini et al. (2014) developed an integer programming model and a hybrid heuristic solution approach to minimize shipping cost by reducing the number of required identical vehicles. Sadjadi et al. (2009) included due dates and inventory costs in their analysis. They considered a waiting cost for delivery trucks at each supplier as well as an assumption for shipping loads being increments of pallet loads for each item. A mathematical model was developed with a genetic algorithm solution approach along with a computational study to solve a small-scale problem that includes a maximum of ten suppliers and fifty parts.

Multi stop truckload (MSTL) in the U.S. is priced using the annual contract pricing structure of the truckload (FTL) market. Such contracts set per mile rates ($/mile) for origin and destination city pairs (i.e. lanes) between shippers and carriers. The only difference for MSTL lies in the "stop-off charges" (SOC), which are per stop charges ($/stop) for additional intermediate stops (Chen and Tsai Yang, 2016). Although annual contracts add certainty for shippers, carriers are not obligated to accept all the loads offered by shippers. Once a load is rejected, shippers need to go to the next preferred carriers in their routing guide. Chen and Tsai Yang (2016) showed that multi stop loads have higher rejection rates from carriers than direct loads. They used a supervised classification algorithm known as logistic regression to model carrier behavior towards multi stop loads. They identified significant factors such as out-of-route miles, proximity of stops, number of stops, SOC, and continuous move potential to predict the chance of acceptance. According to Caldwell and Fisher (2008), load tenders rejection can increase the cost of transportation because shippers normally place the cheaper carriers at higher positions in their routing guide. They observed a 7.9% cost increase in initial rejection followed by 3.2% in subsequent rejections. Load rejections not only increase the transportation cost but also disrupt the delivery plans and negatively impact other dependent operations. Therefore, the main challenge for shippers is not only to offer the most profitable multi-stop loads but have them be accepted by the carriers. In this chapter, we develop a model that helps shippers to select the best two-stop routes with respect to both aspects of this decision problem.

4.3 MSTL Cost Savings

This study focuses on the manufacturers (i.e. shippers) in the MSTL market and develops a model that enable them to integrate cost savings maximization with the most important carrier acceptance criteria. One of the typical distribution systems in any sector is a network of distributors that receive products from manufacturers and send them downstream to retailers or wholesalers or hospitals (Figure). Manufacturers sell their products to numerous distributors and ship them via different carriers with whom the manufacturer maintains annual contracts. Both manufacturers and distributors own DCs, typically spread around their target geographical areas. In this section, we discuss the process of cost savings calculation to identify the most profitable pairs of distributor locations for two-stop deliveries over time.

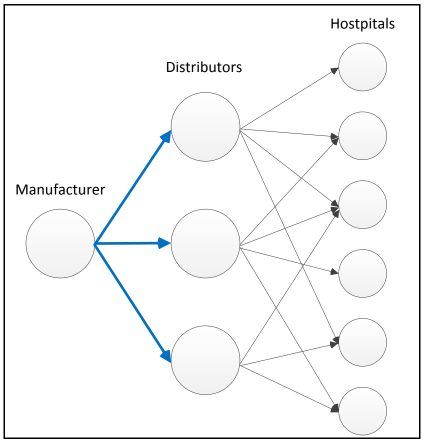

Figure 1: A representation of healthcare supply chain

The following list represents the critical assumptions of the study along with the logic behind using them:

1. The replenishment system resembles a VMI or CRP system in which the manufacturer makes the replenishment quantity and timing decisions,
2. The demand pattern across the network at the aggregate level is stationary over a yearly time period. This allows the cost savings estimates of one year to be valid for the next

year planning. This assumption can be valid depending on the industry or product family or etc.

3. Two-stop delivery system: vehicles are only allowed to stop for two drop-offs. Therefore, once a vehicle is dispatched, it will only visit delivery locations,
4. The demand at two distributor locations can only be combined and delivered via a two-stop trip if their order dates are not apart by more than a week,
5. MSTL rates are the same as the FTL rates ($/mile) associated with the final stop plus an additional stop-off charge that normally varies between $50 and $100 per extra stop.

The model is developed for a network with one manufacturer DC because the two-stop delivery system only allows multiple drop-offs. Therefore, to analyze a network with multiple manufacturer DC's, the model needs to be executed multiple times. Since this is a transportation problem, load size is the main characteristic that needs to be considered. Therefore, for each distributor location, the demand is aggregated across SKUs and bucketed into weekly periods (assumption 4). Then, given the total weekly demand at distributor locations, all possible combinations of two-stop routes per week are evaluated. For each pair of distributor locations, the delivery sequence with the shorter travel distance is considered (i.e. depot → A → B or depot → B → A). This process is an exhaustive search to compute the cost savings for every two-stop delivery over the planning period. For populated networks, where the number of delivery locations does not allow an exhaustive search, heuristic algorithms such as the Clark-Wright algorithm can be used (You and Jiao, 2014). The savings potential of every two-stop route is the difference between the shipping costs of i) two separate direct deliveries and ii) a two-stop delivery. Let

D_x the total weekly demand at location x (ft^3)

D_x^{MSTL} the weekly demand at location x that can be sent via a two-stop delivery (ft^3)

W_x^{MSTL} the weekly demand at location x that can be sent via a two-stop delivery ($lbs.$)

LL the lower cube limit for dispatching a FTL or MSTL truck (ft^3)

UL the upper cube limit for dispatching a FTL or MSTL truck (ft^3)

r_x the FTL shipping rate from the manufacturer DC to location x ($\$/mile$)

ST the stop-off charge ($/intermediate stop)

$l_{d,w}$ the LTL rate for a load with delivery distance of d and weight of w (cwt)

$d_{x,y}$ the driving distance from location x to location y ($mile$)

d_x the driving distance from depot to location x ($mile$)

$C_{x,y}^{direct}$ the total shipping cost to location x and y via a single stop/direct system

$C_{x,y}^{MSTL}$ the total shipping cost to location x and y via a two-stop system

$S_{x,y}$ the total savings associated with two-stop delivery to location x and y

The highlighted parameters above are known and do not need to be estimated. D_x is known from the weekly bucketed demand data. LL and UL are shipping policy parameters of the manufacturer DC. r_x is the FTL rate of each lane that is negotiated between the manufacturer and contract carriers. $d_{x,y}$ (or d_x) is essentially the distance matrix of the distribution network. The remaining variables need to be estimated for each route. To compute the savings, LTL rates are critical because the main target of MSTL is the LTL market. LTL cost structure is more complicated than FTL and is a function of both distance and weight (Mendoza and Ventura, 2009). Therefore, we estimated the LTL rates by developing a regression model on a dataset of 3,569 LTL shipments, obtained from a healthcare manufacturer, that contains total cost figures, lane distance, and weight. Equation (1) is the generic regression equation that we obtained.

$$l_{d,w} = w^{-\alpha} e^{\beta d} \quad (1)$$

where w and d represent the weight of a load and distance of the corresponding lane, α and β are constant values that cannot be disclosed for confidentiality reasons. Figure 2 illustrates the LTL dataset in blue dots as well as the fitted values of the regression model in red dots. The regression model in Figure 2 shows a good fit with R^2 of 0.71 and a good accuracy with relative percent error of -5% which means over-estimating CWT by 5%. The residuals of the model (Figure 3) also do not show any systematic pattern indicating that the model has captured the patterns in the data quite well.

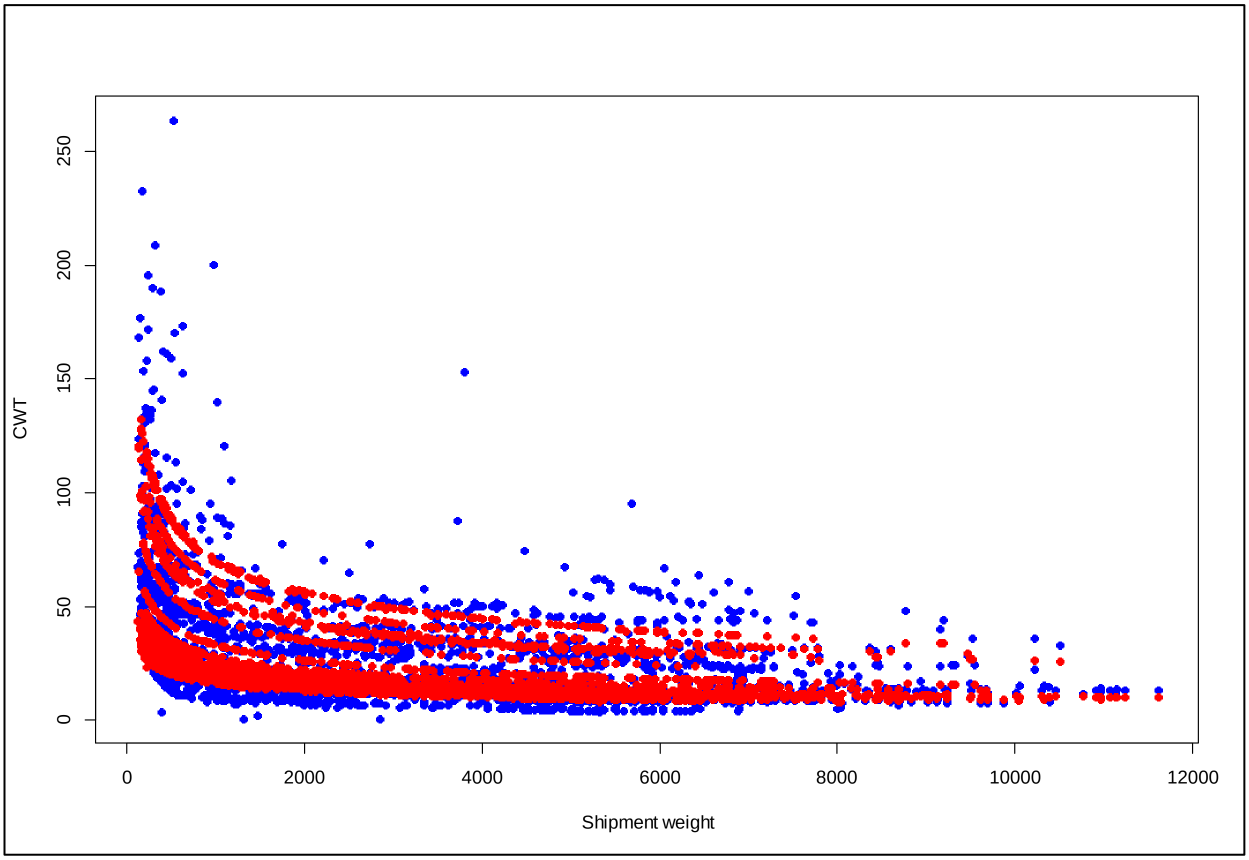

Figure 2: LTL shipment dataset (blue dots) and the fitted values from the model (red dots)

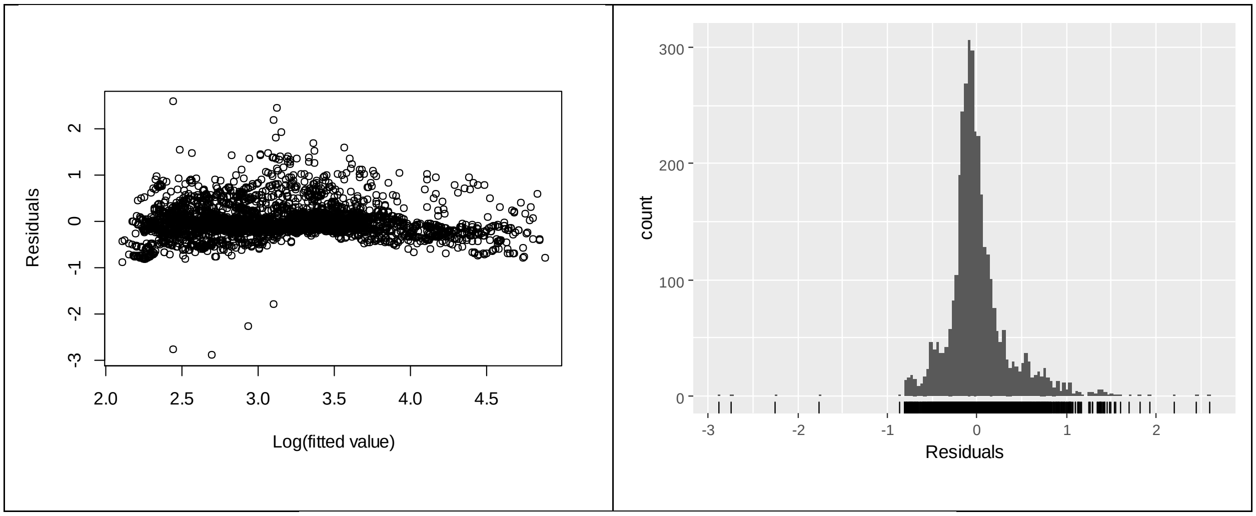

Figure 3: Residual analysis of the regression model

To compute the savings associated with a two-stop delivery, we need to estimate the available demand for the delivery, D_x^{MSTL}. The total weekly demand at a location (i.e. D_x) could be so high that direct, fully utilized FTL shipments can be used without needing two-stop delivery planning. In such case, D_x^{MSTL} becomes the remaining volume that cannot be shipped via direct FTL. By comparing the total weekly demand (i.e. D_x) with the upper and lower cube limits of the manufacturer DC (i.e. LL and UL), D_x^{MSTL} can be determined:

$$D_x^{MSTL} = \begin{cases} D_x & if\ D_x < LL \\ D_x & if\ LL \leq D_x \leq UL \\ D_x \bmod UL & if\ D_x > UL \end{cases} \tag{2}$$

Given D_x^{MSTL}, the difference between the cost of direct shipments ($C_{x,y}^{direct}$) and the shipping cost of two-stop delivery ($C_{x,y}^{MSTL}$) is the expected savings (Equation 3). To compute $C_{x,y}^{direct}$, we need to define it for all possible scenarios. Since either of direct FTL or LTL can be used for delivering to location *x* and *y*, there are four different ways of shipping via two direct single-stop deliveries (i.e. FTL+FTL, FTL+LTL, LTL+FTL, LTL+LTL). Equation 4 lists these four scenarios, the conditions associated with each, and their corresponding cost equations.

$$S_{x,y} = C_{x,y}^{direct} - C_{x,y}^{MSTL} \tag{3}$$

$$C_{x,y}^{direct} = \begin{cases} r_x\, d_x + r_y d_y & if\ D_x^{MSTL} > LL\ and\ D_y^{MSTL} > LL \\ r_x\, d_x + \left(W_y^{MSTL} \times \frac{l_{d,w}}{100}\right) & if\ D_x^{MSTL} > LL\ and\ D_y^{MSTL} < LL \\ \left(W_x^{MSTL} \times \frac{l_{d,w}}{100}\right) + r_y\, d_y & if\ D_x^{MSTL} < LL\ and\ D_y^{MSTL} > LL \\ \left(W_x^{MSTL} \times \frac{l_{d,w}}{100}\right) + \left(W_y^{MSTL} \times \frac{l_{d,w}}{100}\right) & if\ D_x^{MSTL} < LL\ and\ D_y^{MSTL} < LL \end{cases} \quad (4)$$

The only rational feasibility condition for a two-stop delivery to location *x* and *y*, is $LL < D_x^{MSTL} + D_y^{MSTL} < UL$, which ensures the capacity constraint of the shipping vehicles. For each pair of x and y that satisfies this condition, $C_{x,y}^{MSTL}$ can be computed as follows:

$$C_{x,y}^{MSTL} = r_y\left(d_x + d_{x-y}\right) + ST \quad (5)$$

The value of $S_{x,y}$ can be computed for all possible *xy* pairs over a sufficiently long period of time (e.g. a year to capture possible seasonality effects) to evaluate the savings potential of each pair. Savings potential has two key indicators: I) average expected weekly savings ($\bar{S}_{x,y}$ in Equation 6) II) expected savings frequency ($\hat{F}_{x,y}$ in Equation 7). While the first one is an indicator of savings magnitude, the second one is an indicator of savings consistency over time. They are both important indicators for shippers to evaluate the profitability of routes.

$$\bar{S}_{x,y} = \frac{\sum_{t=1}^{T} S_t^{x,y}}{T} \quad (6)$$

$$\hat{F}_{x,y} = \sum_{t=1}^{T} I_t \quad where \quad I_t = \begin{cases} 1 & if\ \ S_{x,y}^t > 0 \\ 0 & if\ \ S_{x,y}^t \le 0 \end{cases} \quad (7)$$

where *t* is the week index and $T = 52$ if one year of data is used to estimate $\bar{S}_{x,y}$ and $\hat{F}_{x,y}$.

If savings potential was the only criteria to succeed in offering MSTL tenders, the above two indicators would be the only representatives of the decision factors. The problem becomes more interesting knowing that savings only matters to the shippers. Carriers have their own criteria for accepting loads, which oftentimes have nothing to do with the savings potential for the shippers! As discussed in the literature review section, Chen and Tsai Yang (2016) identified these criteria through an empirical study. The two crucial decision factors for carriers are *out-of-route miles* and *proximity of stops*. Out-of-route miles is the extra distance that needs to be traveled to visit an intermediate stop before arriving at the final destination.

4.4 Multi-objective Decision Analysis (MODA)

We propose a multi-objective decision model to assist shippers in selecting the best two-stop routes considering i) savings magnitude ($\bar{S}_{x,y}$) ii) savings consistency over time ($\hat{F}_{x,y}$) iii) out-of-route miles ($R_{x,y}$), and iv) proximity of stops ($d_{x,y}$). While the first two seek the shipper's benefit, the last two affect carriers' acceptance chance. In this section, we first discuss the components of MODA in general and how they support the process of selecting routes. Then we focus on applying MODA on the problem to integrate the four above factors to select the best routes.

MODA is an evaluation methodology that uses the decision criteria to measure how well different candidate solutions (e.g. two-stop routes in our problem) satisfy the fundamental objective of the stakeholders in the decision-making problem. Prior to measuring the value of any candidate solution, a qualitative model must be constructed that captures the critical objectives of the stakeholders. Then MODA will quantify the value of each candidate solution with respect to the objectives that are identified in the qualitative model. The process can be listed as follows:

1. *Collect the views of stakeholders*

   The stakeholders of a decision problem are the parties that are influenced by the decision (e.g. shipper and carriers in our problem).

2. *Determine the fundamental objective*

The most basic and high-level objective the stakeholders are trying to achieve.

3. *Determine the objectives and sub-objectives*

   Any fundamental objective contains various aspects so it can be broken out to specific objectives that each points out to one aspect of the fundamental objective.

4. *Construct the value tree*

   A pictorial representation of hierarchy of identified objectives. The fundamental objective is placed on the top while the relations between objectives and value measures shape the tree.

5. *Determine the value measures*

   Each objective at the lowest level of the value tree needs a value measure. A measure assesses how well each candidate solution attains the corresponding objective.

6. *Determine the value functions*

   Each value measure has a different scale with different units. Value functions are used to convert candidate solution scores on the value measures to a standard unit.

7. *Weights*

   A value, normally between 0 and 1, that represents the importance of a value measure. The weights should sum to 1.

8. *Quantitative value model*

   This is a quantitative method for trading off conflicting objectives (Kirkwood, 1996). Although different mathematical relationships can achieve this, we use the most common method called the additive value model to calculate the value of each candidate solution.

A complete review of MODA can be found at Parnell et al. (2011), where various aspects of the methodology are discussed along with examples from multiple application areas.

4.4.1 Model Development

The views of shippers and carriers, as the primary stakeholders of MSTL business, lead to the fundamental objective. Our collaboration with a major manufacturer in the healthcare industry indicates that cost savings is the main objective of shippers because on the customer service side, they are almost certain that transitioning from LTL to MSTL would increase the quality of delivery for customers. As discussed earlier, $\bar{S}_{x,y}$ and $\hat{F}_{x,y}$ are appropriate value measures for the shipper's objective. On the other hand, studies such as Chen and Tsai Yang (2016) show that carriers' main objectives are to avoid extra travel miles and extra detention or dwell time (which refers to the time periods that drivers become idle due for over-night stays or unloading delays). Out-of-route miles and proximity of stops are the value measures that capture carriers' objectives. Given such information, we can develop the value tree, where the fundamental objective is at the top, objectives at the second tier, and value measures at the third tier (Figure 4).

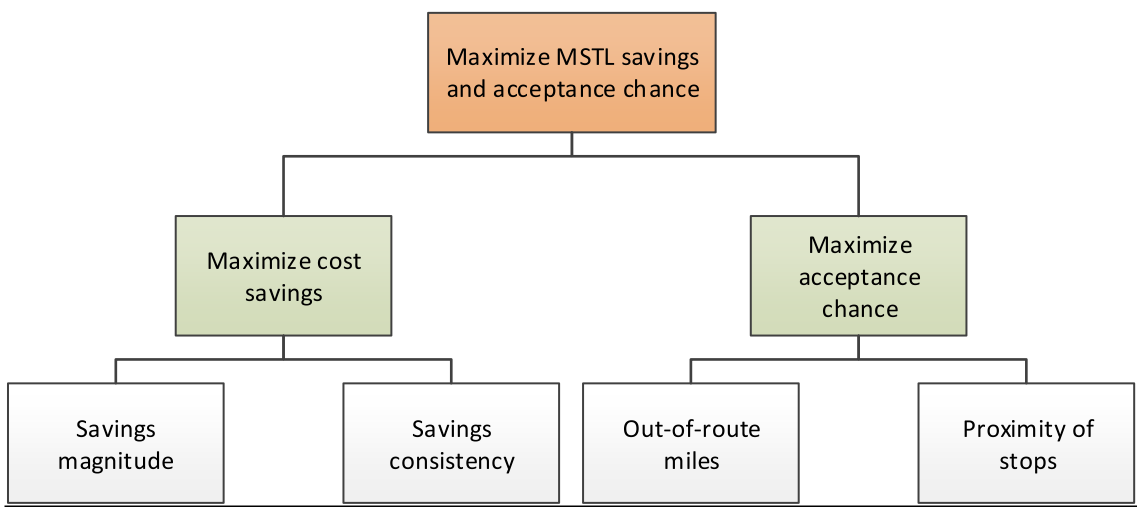

Figure 4: Value tree

All four value measures in Figure 4 can be obtained for any two-stop route. While savings magnitude ($\bar{S}_{x,y}$) and consistency ($\hat{F}_{x,y}$) are discussed in section 4.3, out-of-route miles ($R_{x,y}$) and proximity of stops ($d_{x,y}$) can be obtained from a map service such as Google maps.

As indicated in the development process of MODA, at this stage we need to develop value functions for the value measures. Value functions convert different scales of value measures to a common scale that ranges from 0 to 100. This determines a *value* for each route by adding up its scores on the value measures. Each value function has an *x*-axis and a *y*-axis, where the *x*-axis is the scale of the value measure (e.g., dollar savings) and *y*-axis is a standard unit-less scale from 0 to 100. Continuous value functions typically follow four basic shapes of linear, concave, convex, and S curve (Figure 5). Depending on the impact of each value measure, value functions could be either monotonically increasing, as indicated in Figure 5, or decreasing. As suggested in Kirkwood (1996), the shape of value functions is determined by consulting with subject experts. Once the general shape is determined, the experts should identify the increase/decrease in value from a specific incremental increase in the measure scale. Repeating this multiple times up to the maximum on the measure scale will produce a piecewise linear function. The instance functions in Figure 5 are produced in a linear piecewise fashion. Let

k_i be the score of measure i on the *x*-axis of the value function

$v(k_i)$ be the value of measure i on the *y*-axis of the value function

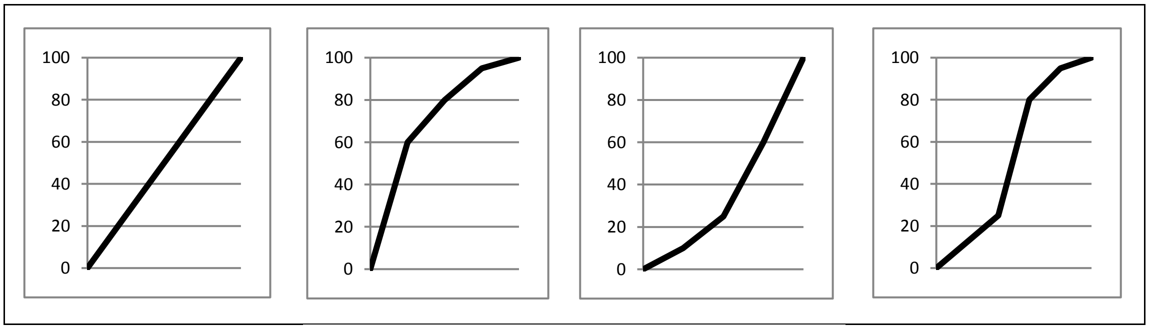

Figure 5: Value function types for MODA

A two-stop route would have k_1, …, k_4 scores that correspond to the four value measures shown in Figure 4. Value functions would then convert k_1, …, k_4 scores to $v(k_1)$, …, $v(k_4)$ values that all have common scales (i.e. [0, 100]).

Typically, decision makers do not view all value measures equally. Measure weights (w_i) are supposed to capture the importance of measures to the decision makers and incorporate them to the value model (Equation 8). The weights depend on both the importance of the value measure and the impact of varying the score of value measures. Swing weight matrix is one of the well-known methods to determine the weights. This method assesses measure weights by "swinging" the value measure score from its worst to its best. Parnell and Trainor (2009) discusses this method in detail with examples. There are various ways besides the swing weight matrix to elicit weights from stakeholders which are discussed in (Clemen and Reilly, 2001; Kirkwood, 1996). Once the weights are determined, we can evaluate two-stop routes using a value model that generates total *value* for each route (Equation 8). A value model is a mathematical expression that provides trading off capability among objectives. MODA has many different relationships to do this but we will use the most common method called the *additive value model* to calculate how well each route satisfies identified objectives:

$$V(x,y) = \sum_{i=1}^{4} w_i v(k_i) = w_1 v(\bar{S}_{x,y}) + w_2 v(\hat{F}_{x,y}) + w_3 v(R_{x,y}) + w_4 v(d_{x,y}) \qquad (8)$$

A higher total value indicates a better two-stop route given the weights, value functions and its measure scores. MODA enables us to evaluate the routes based on multiple objectives that influence the success of MSTL. We can gain valuable insights by performing a sensitivity analysis on the elements of MODA and observe how the results change. Weights and value functions are the elements that can be subject to sensitivity analysis to gain further insights on the most robust high-performing routes.

4.5 Case Study

This study is focused on the distribution network of a major healthcare manufacturer that has four DC’s across the U.S. They ship their products to numerous independent distributors via different carriers with whom the manufacturer maintains annual contracts. Distributors operate their own distribution networks to replenish hospitals as the end customers in the supply chain. The relationship between manufacturer and its distributors constitutes a distribution network with four manufacturer DC’s that support the demand at 96 distributor DC locations (Figure 6). As discussed, the purpose is to identify the best pairs of distributor DC locations for two-stop deliveries over time. The distributors are in a continuous replenishment program which allows the manufacturer to make the decision about replenishment quantities and transportation timing.

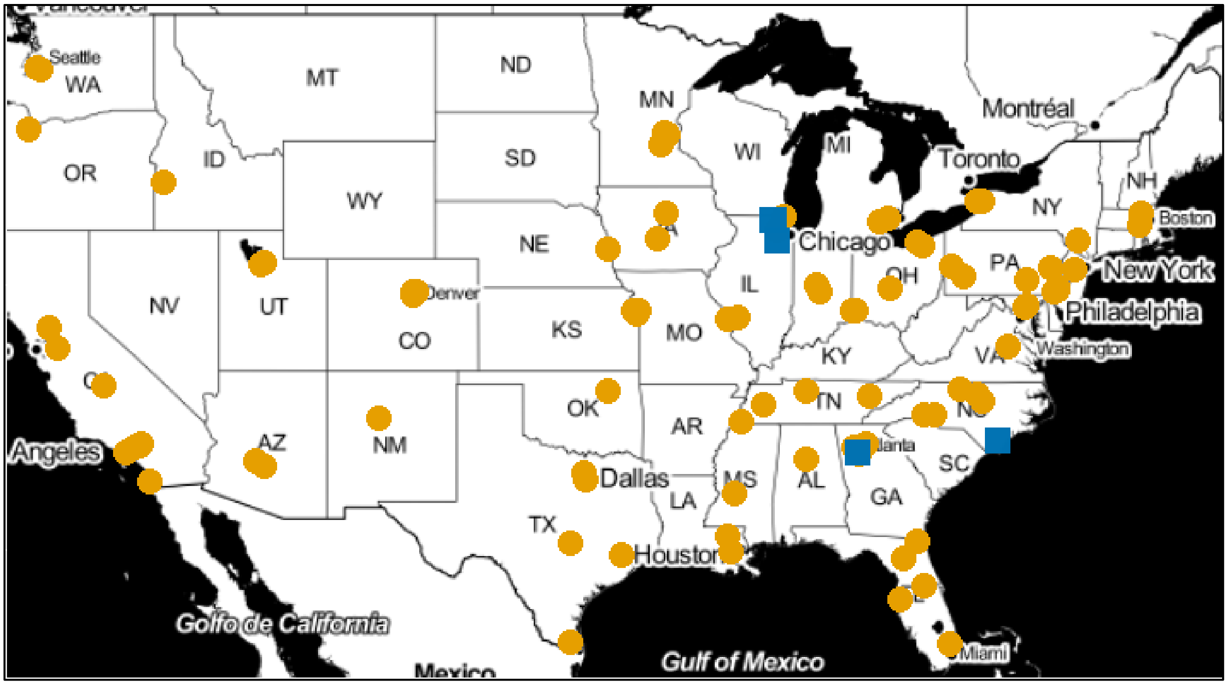

Figure 6: Four blue squares and 96 yellow circles indicate the manufacturer DC's and distributors DC's respectively

4.5.1 Results and Discussion

We begin with cost savings estimations for all possible pairs out of the manufacturer DC is Atlanta, GA. Table lists the given constants and variables as well as the variables that need to be estimated for the routes on a weekly basis to finally estimate the savings magnitude ($\bar{S}_{x,y}$) and consistency ($\hat{F}_{x,y}$) of routes. The given constants are the stop-off cost (*ST*), lower and upper cube limits of the manufacturer DC (*LL* and *UL*). The given variables are the distance matrix (driving distances obtained from the Google maps) and FTL rates obtained from carriers. The formulas discussed in section 4.3 are used to estimate the rest of the variables. We used the demand of the 96 distributor locations over time for one calendar year and computed the savings for every week.

Table 1: constants and variables for cost savings estimation

Given constants	Given variables	Estimated variables for each route on each week	Estimated variables for each route
$ST = \$100$	$d_{x,y}$	D_x^{MSTL}	$\bar{S}_{x,y}$
$LL = 1200\ ft^3$	r_x	W_x^{MSTL}	$\hat{F}_{x,y}$
$UL = 2000\ ft^3$		$l_{d,w}$	
		$C_{x,y}^{direct}$	
		$C_{x,y}^{MSTL}$	
		$S_{x,y}$	

The cost savings analysis leads to identifying 80 two-stop routes that generate total annual savings of \$590K over 52 weeks. Figure 7 illustrates the distribution of annual savings by route. One noticeable pattern, which is expected given the fact that shipments are dispatched from Atlanta, is the identification of high savings routes with delivery locations in the west coast.

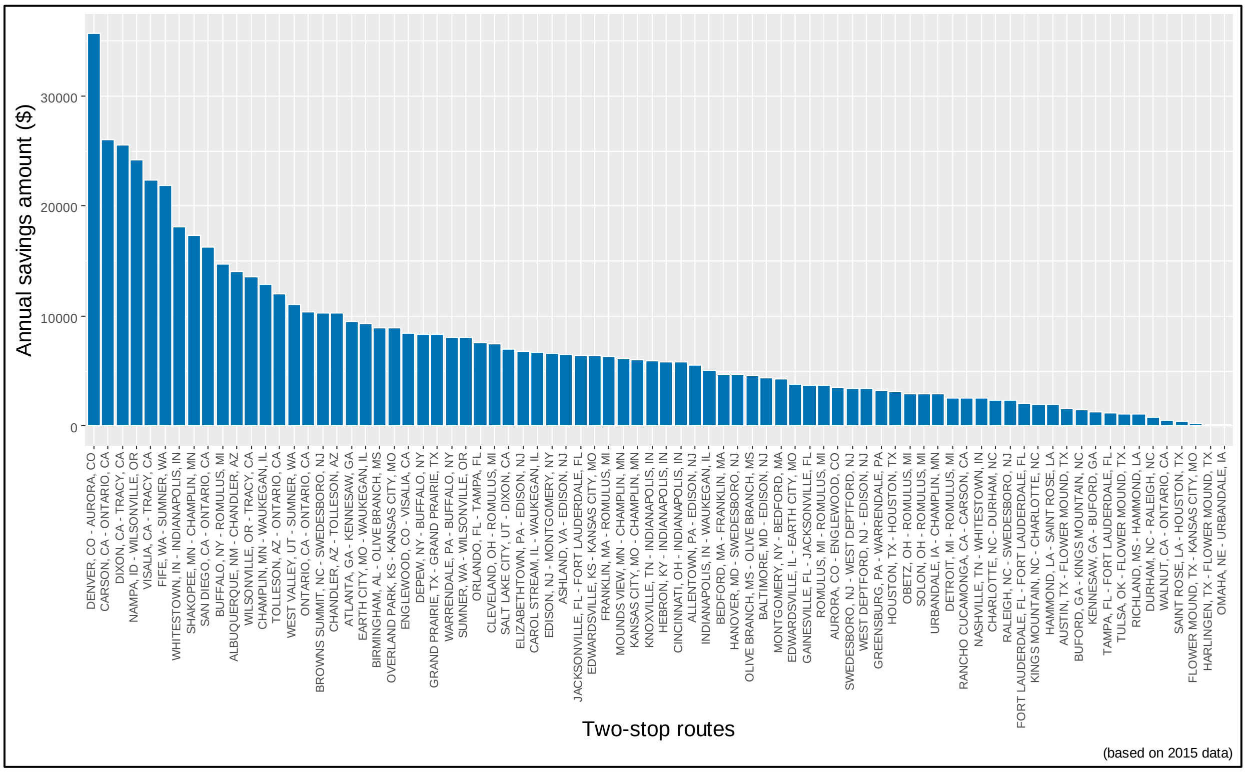

Figure 7: Distribution of cost savings from the manufacturer DC in Atlanta, GA

The realization of savings over time is illustrated in Figure 8. The graph shows a fairly consistent average of \$10K to \$12K per week across the network, except from week 23 to week 27 where a major decline is realized. The savings trend shows a significant correlation with the demand pattern over time (correlations coefficient = 0.69). Thus, demand can be used as a strong predictor of savings if needed.

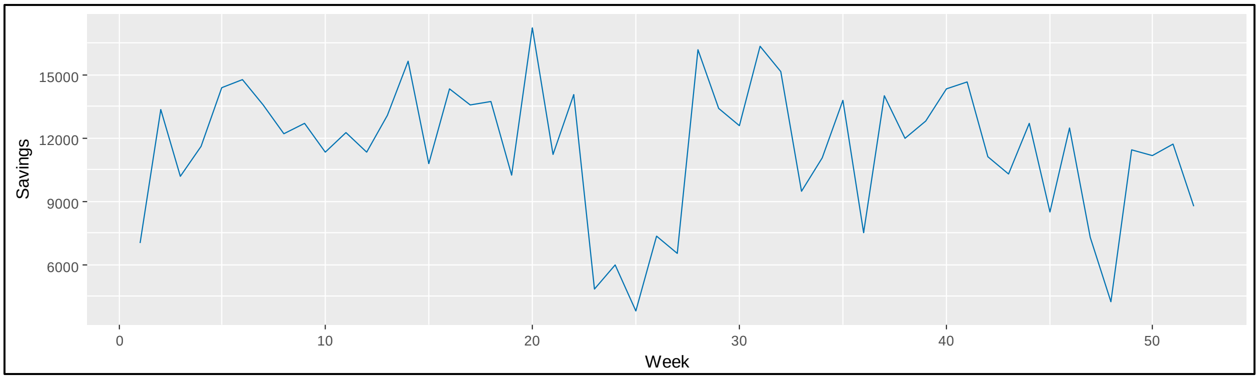

Figure 8: Time series realization of cost savings from the entire network

It is worth indicating that there are location overlaps between the identified set of 80 routes because a location can generate savings by getting paired with different locations over time. That is why it is important to consider consistency of savings generation ($\hat{F}_{x,y}$) over time as the other important savings indicator. Figure 9 shows the distribution of savings consistency ($\hat{F}_{x,y}$) for the same 80 two-stop routes. The x-axis of the figure does not indicate the same high performing routes that are observed in Figure 7. In fact, the correlation between the savings magnitude indicator ($\bar{S}_{x,y}$) and the savings consistency indicator ($\hat{F}_{x,y}$) is - 0.11. Figure 10 depicts this relationship for the 80 channels that are identified. The insignificant correlation reassures the necessity of using both indicators as the cost savings value measures in MODA.

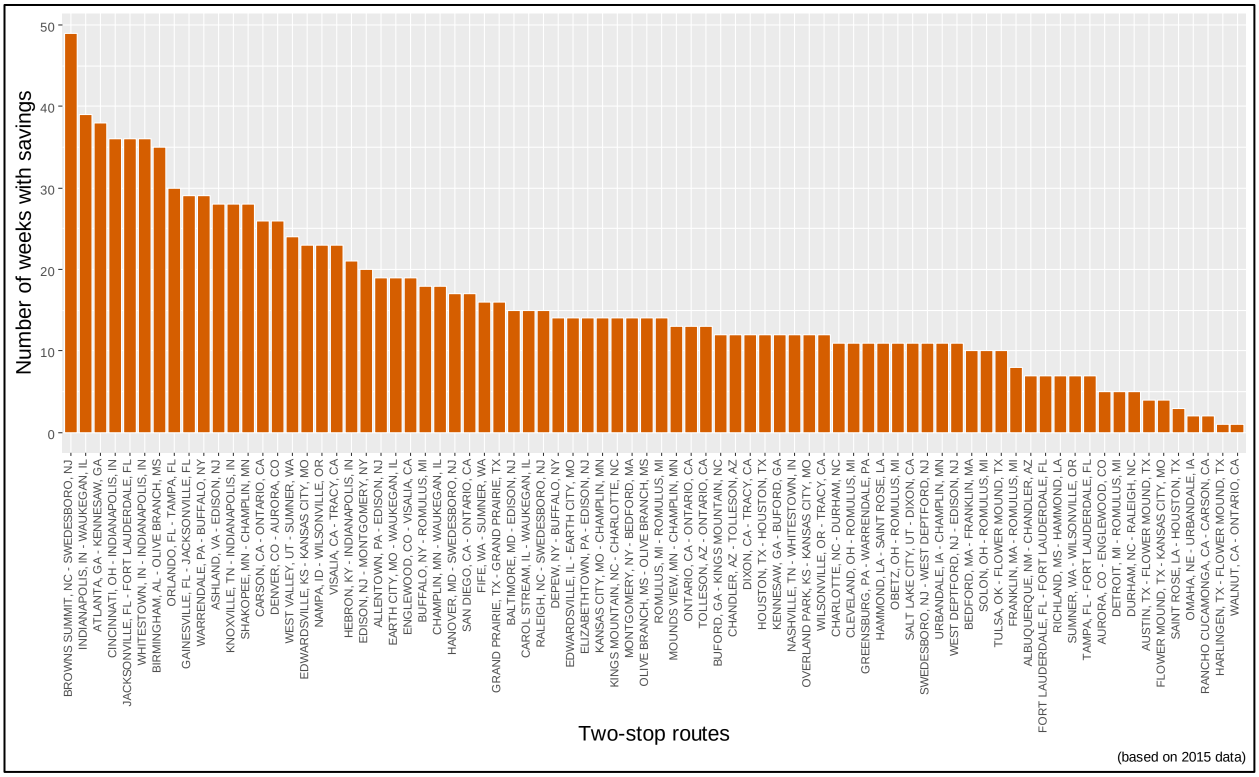

Figure 9: Distribution of number of weeks with savings for two-stop routes out of the DC in Atlanta, GA

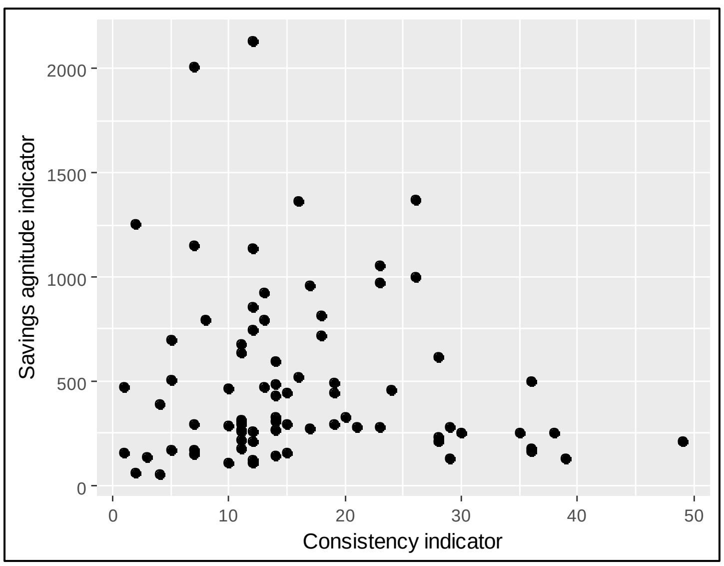

Figure 10: The relationship between $\bar{S}_{x,y}$ and $\hat{F}_{x,y}$

Before discussing the developed MODA for the case study, it is insightful to look at the summary statistics of the two other decision factors, out-of-route miles ($R_{x,y}$) and proximity of stops ($d_{x,y}$). Table 2 and Figure 11 indicate that even though the stops might be far away from each other, the constructed routes do not create much out-of-route miles. This can still be problematic for carriers because far away stops increase the chance of over-night stays for drivers which impose detention charges and make the drivers unavailable for longer times.

Table 2: Case study results: summary statistics of out-of-route miles and proximity of stops

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
$R_{x,y}$	-19.840	7.252	31.710	102.300	69.860	1074.000
$d_{x,y}$	6.59	32.78	124.30	193.60	282.30	1115.00

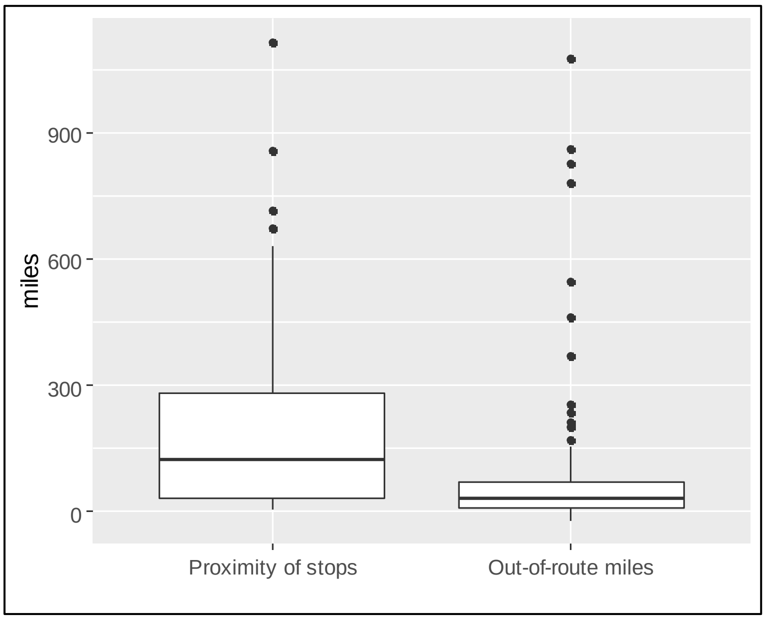

Figure 11: Case study results: boxplot of out-of-route miles ($R_{x,y}$) and proximity of stops ($d_{x,y}$)

At this point, we have all four criteria of the MODA for all the routes. Savings magnitude ($\bar{S}_{x,y}$) and consistency ($\hat{F}_{x,y}$) are estimated, while out-of-route miles ($R_{x,y}$) and proximity of stops ($d_{x,y}$) are obtained from the Google maps. These are the value measures for the MODA as indicated earlier (Figure 4). They need to be converted in standard units using value functions. Table 3 lists the range of observations for each criterion in the case study, as well as the value functions and weights that are assumed for the model. The value functions are assumed to be linear for this case study based on the opinions of subject experts. The weights indicate more importance toward the carrier acceptance in order to convince them by offering more appealing routes. Given the weights and value function forms, the total value for each route *x*,*y* can be computed as shown in Equation 7. Table 4 shows the value measures and $V(x,y)$ figures for all the 80 routes.

Table 3: MODA elements for the case study

Criteria	Range	Value function	Weight ($\boldsymbol{w_i}$)
$\bar{S}_{x,y}$	\$52 – \$2,132	Linear increasing	0.15
$\hat{F}_{x,y}$	1 – 49 weeks	Linear increasing	0.15
$R_{x,y}$	0 – 1,073 mi	Linear decreasing	0.4
$d_{x,y}$	7 – 1,115 mi	Linear decreasing	0.3

$$V(x,y) = 0.15\frac{100(\bar{S}_{x,y})}{\max_{x,y}\bar{S}_{x,y} - \min_{x,y}\bar{S}_{x,y}} + 0.15\frac{100(\hat{F}_{x,y})}{\max_{x,y}\hat{F}_{x,y} - \min_{x,y}\hat{F}_{x,y}} + 0.4\times 100\left(1 - \frac{R_{x,y}}{\max_{x,y}R_{x,y} - \min_{x,y}R_{x,y}}\right) + 0.3\times 100\left(1 - \frac{d_{x,y}}{\max_{x,y}d_{x,y} - \min_{x,y}d_{x,y}}\right) \tag{9}$$

We indicated earlier that there is location overlap among the 80 routes in Table 4, which means there are locations that exist in multiple routes. Now that $V(x,y)$ values are available for every route, we should select a non-overlapping list of routes by prioritizing ones with the higher $V(x,y)$ figures. This process will select a group of distinct routes that can be offered to carriers. Table 5 lists 29 routes that are selected from the original set in Table 4, to exclude location overlaps and prioritize higher $V(x,y)$ values.

Table 4: Total value $V(x, y)$ calculation for the 80 two-stop routes

STOP 1	Cust	STOP 2	Cust	$\overline{S}_{x,y}$	$\widehat{F}_{x,y}$	$d_{x,y}$	$R_{x,y}$	$V(x,y)$
DENVER, CO	A	AURORA, CO	C	1372.51	26	17.24	26.95	86.57
FIFE, WA	C	SUMNER, WA	A	1365.91	16	6.59	4.99	84.49
ATLANTA, GA	C	KENNESAW, GA	A	249.77	38	27.64	0	82.93
WHITESTOWN, IN	C	INDIANAPOLIS, IN	A	502.52	36	21.81	42.31	82.74
SHAKOPEE, MN	C	CHAMPLIN, MN	B	617.58	28	34.86	-3.05	82.37
DIXON, CA	B	TRACY, CA	A	2132.50	12	75.37	148.95	81.64
CARSON, CA	B	ONTARIO, CA	B	1000.84	26	52.9	101.18	80.21
ONTARIO, CA	B	ONTARIO, CA	A	796.55	13	10	10	79.17
VISALIA, CA	C	TRACY, CA	A	971.64	23	165.51	20.23	78.97
CHANDLER, AZ	B	TOLLESON, AZ	A	854.32	12	36.94	7.51	78.64
GRAND PRAIRIE, TX	B	GRAND PRAIRIE, TX	C	519.12	16	10	10	78.11
OVERLAND PARK, KS	B	KANSAS CITY, MO	A	744.10	12	11.57	21.58	78.01
CINCINNATI, OH	B	INDIANAPOLIS, IN	A	162.09	36	112.22	38.52	77.97
RANCHO CUCAMON, CA	B	CARSON, CA	B	1255.35	2	58.46	5.83	77.88
DEPEW, NY	B	BUFFALO, NY	D	596.09	14	11.9	16.42	77.75
EDWARDSVILLE, KS	C	KANSAS CITY, MO	A	277.99	23	16.19	31.07	77.62
MOUNDS VIEW, MN	A	CHAMPLIN, MN	B	473.93	13	13.54	-0.36	77.13
BIRMINGHAM, AL	B	OLIVE BRANCH, MS	A	255.41	35	216.62	1.68	76.86
INDIANAPOLIS, IN	B	WAUKEGAN, IL	B	129.32	39	233.09	-0.03	76.81
ORLANDO, FL	C	TAMPA, FL	B	251.09	30	84.52	66.66	76.46
BROWNS SUMMIT, NC	C	SWEDESBORO, NJ	B	210.56	49	408.58	-14.44	76.30
AURORA, CO	C	ENGLEWOOD, CO	B	696.91	5	14.83	1.21	76.14
OLIVE BRANCH, MS	B	OLIVE BRANCH, MS	A	326.18	14	10	10	76.09
ALBUQUERQUE, NM	A	CHANDLER, AZ	B	2010.80	7	420.78	-19.84	76.03
GAINESVILLE, FL	A	JACKSONVILLE, FL	B	127.81	29	71.14	60.76	75.84
WARRENDALE, PA	B	BUFFALO, NY	D	278.82	29	198.15	1.81	75.64
ROMULUS, MI	B	ROMULUS, MI	A	263.88	14	10	10	75.64
EDISON, NJ	B	MONTGOMERY, NY	B	329.80	20	93.11	14.31	75.58
CAROL STREAM, IL	A	WAUKEGAN, IL	B	447.56	15	54.68	30.6	75.32
SWEDESBORO, NJ	B	WEST DEPTFORD, NJ	A	311.39	11	11.34	4.47	75.21
ALLENTOWN, PA	A	EDISON, NJ	B	293.68	19	72.41	24.16	75.21
HOUSTON, TX	B	HOUSTON, TX	A	260.97	12	10	10	75.00
EDWARDSVILLE, IL	A	EARTH CITY, MO	B	267.96	14	32.43	15.27	74.87
SAN DIEGO, CA	A	ONTARIO, CA	B	956.89	17	113.82	117.09	74.85
HANOVER, MD	A	SWEDESBORO, NJ	B	272.46	17	96.01	2.01	74.61
KINGS MOUNTAIN, NC	A	CHARLOTTE, NC	B	143.50	14	32.89	2.11	74.44
HEBRON, KY	A	INDIANAPOLIS, IN	A	279.53	21	113.52	34.72	74.24
FT LAUDERDALE, FL	B	FT LAUDERDALE, FL	A	290.24	7	10	10	73.64
WEST DEPTFORD, NJ	A	EDISON, NJ	B	306.63	11	71.7	4.88	73.53
DETROIT, MI	B	ROMULUS, MI	A	505.44	5	23.78	35.43	73.27
GREENSBURG, PA	A	WARRENDALE, PA	B	291.56	11	51.29	25.62	73.21
BEDFORD, MA	B	FRANKLIN, MA	A	463.84	10	45.63	58.52	73.09

Table 4 (Cont.): Total value $V(x,y)$ calculation for the 80 two-stop routes

STOP 1	Cust	STOP 2	Cust	$\bar{S}_{x,y}$	$\hat{F}_{x,y}$	$d_{x,y}$	$R_{x,y}$	$V(x,y)$
NAMPA, ID	C	WILSONVILLE, OR	A	1052.37	23	432.68	12.22	72.62
JACKSONVILLE, FL	B	FT LAUDERDALE, FL	B	177.68	36	326.49	32.35	72.51
ASHLAND, VA	A	EDISON, NJ	B	233.90	28	294.37	2.82	72.37
WALNUT, CA	A	ONTARIO, CA	B	470.47	1	16.42	33.03	72.05
BALTIMORE, MD	B	EDISON, NJ	B	291.54	15	163.8	8.46	72.05
DURHAM, NC	B	RALEIGH, NC	A	166.74	5	24.69	1.73	72.03
ELIZABETHTOWN, PA	C	EDISON, NJ	B	484.99	14	154.79	47.95	71.93
HAMMOND, LA	B	SAINT ROSE, LA	A	178.01	11	50.45	46.37	71.66
KENNESAW, GA	A	BUFORD, GA	B	109.26	12	52.16	41.23	71.62
CHARLOTTE, NC	B	DURHAM, NC	B	214.85	11	142.91	5.39	70.92
TOLLESON, AZ	A	ONTARIO, CA	A	924.77	13	325.73	44.43	70.29
BUFORD, GA	B	KINGS MOUNTAIN, NC	A	124.94	12	181.93	6.48	69.49
MONTGOMERY, NY	B	BEDFORD, MA	B	305.48	14	207.1	45.51	69.31
KNOXVILLE, TN	A	INDIANAPOLIS, IN	A	211.19	28	357.33	36.4	69.27
SUMNER, WA	A	WILSONVILLE, OR	A	1152.05	7	167.24	201.51	68.60
RICHLAND, MS	A	HAMMOND, LA	B	151.14	7	134.83	32.65	68.43
CLEVELAND, OH	A	ROMULUS, MI	B	675.76	11	157.26	156.29	68.34
OBETZ, OH	B	ROMULUS, MI	A	265.73	11	196.73	54.13	68.05
URBANDALE, IA	A	CHAMPLIN, MN	B	262.07	11	265.71	29.87	67.04
NASHVILLE, TN	A	WHITESTOWN, IN	C	208.92	12	310.4	5.07	66.67
EARTH CITY, MO	B	WAUKEGAN, IL	B	489.64	19	339.32	144.98	64.98
SOLON, OH	B	ROMULUS, MI	B	290.12	10	168.71	170.06	64.43
TAMPA, FL	B	FT LAUDERDALE, FL	A	168.23	7	263.33	79.46	63.37
RALEIGH, NC	A	SWEDESBORO, NJ	B	154.31	15	406.5	45.25	63.14
SAINT ROSE, LA	A	HOUSTON, TX	A	139.16	3	339.78	33.67	61.51
KANSAS CITY, MO	A	CHAMPLIN, MN	B	432.87	14	455.52	109.35	61.17
OMAHA, NE	B	URBANDALE, IA	A	60.55	2	135.73	213.92	59.56
TULSA, OK	A	FLOWER MOUND, TX	A	109.32	10	259.8	235.19	58.28
SALT LAKE CITY, UT	B	DIXON, CA	B	634.59	11	671.42	58.92	57.69
BUFFALO, NY	D	ROMULUS, MI	B	817.87	18	278.25	460.19	57.16
AUSTIN, TX	A	FLOWER MOUND, TX	A	387.07	4	221.14	369.48	54.54
WEST VALLEY, UT	A	SUMNER, WA	A	460.39	24	857.36	88.11	54.39
CHAMPLIN, MN	B	WAUKEGAN, IL	B	718.21	18	402.09	781.33	41.34
ENGLEWOOD, CO	B	VISALIA, CA	C	445.02	19	1115.1	253.54	39.69
FLOWER MOUND, TX	A	KANSAS CITY, MO	A	52.37	4	537.87	545.43	37.12
WILSONVILLE, OR	A	TRACY, CA	A	1134.05	12	628.54	825.79	34.71
HARLINGEN, TX	A	FLOWER MOUND, TX	A	156.29	1	545.94	858.79	25.25
FRANKLIN, MA	A	ROMULUS, MI	B	793.31	8	716.49	1073.55	19.56

Table 5: The set of 29 non-overlapping two-stops routes

STOP 1	Cust	STOP 2	Cust	$\bar{S}_{x,y}$	$\hat{F}_{x,y}$	$d_{x,y}$	$R_{x,y}$	$V(x,y)$
DENVER, CO	A	AURORA, CO	C	1372.51	26	17.24	26.95	86.57
FIFE, WA	C	SUMNER, WA	A	1365.91	16	6.59	4.99	84.49
ATLANTA, GA	C	KENNESAW, GA	A	249.77	38	27.64	0	82.93
WHITESTOWN, IN	C	INDIANAPOLIS, IN	A	502.52	36	21.81	42.31	82.74
SHAKOPEE, MN	C	CHAMPLIN, MN	B	617.58	28	34.86	-3.05	82.37
DIXON, CA	B	TRACY, CA	A	2132.50	12	75.37	148.95	81.64
CARSON, CA	B	ONTARIO, CA	B	1000.84	26	52.9	101.18	80.21
CHANDLER, AZ	B	TOLLESON, AZ	A	854.32	12	36.94	7.51	78.64
GRAND PRAIRIE, TX	B	GRAND PRAIRIE, TX	C	519.12	16	10	10	78.11
OVERLAND PARK, KS	B	KANSAS CITY, MO	A	744.10	12	11.57	21.58	78.01
DEPEW, NY	B	BUFFALO, NY	D	596.09	14	11.9	16.42	77.75
BIRMINGHAM, AL	B	OLIVE BRANCH, MS	A	255.41	35	216.62	1.68	76.86
INDIANAPOLIS, IN	B	WAUKEGAN, IL	B	129.32	39	233.09	-0.03	76.81
ORLANDO, FL	C	TAMPA, FL	B	251.09	30	84.52	66.66	76.46
BROWNS SUMMIT, NC	C	SWEDESBORO, NJ	B	210.56	49	408.58	-14.44	76.30
GAINESVILLE, FL	A	JACKSONVILLE, FL	B	127.81	29	71.14	60.76	75.84
RAMULUS, MI	B	RAMULUS, MI	A	263.88	14	10	10	75.64
EDISON, NJ	B	MONTGAERY, NY	B	329.80	20	93.11	14.31	75.58
HOUSTON, TX	B	HOUSTON, TX	A	260.97	12	10	10	75.00
EDWARDSVILLE, IL	A	EARTH CITY, MO	B	267.96	14	32.43	15.27	74.87
KINGS MOUNTAIN, NC	A	CHARLOTTE, NC	B	143.50	14	32.89	2.11	74.44
FT LAUDERDALE, FL	B	FT LAUDERDALE, FL	A	290.24	7	10	10	73.64
GREENSBURG, PA	A	WARRENDALE, PA	B	291.56	11	51.29	25.62	73.21
BEDFORD, MA	B	FRANKLIN, MA	A	463.84	10	45.63	58.52	73.09
NAMPA, ID	C	WILSONVILLE, OR	A	1052.37	23	432.68	12.22	72.62
DURHAM, NC	B	RALEIGH, NC	A	166.74	5	24.69	1.73	72.03
HAMMOND, LA	B	SAINT ROSE, LA	A	178.01	11	50.45	46.37	71.66
AAHA, NE	B	URBANDALE, IA	A	60.55	2	135.73	213.92	59.56
TULSA, OK	A	FLOWER MOUND, TX	A	109.32	10	259.8	235.19	58.28

Given the information in Table 5, we can compute the total annual savings of the selected 29 routes (Equation 10). Figure 12 depicts the distribution of the cost savings among the routes. The routes are sorted based on the MODA score, not savings.

$$\sum_{x,y} \bar{S}_{x,y} \times \hat{F}_{x,y} = \$282{,}667 \tag{10}$$

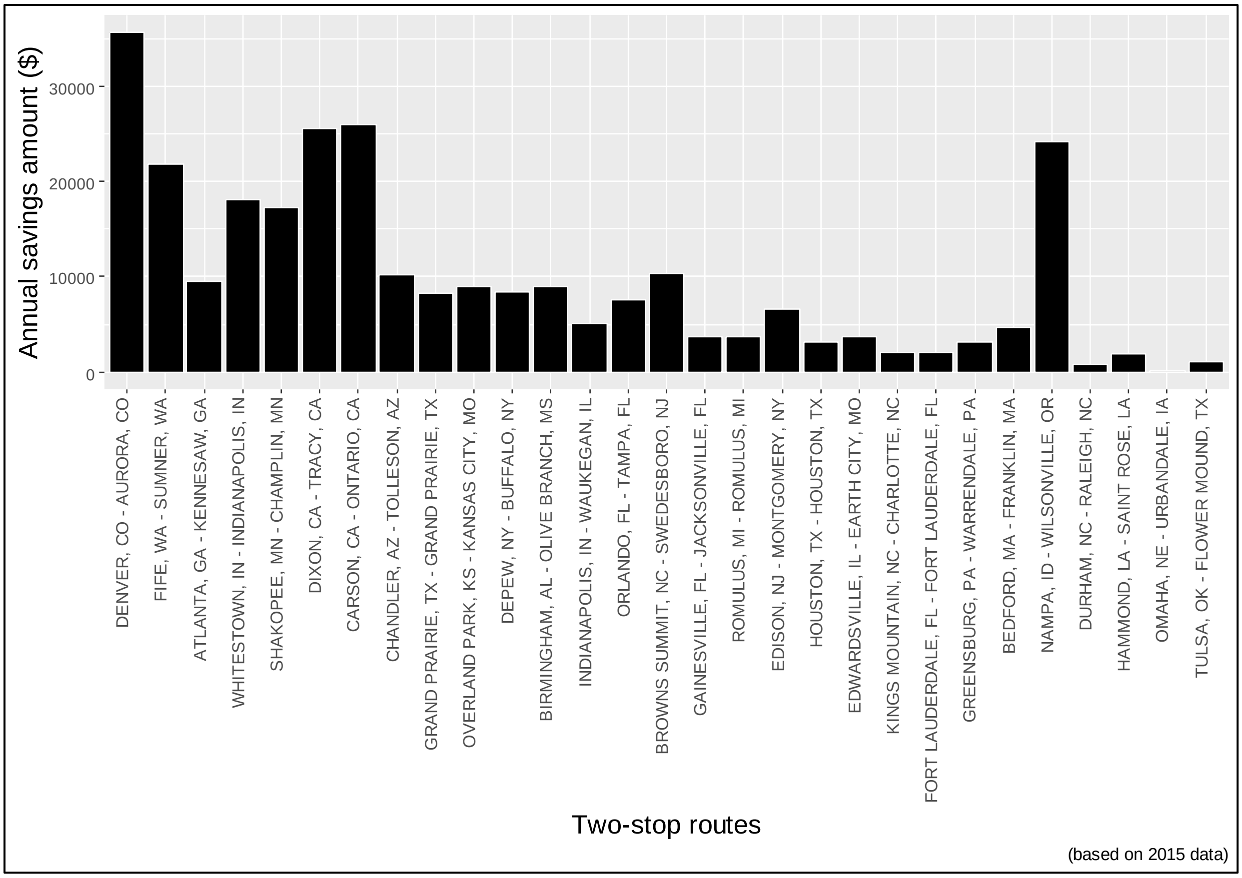

Figure 12: Distribution of annual savings for the 29 selected two-stop routes. Sorted based on MODA score

4.5.2 Sensitivity Analysis

There are various sensitivity analyses that can be insightful for the problem, decision factors and the existing relationships between them. In this section, we discuss some of them to provide a better understanding on the robustness of results and the relationships between parameters of the model.

As previously indicated, the result of the case study is based on devoting more importance to carrier acceptance rather than cost savings (Equation 7. We gave 70% total weight to carrier acceptance criteria ($d_{x,y}, R_{x,y}$) and 30% to cost savings criteria ($\bar{S}_{x,y}, \hat{F}_{x,y}$) in order to generate more appealing results for carriers. A key question is what if we flip the importance weights and give more importance to cost savings? To answer, we need to re-compute $V(x, y)$

using new weights in Equation 7. (i.e. 0.4 for savings magnitude, 0.3 for savings consistency, 0.15 for out-of-route miles, and 0.15 for proximity of stops.

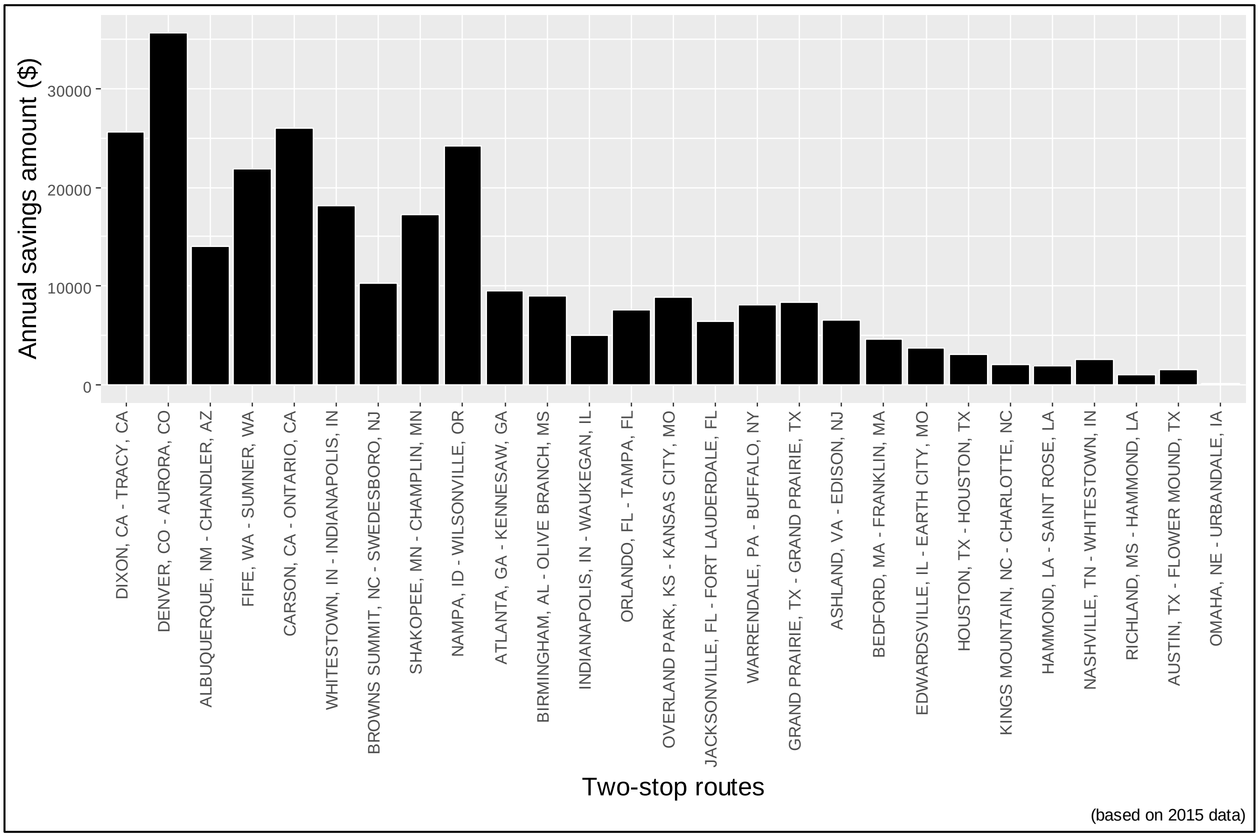

Figure 13: Distribution of the annual savings for the 27 selected two-stop routes based on more importance toward cost savings impact. Sorted based on MODA score

The model selects 27 routes with the total annual cost savings of \$ 283,128, slightly higher than the original scenario (Figure 13). The new set contains 7 new routes that did not exist in the original set of 29 routes. Another important change is the order of routes in Figure 13. The new set places different routes in high priority to create more appeal to the shipper rather than carriers.

While the estimated cost savings of both scenarios are almost the same ($\cong \$283K$), it is essential to compare their performances in terms of proximity of stops and out-of-route miles. Figure 14 illustrates this comparison clearly. The comparison suggests that flipping the weights to favor shippers' criteria increases proximity of stops significantly, while no noticeable change is realized on out-of-route miles. This change would certainly impact carriers' behavior toward accepting loads. Therefore, since the estimated cost savings of both scenarios are virtually the same, it makes sense for the shipper to offer the original set of routes to carriers.

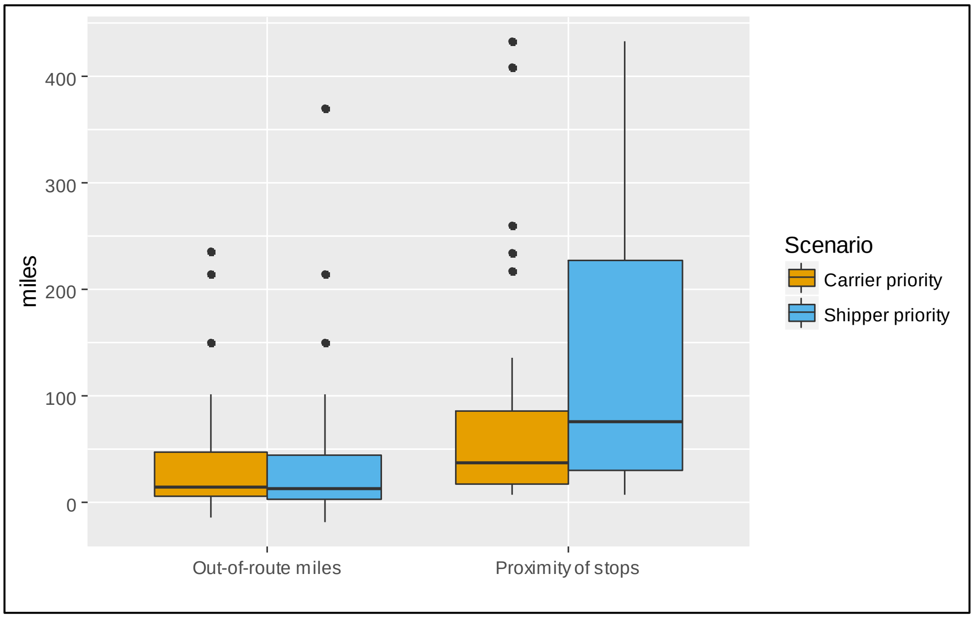

Figure 14: Performance of two scenarios, that are based on weight assignment, on carrier acceptance criteria

Additional sensitivity analyses, especially on the value function forms, would be helpful to provide further insights.

4.6 Summary

Multi-stop truckload (MSTL) is a new mode of transportation that is gaining market share rapidly due to numerous benefits and in some sectors due to necessity. However, there are multiple, perhaps conflicting, objectives associated with selecting and operating a multi-stop transportation system. The decision-making process is bi-lateral, where both shipper's and carrier's objectives need to be considered. While shippers seek to minimize the shipping cost, carriers want to avoid extra costs and risks associated with MSTL delivery. In this chapter, we identified the key objectives of both parties, offered a procedure to estimate the potential cost savings of two-stop delivery system, developed a multi-objective decision analysis (MODA) model for the problem, and applied the model on a case study in the healthcare supply chain sector. The model enables the decision-makers to incorporate all their preferences into the problem-solving process and trade-off between objectives to obtain a solution that satisfies both parties. The case study revealed the potential of MSTL in a sizeable supply network as well as the capability of MODA in producing desirable solutions for both parties. Solutions are designed from the perspective of shippers to offer multi-stop routes that minimize the shipping cost and are appealing to carriers.

4.7 Future Work

This area has abundant future research potential as various industries are becoming more interested in multi-stop transportation. Here we indicate some of the immediate future research related to this chapter:

- Besides MODA, there are other methodologies that could be more useful depending on the context and the application area. We think quadratic assignment is another suitable modeling approach for the ranking of the identified routes based on the decision criteria.
- A comprehensive sensitivity analysis on additional key elements of the MODA would be useful to better understand the robustness of solutions. To achieve this goal, another case study on a larger network with more than two stops would be substantially insightful.
- Two of the main assumptions in the case study are lower cube limit (*LL*) of 1200 ft^3 and upper cube limit (*UL*) of 2000 ft^3. The assumption for *UL* is very conservative, given the fact that the normal capacity of a truck is 3300 ft^3. It would be interesting to investigate the relationship of *UL* with various performance measures such as cost savings, proximity of stops, or out-of-route miles.

References

Akiyama, H., Yano, Y., 2010. Current conditions pertaining to truck deliveries to large retail stores in Japan, and efforts to reduce them. Procedia - Soc. Behav. Sci. 2, 6343–6354. doi:10.1016/j.sbspro.2010.04.043

Brar, G.S., Saini, G., 2011. Milk Run Logistics : Literature Review and Directions, in: Proceedings of the World Congress on Engineering.

Caldwell, E.R., Fisher, B.C., 2008. The Impact of Lead Time on Truckload Transportation Rates.

Chen, X., Tsai Yang, S. lin, 2016. Multi-Stop Trucking : A Study on Cost and Carrier Acceptance. MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

Clemen, R.T., Reilly, T., 2001. Making Hard Decisions with Decision Tools®. Isbn 0-534-36597-3. doi:10.1198/tech.2002.s729

Du, T., Wang, F.K., Lu, P.Y., 2007. A real-time vehicle-dispatching system for consolidating milk runs. Transp. Res. Part E Logist. Transp. Rev. 43, 565–577. doi:10.1016/j.tre.2006.03.001

Eroglu, D.Y., Rafele, C., Cagliano, A.C., Sevilay, M.S., Ippolito, M., 2014. Simultaneous Routing and Loading Method for Milk-Run Using Hybrid Genetic Search Algorithm. XII Int. Logist. Supply Chain Congr. 48–57.

Hosseini, S.D., Akbarpour Shirazi, M., Karimi, B., 2014. Cross-docking and milk run logistics in a consolidation network: A hybrid of harmony search and simulated annealing approach. J. Manuf. Syst. 33, 567–577. doi:10.1016/j.jmsy.2014.05.004

Kirkwood, C.W., 1996. Strategic decision making. Wadsworth Publ. Co.

Mendoza, A., Ventura, J.A., 2009. Estimating freight rates in inventory replenishment and supplier selection decisions. Logist. Res. 1, 185–196. doi:10.1007/s12159-009-0018-5

Parnell, G.S., Driscoll, P.J., Henderson, D.L., 2011. Decision making in systems engineering and management. John Wiley & Sons.

Parnell, G.S., Trainor, T.E., 2009. Using the swing weight matrix to weight multiple objectives. 19th Annu. Int. Symp. Int. Counc. Syst. Eng. INCOSE 2009 1, 283–298. doi:10.1002/j.2334-5837.2009.tb00949.x

Sadjadi, S.J., Jafari, M., Amini, T., 2009. A new mathematical modeling and a genetic algorithm search for milk run problem (an auto industry supply chain case study). Int. J. Adv. Manuf. Technol. 44, 194–200. doi:10.1007/s00170-008-1648-5

Xu, Q.H., 2003. Milk Run Practice and Application about Cycle Pick Up Model in Shangahi GM (J) Automotive Accessories.

Yildiz, H., Ravi, R., Fairey, W., 2010. Integrated optimization of customer and supplier logistics at Robert Bosch LLC. Eur. J. Oper. Res. 207, 456–464. doi:10.1016/j.ejor.2010.03.044

You, Z., Jiao, Y., 2014. Development and application of milk-run distribution systems in the express industry based on saving algorithm. Math. Probl. Eng. 2014. doi:10.1155/2014/536459

5 CONCLUSION AND FUTURE WORK

This dissertation contributes to the area of supply chain management by developing various models for the current challenges in the area. The first chapter presented a comprehensive model to accurately approximate the cost savings of continuous replenishment programs (CRP). Unlike other published works in the literature, the model did not impose assumptions that normally do not hold in practice, especially in transportation and handling cost components. The model provides a full perspective to the projected cost savings of both partners which helps them to further understand the financial benefits of CRP and reach a sustainable profit sharing agreement. The model is implemented for a healthcare manufacturer to analyze its CRP relationships with its independent distributors. The results revealed that savings significantly vary across the channels depending on product mix, demand characteristics, handling and transportation requirements, etc. The results showed that distributors generally gain more savings within the shared cost components, which are inventory holding and order processing. The results also indicated that small channels have more relative savings potential but do not generate substantial monetary savings. This chapter can be expanded in different ways. One of the future directions is modeling the decision of starting a CRP relationship based on both cost savings and qualitative factors that may outweigh the financial benefits. Factors such as trust, team attitude, cooperation, power shift, implementation capability and shared business philosophy are amongst them. Another future work is performing a comprehensive sensitivity analysis on the key elements of the model to provide useful insights for various types of supply chain arrangements.

The second chapter developed multi-functional logistics metrics to monitor the efficiency of inventory holding, transportation, and order processing operations. The key contribution of this study is providing data-driven and traceable metrics that quantify possible tendencies in favoring efficiency of certain functions over others. In addition, a statistical process control (SPC) system is devised for monitoring the metrics over time. Various time series behaviors of the metrics are discussed and appropriate SPC systems are suggested for them. The metrics are particularly useful for monitoring CRP/VMI relationships. Our collaboration with a group pf active VMI partners in the healthcare sector confirmed that VMI relationships need a verifiable efficiency measurement system that measures the state of efficiency within different functions of logistics. The metrics capture the trade-off in gaining efficiency between two functions while optimal trade-off levels are estimated as target levels. The metrics enable decision makers to communicate factually and work collectively toward the optimal efficiency levels using statistical control systems. One of the most immediate future works for this chapter is studying the correlation of these metrics and identify the circumstances that lead the metrics to provide redundant information. Besides, analytical investigations, applying the metrics on different case studies would help in achieving a better understanding of those circumstances. Investigating whether companies see the suggested optimal trade-off levels aligned with their strategic objectives and other considerations would also be a valuable future work.

The third chapter is focused on an area that has significant research potential. Multi-stop truckload (MSTL) is gaining market share in the transportation sector and that is going to continue in the coming years due to an increasing shift in moving distribution centers to urban areas. Carriers are the operator of delivery tasks and their behavior toward multi-stop truckload offers is the main focus of this chapter. In this chapter, we model the decision of offering MSTLs considering both the cost savings potential and carrier acceptance criteria. While shippers seek to minimize the shipping cost, carriers want to avoid extra costs and risks associated with MSTL delivery. We showed how to compute the savings potential of two-stop (drop-offs) delivery routes, which is the most common form of MSTL business. We developed a multi-objective decision analysis (MODA) model to incorporate shippers' priorities with carriers' priorities to select the most desirable two-stop routes. While savings magnitude and savings consistency are important decision criteria for shippers, out-of-route miles and proximity of stops are considered as the key decision factors for carriers. The application of the model is discussed in a case study, where the potential of MSTL in a sizeable supply network as well as the capability of MODA in producing desirable solutions for both parties are shown. The results and sensitivity analysis showed that changing the elements of MODA causes the model to select a different set of routes with a different preference order. The sensitivity analysis suggested that proximity of stops is much more sensitive than other performance measures to changing the weights of the model. Among the immediate future works, modeling the problem with alternative approaches such as the quadratic assignment, and sensitivity analysis on various elements of the model would be insightful.