Centralized and Decentralized Warehouse Logistics Collaboration

By

Shiman Ding

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Philip M Kaminsky, Chair
Professor Zuo-Jun Max Shen
Professor Shachar Kariv

Summer 2017

Abstract

Centralized and Decentralized Warehouse Logistics Collaboration

by

Shiman Ding

Doctor of Philosophy in Industrial Engineering and Operations Research

University of California, Berkeley

Professor Philip M Kaminsky, Chair

In an emerging trend in the grocery industry, multiple suppliers and retailers share a warehouse to facilitate horizontal collaboration, in order to lower transportation costs and increase outbound delivery frequencies. Typically, these systems (sometimes known as Mixing and Consolidation Centers) are operated in a decentralized manner, with little effort to coordinate shipments from multiple suppliers with shipments to multiple retailers. Indeed, implementing coordination in this setting, where potential competitors are using the same logistics resources, could be very challenging. In this thesis, we characterizes the loss due to this decentralized operation, in order to develop insight into the value of making the extra effort and investment necessary to implement some form of coordinated control. To do this, we consider a setting where several suppliers ship to several retailers through a shared warehouse, so that outbound trucks from the warehouse contain the products of multiple suppliers. We extend the classic one warehouse multi-retailer analysis of Roundy (1985) to incorporate multiple suppliers and per truck outbound transportation cost from the warehouse, and develop a cost lower bound on centralized operation as benchmark. We then analyze decentralized versions of the system, in which each retailer and each supplier maximizes his or her own utility in a variety of settings, and we analytically bound the ratio of the cost of decentralized to centralized operation, to bound the loss due to decentralization. We find that easy-to-implement decentralized policies are efficient and effective in this setting, suggesting that centralization (and thus, coordination effort intended to lead to some of the benefit of centralization) does not bring significant benefits. In a computational study, we explore how system parameters impact the relative performance of this system under centralized and decentralized control. Finally, we consider a stochastic version of this model of decentralized collaboration, where we assume independent Poisson demand occurs at each retailer for all products. To coordinate replenishment, each retailer follows an aggregate (Q,S) policy, i.e., an order is placed to raise inventory position to S whenever total demand since the last order at that retailer reaches Q.

In this, setting demand at the warehouse can be well-approximated by a compound Poisson process, and thus inventory at the warehouse is managed via an (s,S) policy. We develop optimal and heuristic algorithms to optimize parameter settings in this model.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my advisor, Professor Philip Kaminsky, the smartest and nicest person I know at Berkeley, for his wise guidance, creative insights, kind understanding and strong support of my doctoral studies. Phil is always supportive and available when I need advice in research, and I have learned a lot from him professionally and personally via numerous discussions, which have extraordinarily improved my research capability. Next I am really grateful to Professor Zuo-Jun Max Shen, Professor Anil Aswani, and Professor Shachar Kariv for being my qualifying and dissertation committee and for their helpful feedback and encouragement of my research. I also would like to express my special gratitude to Professor Xin Guo, who has always supported me both in research and in life, which really warms and guides me when I have a hard time.

I am so lucky to receive support and encouragement from friends at Cal: Dimin and Kelly, all the night chats and gym talks we had boosted me up and made me strong; Zhao, your accompanyment was so treasured that all of our best and worst memories shape me. IEOR is a relatively small but great family, Birce, Cheng, Haoyang, Jiaying, Kevin, Long, Min, Mo, Nan, Nguyen, Renyuan, Stewart, Sheng, Tugce, Wei, Xu and Ying, all those days with you guys in Etch on course studies, research discussions, and "after-seminar" activities made my life so sweet and memorable at Cal.

I would also like to thank my friends outside Cal. Though we were not in the same place in the last five years, I never felt we were apart. Po, you are always the one who calmed me down when I was in anxious, just like in high school:) Tiantian and Qin, thanks for sharing my joyful times and still being my friends after enduring all my complaints for the five years:) Runqi, my best moment in 2016 is to know "another me" in Alberta! Jun, thanks for discussing all those interesting problems with me, and watched me crying out when I broke up:). Yuanjun, it is enjoyable that every one of our random chats would turn into research discussions:)

Next I would like to thank my parents, who devote their deepest love to me without expecting any return. Though we are separated by the Pacific Ocean, my dear baba and mama are always there whenever I need them. You always let me choose whatever I want and fully support many of my seemingly stupid decisions ever since childhood. Your love makes me strong enough to face all the difficulties in life.

Now, thank you Jerry:) Every moment we spent together, on badminton, on hiking, on (random) research chats, on cooking and food tasting, on everything, is so enjoyable and unparalleled. We've been friends for years (can't believe...), and I'm looking forward:)

Hmm... finally thank YOU. Thanks for choosing this adventure five years ago; thanks for being patient and never giving up; thanks for always following your heart! You know, life is fun:)

Chapter 1

Introduction

An emerging paradigm for horizontal logistics collaboration in the grocery industry centers on large third-party warehouses, sometimes called *mixing and consolidation centers (MACC)* that multiple suppliers use as warehouses or mixing centers, and from which multiple retailers order mixed-supplier truckloads. Anecdotal evidence (and common sense) suggests that these warehouses not only lower transportation costs by more fully utilizing outbound transportation (that is, sending fuller trucks to retailers), but also increase service level by increasing the frequency of deliveries to retailers. Typically, systems like these are operated in an effectively decentralized fashion – individual suppliers decide when to make deliveries to the warehouse, and individual retailers order from the warehouse, where ordering information goes directly to the warehouse to assemble deliveries, and also to the suppliers for billing and planning purposes. There is no coordination between different suppliers or between different retailers. Indeed, implementing coordination in this setting would be quite challenging, as firms would be required to share order information, and to coordinate deliveries.

Our goal in this thesis is to develop some insight into the value of working to overcome these challenges to implement coordination in this type of system. Would the additional effort (in contract design, information technology, trust development, etc.) be worth it? While this is obviously a complex question, in this thesis we analyze two scenarios: models of deterministic demand, and models of stochastic and stationary demand.

Using deterministic models, we begin to explore the value of centralized coordination by developing and analyzing a stylized continuous time constant demand model. Specifically, we consider a set of suppliers each of which individually ships a supplier-specific product to a single warehouse. In turn, trucks containing products from multiple suppliers are shipped to retailers, each of which faces constant, deterministic demand for each of (or a subset of) the products. Outbound shipping costs are charged both in the simple case per shipment (Chapter 3) as well as in a more complicated model per capacitated truck (Chapter 4), so that we can explore how much this warehouse facilitates the effective utilization of outbound transportation. Additionally, holding cost is charged at the warehouse and retailers. In this stylized setting, we explore the following question: How much can system costs potentially be reduced if decentralized control, where each retailer places its own (uncoordinated) order and suppliers independantly react to these orders, is replaced with centralized control that coordinates each order and delivery?

We also consider a stochastic model setting in Chapter 5, where demands at different retailers occur randomly. Collaboration can still be facilitated via warehouse and truck sharing, and we still focus on decentralized policies in which suppliers and retailers can keep their information private. In our model, we assume unmet demand at retailers is fully backordered, but suppliers must meet all the orders to maintain system stability. Thus, when inventory at the warehouse is insufficient to cover retailer orders, suppliers need to pay additional fees to expedite supplies from other sources. Under such an assumption, we propose the following decentralized policy: retailers all follow an aggregate (Q, S) policy so that an order is placed if total demand for all products reaches Q; similarly, suppliers implement a typical (s, S) policy. We provide an algorithm to solve for the optimal policy parameters that is both efficient and easy to implement.

1.1 Background

A significant portion of supply chain cost and environmental impact can be attributed directly to logistics costs. In 2013, transportation represented 8.8% of U.S. Gross Domestic Product, or roughly \$1.4 T (trillion), and truck transportation specifically accounted for nearly 31% of that (US Department of Transportation, 2014). Additionally, packaging and commercial warehousing account for \$141B (billion) and \$39B in annual expenditures respectively, excluding all costs directly incurred by manufacturers, distributors and retailers (Armstrong Associates, 2016). In spite of this, the current logistics network is far from efficient. For example, in spite of the obvious economies of scale, trucks often operate at on average 60% of capacity (McKinnon, 2010), because firms often find it more cost effective to either use their own fleets, or to pay truckload carriers, even if they don't have full loads to ship, and indeed, trucks often travel completely empty, because shippers often have more goods to ship in one direction than in another. At the same time, inventory levels are often significantly higher than they have to be, as firms attempt to reduce transportation costs at the expense of inventory costs, and storage facilities are used inefficiently, due to the time required to fill and pick from these facilities, and the incompatible shapes and sizes of packaging.

Indeed, over the past decades, individual firms have increasingly optimized their supply chains, so that resources are used as efficiently as possible given the individual firm's requirements. It is becoming clear, however, that in order to achieve necessary scale to efficiently utilize logistics resources, all but the largest firms will have to work together, and specifically, focus on horizontal collaboration. While much of supply chain management focuses on collaboration, the traditional focus, both in industry and in academia, has been on so-called vertical collaboration. Vertical collaboration refers to the development of partnerships, alliances, and strategic contracts between firms at different stages of production or distribution within the same supply chain, from manufacturing in factories, transportation to warehouses, to sales in retail stores. While vertical collaboration leads to considerable supply chain efficiency, in many cases it does not lead to sufficient volume to enable firms to fully and effectively utilize supply chain assets. Horizontal collaboration, on the other hand, can be an effective approach for addressing these concerns.

Specifically, horizontal collaboration is the cooperation among companies with similar customers and consumers that share assets at the same stage of the supply chain – production, transportation, warehouse storage, local retail selling, etc. It is collaboration across rather than along the supply chain. For example, firms could share trucks, warehouses and other logistic resource. Utilizing this type of approach, each individual company could cost-effectively increase the replenishment frequency with fewer products transferred each time.

Articles in trade magazines occasionally describe examples of ongoing ad-hoc horizontal collaboration efforts. Hershey and Ferrero, two competing chocolate makers, have a collaborative logistics agreement focusing on shared warehousing, transportation, and distribution in North America (Cassidy, 2011). Colgate actively seeks opportunities to collaborate on shipping, and has an effort in place in the Los Angeles area with Sunny Delight (Trunick, 2011). According to a case study published by KANE, a midsized third-party logistics provider, raisin and dried fruit distributor Sun-maid collaborated with manufacturers of candy, pet foods, condiments, and others, to shared shipping resources, leading to a 62% reduction in Sun-maid's outbound logistics costs (KANE Is Able, Inc., 2011). In a pilot program in the UK, Nestle and Mars combined deliveries to Tesco, a large UK grocery chain. Over three months, over 7500 miles of truck travel were removed from the system (Meall, 2010). Four retail companies in France (Ballot and Fontane, 2010) shared warehouses and trucks, reallocated gains, costs and tariffs, and averaged savings of 29%. JSP, a manufacture of lightweight plastic bags, and Hammerwerk Fridingen, a manufacture of advanced medal components (CO3, 2011) shared transportation and significantly increased truck fill rate in the term of both volume and weight.

Indeed, these types of collaborative logistics arrangements are not new, although existing efforts such as those described above tend to be small and involve a few entities with a narrowly focused scope (Benavides and Swan, 2012). Industry professionals have recognized that significant levels of collaboration in the industry could lead to breakthrough reductions in the cost and environmental impact of logistics (Gue and Forger., 2014). Even with the significant potential, however, real obstacles stand in the way of widespread implementation of collaboration. Most interestingly, different stakeholders have strong views of why this type of system should work, but currently does not – evidence suggests that collaboration efforts are more likely to fail than to succeed with participants in a recent survey suggesting a 20% success rate for these efforts (Benavides and Swan, 2012). Surveys suggest that technological obstacles related to data sharing and securities are certainly not insurmountable. The critical challenge seems to be the need for companies (sometimes competing companies) to trust one another enough to achieve the needed levels of asset and information sharing (Cassidy, 2011). The MACCs introduced in the first paragraph of this paper may be a way to address this issue, at least to the extent that they can facilitate some of the benefits of collaboration, such as effectively utilizing outbound transportation, while limiting the need for information sharing and coordination.

This research is motivated by our work with a specialized third party logistic provider that has established large warehouse centers, where both manufacturers and retailers combine their mixing centers and distribution centers into large facilities. As shown in Figure 1.1, the warehouse center serves both as inbound warehouse for retailers and outbound warehouse for suppliers.

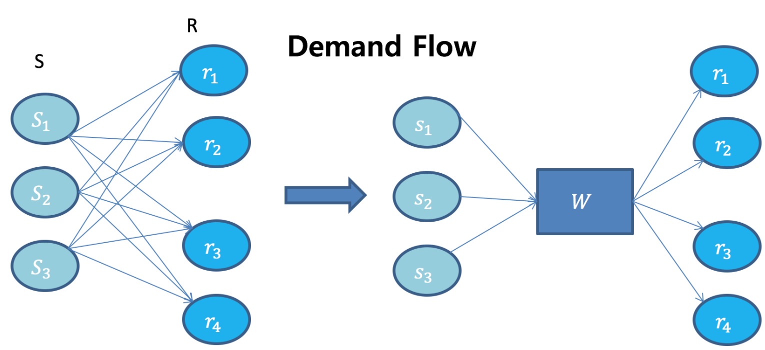


Figure 1.1: 3PL Warehouse Collaboration

Currently, such 3PL systems are operated in a decentralized way. Suppliers can send their products to the warehouse center and pay for inventory holding cost, until it is ordered by some down stream retailer. Retailers can order from all suppliers and combine their products in the same shipment to increase truck load. According to Ryder, their solution involves "Harnessing the power of collaboration to ship less-than-truckload quantities at truckload prices".

By sharing these warehouse centers, each retailer and supplier reduces supply chain inventory, shortens their replenishment cycle length thus reducing inventory uncertainty, and increases truck usage since each company only needs to fill in a portion of truck's capacity. Effectively, these systems enable horizontal collaboration by providing necessary information sharing and limited coordination. A different provider, Ryder Integrated Logistics, estimates that the use of their MACCs leads to the following gains Ryder System (2014): "Average savings of 6 to 22% on freight costs, 99.8% on-time delivery, reductions in out-of-stocks from 2 to 14% and lead-time reductions of 3 to 7 days."

In this thesis, we explore how information affects horizontal collaboration and how should we operates such system in reality. To assess the value of information sharing, we compare two kinds of models: centralized systems and decentralized ones. As we illustrate in Figure 1.2, in a centralized system, every party shares all the information and makes collective decisions, to optimize system wide performance. In a decentralized systen, each supplier and retailer makes their own manufacturing and shipping plans, without giving out any private information, just as how ES3 and Ryder operates their warehouse centers now.

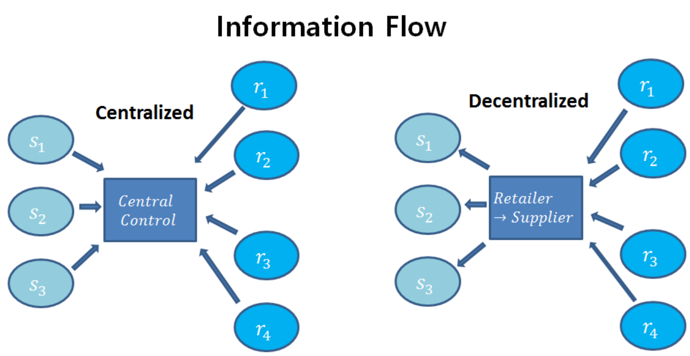


Figure 1.2: Centralized and Decentralized Collaboration

Chapter 2

Literature Review

Several streams of literature are closely related to our problem. For deterministic models, the One Warehouse Multi-Retailer Problem (OWMR) and the Joint Replenishment Problem (JRP) are two models that are the building blocks of our centralized model. The OWMR Problem considers collaboration between a single supplier and many retailers, while the JRP focuses on coordination between a single retailer and many suppliers. In addition, several authors have considered logistics collaboration in a decentralized setting and logistic models with generalized transportation costs, which are closely related to our model in Chapter 4. For stochastic models, we reviewed many classical policies as well as multi-echelon collaboration.

2.1 The One Warehouse Multi-Retailer Problem

The One Warehouse Multi-Retailer problem has been widely studied, most notably by Roundy (Roundy, 1985). Roundy considered a distribution system with one warehouse and multiple retailers. Constant demands occur at each retailer and no backorder or shortage is allowed. The warehouse orders from an outside supplier with unlimited supply and replenishes the retailers' inventories.

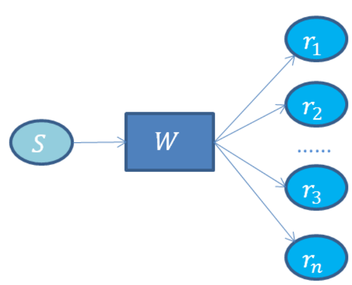


Figure 2.1: Classic One Warehouse Multi-Retailer (OWMR) Problem

Arkin et al. showed that the OMWR problem is NP-hard (Arkin et al., 1989). Roundy introduced both the classic Power-of-Two (PoT) policy and the q-optimal integer ratio policy and proved worst-case effectiveness of 94% and 98% respectively (Roundy, 1985). He also developed an algorithm that obtains the optimal PoT solution efficiently (Roundy, 1985). Later, Roundy extended the results to a multi-stage production problem, where he viewed product at different stages as "different products" and showed that for assembly systems, a PoT policy can be found in $O(NlogN)$ time with 98% effectiveness (Roundy, 1986). Lu and Posner developed additional heuristics for the problem,one of which finds a solution in $O(N)$ with error bound 2.014%, and another of which finds a solution in $O(NlogN/\sqrt{\epsilon})$, where ϵ is the error bound. Mitchell and Joseph extended the model to include backlogging (Mitchell, 1987). Gallego and Simchi-Levi extended the cost structure to incorporate transportation costs, where trucks are capacitated and each truck will generate a fixed transportation cost independent of its truckload (Gallego and Simchi-Levi, 1990). They demonstrated that fully loaded direct shipping policy is at least 94% effective whenever the economic lot size of each retailer is at least 71% of truck capacity (Gallego and Simchi-Levi, 1990).

The research detailed above is restricted to the deterministic demand single-supplier setting. For the multi-supplier problem, Maxwell and Muckstadt (Maxwell and Muckstadt, 1985) and Muckstad and Roundy (Muckstadt and Roundy, 1987) considered a class of nested and stationary policies. In this context, stationary means that the order intervals are constant for each specific retailer, and nested means that whenever the warehouse replenishes an item, a shipment will be sent to each retailer. Both provided $O(NlogN)$ algorithms to compute a PoT policy with 94% effectiveness compared to any possible nested policy (Maxwell and Muckstadt, 1985), (Muckstadt and Roundy, 1987). Viswanathan and Mathur considered vehicle routing together with the multi-item one warehouse inventory problem, where the warehouse is a break-bulk center and does not keep any inventory. They presented a heuristic that develops a stationary and nested joint replenishment policy (Viswanathan and Mathur, 1997).

The OWMR with warehouse capacity constraints has also received much attention. Alhough our primary focus is on transportation capacity, these warehouse-capacity models share characteristics with our models, in that they both restrict the quantity that can be shipped to retailers. Jackson et al. considered a more general supply chain distribution network, where supply is limited because of production capacity (Jackson et al., 1988). They derived a closed-form solution for model with one capacity constraint, and lagrangian multiplier method for model with multiple capacity constraints (Jackson et al., 1988). Federgruen and Zheng analyzed a similar system, and characterized the effectiveness of an algorithm they developed within class of PoT policies (Federgruen and Zheng, 1993). Konur analyzed an integrated inventory control and transportation problem, with capacitated order quantity (Konur, 2014).

In simple model in Chapter 3, our centralized model setting is related to Roundys multi-item model, but we consider a more a general class of policies (rather than just nested policies), and we consider inventory control at retailers and the warehouse. We provide an efficient algorithm to find a Power- of-Two policy with 94% effectiveness compared to any optimal policy.

In Chapter 4, our centralized setting is most closely related to Roundy's multi-item model, but we incorporate a more general transportation cost structure and consider a more a general class of policies (rather than just nested policies), and we consider inventory control at retailers and the warehouse.

2.2 The Joint Replenishment Problem (JRP)

The Joint Replenishment Problem is a widely studied special case of the One Warehouse Multi-Retailer problem. The JRP arises when a retailer purchases several items from a single supplier, and pays a so-called major fixed ordering cost that is independent of the number of different products in the order, a minor fixed ordering cost that is incurred for each product included in an order, and holding cost. The retailer must decide when to order and which items to purchase in each order to minimize the total cost while satisfying demand. The assumptions for the classical JRP are similar to those of the EOQ; demand is deterministic and uniform, no shortage or quantity discount allowed, and holding cost is linear.

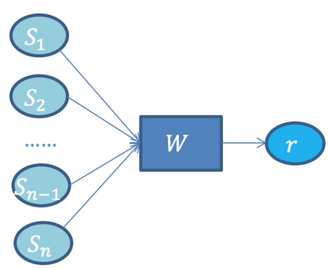


Figure 2.2: Classic Joint Replenishment Problem (JRP) Problem

However, Arkin et al. showed that the JRP is NP-hard (Arkin et al., 1989). Khouja et al. provided a comprehensive literature review of the Joint Replenishment Problem between 1989 and 2005 (Khouja and Goyal, 2008). For the classical JRP, many heuristics have been proposed: Silver's algorithm (Silver, 1976) was improved by Goyal and Belton (Goyal and Belton, 1979), and later, by Kapsi et al. (Kaspi and Rosenblatt, 1991). Jackson et al. considered policies with fixed reorder intervals and restricted policies to PoT policies (Jackson et al., 1985). They proposed a similar sorting algorithm to the one Roundy proposed for the OWMR, and achieved the same 94%-effectiveness.

2.3 Decentralized Logistic Systems

Although centralized models obviously minimize total system costs, they may not be implementable. Thus, several recent papers have focused on decentralized logistics systems. Chen et al. considered pricing and replenishment strategies simultaneously in a system of one supplier and multiple retailers, showed that decentralization may lead to supply chain inefficiency (Chen et al., 2001b), and developed approaches for coordinating the supply chain in this setting (Chen et al., 2001a). Abdul-Jalbar et al. considered the OWMR problem in both centralized and decentralized settings (Abdul-Jalbar et al., 2003). In their retailer-driven decentralized model, they truncated the EOQ order intervals to rational numbers for retailers and applied the algorithm of Wagelmans et al. (Wagelmans et al., 1992) to find the replenishment plan for suppliers (Abdul-Jalbar et al., 2003). Chen and Chen analyzed a multi-item two-echelon supply chain with deterministic demand (Chen and Chen, 2005). However in their decentralized system, orders for different items are separately placed, and hence no joint replenishment or collaboration is achieved (Chen and Chen, 2005). Baboli et al. considered both centralized and decentralized policies for a single-supplier single-retailer model with transportation cost (Baboli et al., 2008). They proposed an algorithm to find the optimal centralized solution, and a cost-sharing mechanism through a quantity discount for the decentralized model (Baboli et al., 2008). Chu and Leon presented a decentralized OWMR model, where coordination is achieved under negotiation (Chu and Leon, 2008). All of these papers considered decentralized performance for a single supplier model within a single class of policies; in contrast, in our decentralized models, we evaluate the performance of decentralized policies for multi-supplier and multi-retailer settings, and specifically develop, analyze, and optimize three stationary policies with guaranteed effectiveness.

2.4 Logistic Collaboration and General Transportation Cost Models

The bulk of papers discussed above consider simple transportation cost structures – either linear or affine in the quantity being shipped – and no transportation collaboration. In fact, truck sharing for collaboration has been widely studied in literature. Burns et al. analytically compared the tradeoff between direct shipping and truck sharing strategies (Burns et al., 1985). Daganzo explored the operational benefits of freight consolidation through a consolidation center (Daganzo, 1988b). Daganzo also compared systems adopting a truck sharing strategy of several delivery stops with systems featuring an intermediate warehouse center, and provided conditions under which the latter system outperforms the former (Daganzo, 1988a).

Logistic systems with more realistic, often nonlinear costs and characteristics of transportation, such as per truck transportation cost and alternative modes, have also been considered in the literature. Aucamp considered a step-wise transportation cost function in an EOQ setting and provided a closed-form solution (Aucamp, 1982). Later, Chan et al. considered a finite horizon OWMR problem with quantity discount for transportation cost (Chan et al., 2002). They showed that there exists a *zero-inventory-ordering* (ZIO) policy with no more than $\frac{4}{3}$ of the optimal cost. They also proposed two heuristics to find an effective ZIO policy (Chan et al., 2002). Jin and Muriel considered a OWMR problem with truckload cost (Jin and Muriel, 2009). They proposed a Lagrangian decomposition approach to solve the finite horizon OWMR. For the continuous OWMR model, Gallego and Simchi-Levi evaluated the effectiveness of a direct shipping strategy for OWMR (Gallego and Simchi-Levi, 1990). Rieksts and Ventura proposed models with two modes of transportation: truckload (TL) and less than truckload (LTL), with corresponding cost structures. They considered simple model with one supplier and one retailer in both finite and infinite horizon setting, as well as a more complicated OWMR system. They provided algorithms to obtain the optimal solution as well as a heuristic PoT policy (Rieksts and Ventura, 2008). Chen derived optimal policies for multi-stage serial and assembly systems where materials flow in fixed batches, like a full truckload (Chen, 2000).

While these papers have elements of either shared transportation or complex logistics cost structures, as far as we know Chapter 4 in this thesis is the first to incorporate both per truck transportation cost along with logistics collaboration in a multiple supplier multiple retailer setting, and characterize the cost of decentralization in this setting.

2.5 Stochastic Multi-Echelon Collaboration

All the work we reviewed above focused mainly on deterministic models, so that inventory position can be relatively easy to control beforehand, because demand rate is fixed and known. But when we come to stochastic version of the problem, much fewer results has been obtained.

2.5.1 Classic Stochastic Inventory Models

Several classes of policies have been widely studied in literature, and two main streams of policies exist, depending on whether inventory is reviewed periodically or continuously.

The (r, Q) policy is a widely used continuous review policy: when inventory position drops to a reorder point r, an order of size Q is placed. When lead time is fixed and unmet demand is back-ordered, the (r, Q) policy has been shown to be optimal Snyder and Shen (2011). Much research has been devoted to calculating optimal (r, Q) policies efficiently when unmet demand is back-ordered Federgruen and Zheng (1992), Axsäter (2000), De Bodt and Graves (1985). Our decentralized retailer policy is an aggregate version of an (r, Q) policy, in which total demand for all products is monitored, and a total of Q units (potentially of different products) are ordered.

The (s, S) policy is commonly used for periodic review systems with non-zero fixed ordering cost. The (s, S) policy is very similar to (r, Q), except that each time the order quantity may be different, as the inventory position is always raised to the same level S.The optimality of (s, S) policy in a system with infinite horizon and with fixed ordering cost was shown inZheng (1991). A variety of researchers have developed efficient algorithms to find the optimal (s, S) policy Federgruen and Zipkin (1984), Zheng and Federgruen (1991). In our decentralized supplier policy, suppliers are facing several streams of demand with Erlang interarrival times and multinomial demand quantity. We develop, analyze, and optimize a continuous review (s, S)-type policy,

A key issue in this stochastic decision relates to the policy implemented by suppliers or the warehouse when there is insufficient inventory to meet downstream orders from retailers. The unmet demand can either be backordered or lost. However in contrast to backorder models, there is considerably less research on lost sales models in general.One reason for this is that lost sales models are more difficult to model explicitly, and thus more complicated to optimize Hadley and Whitin (1963). Although many researchers have simplified the system by allowing only one outstanding order and thus significantly reducing the complexity, the form of optimal solution to the general problem is still unknown Archibald (1981). A comprehensive review of lost sales models can be found in Bijvank and Vis (2011).

2.5.2 The Stochastic One Warehouse Multi-retailer Problem

The One Warehouse Multi-retailer (OWMR) Problem was first studied by Roundy Roundy (1985). This two-echelon system with one upsteam warehouse supplying several downstream retailers is NP-hard, even in the simple deterministic setting without backorder Arkin et al. (1989). When demand is stochastic, the best known exact algorithm is the projection algorithm, which iterates over all possible values of S_0, the warehouse base-stock level Axsäter (1990). However, total cost is nonconvex and each round of cost evaluation is computationally expensive. Several heuristics have been proposed for OWMR problem. Özer and Xiong decomposed the system into several serial systems, solved each decomposed problem individually, and summed up to get a solution Özer and Xiong (2008). Rong proposed a similar decomposition-aggregation heuristic for distributon systesm Rong et al. (2011). To the best of our knowledge, all existing methods are either computationally expensive or heuristic. In contrast, the algorithm we propose later in this paper can solve for aggregate (s, S) policy accurately and effectively, while only requiring linear time cost evaluations.

2.5.3 The Stochastic Multi-supplier Multi-retailer System

Very little research has been conducted in a stochastic setting involving coordination among multiple suppliers and multiple retailers. Hong et al. considered a vender-managed inventory system where unmet demand is lost and show that under certain scenarios, such a system performs better than systems that allow back-orders Hong et al. (2016). Taleizadeh et al. considered a system of multiple suppliers and multiple retailers with uniform demand. To determine optimal reorder point and safety stock, they proposed a harmony search algorithm to solve the corresponding nonlinear integer program Taleizadeh et al. (2011).

In contrast, in Chapter 5 we consider a setting with a single warehouse, multiple suppliers, and multiple retailers, where there is no centralized control. This models the decentralized collaboration we have seen achieved via MACCs. Retailers order different products from the MACC in a single order using an aggregate (Q, S) policy, i.e., inventory position is raised to S whenever the total demand at a retailer reaches Q since its last replenishment. Suppliers replenish inventory at the warehouse in order to satisfy orders from all retailers.

Chapter 3

The Uncapacitated Model

3.1 The Centralized One Warehouse Multi-Supplier Multi-Retailer Problem

For a stylized model of the collaboration between multiple suppliers and retailers when all of the parties are controlled in a centralized fashion, we initially extend the classical OWMR model to a multi-supplier setting (this can alternatively be viewed as a multi-product setting, since we assume that each supplier provides a unique product). We consider centralized control of orders and transportation, so each supplier can benefit from combining deliveries to the warehouse that are ultimately intended for different retailers in order to save on transportation, and retailers can save costs by ordering products from different suppliers simultaneously. In the following, we first extend the integer ratio policy of (Roundy, 1985) to our setting, and we provide an efficient algorithm to solve the relaxed version of the integer-ratio problem, show that this is a lower bound on the cost of an arbitrary policy, and use this to find a 94% effective Power-of-Two policy. Where possible, we follow the development and notation of (Roundy, 1985), extending the analysis to our multiple supplier/multi-item case.

Although Roundy and his co-authors also considered the multi-item case (Muckstadt and Roundy, 1987), they restrict their approach to nested and stationary policies, which as they show can be arbitrarily bad.

3.1.1 The Model

In the One Warehouse Multi-Retailer Multi-Supplier (OWMRMS) problem, we consider a two-echelon supply chain with n suppliers and m retailers sharing a common warehouse. The warehouse serves as both outbound storage for suppliers and a distribution center for the retailers. Each supplier manufactures a unique product (so we can refer to products and suppliers interchangeably in what follows) and supplies that product to all of the retailers. Each retailer faces constant (possibly zero) consumer demand for each product, and this demand must be met without backlogging. A fixed cost is incurred whenever a supplier replenishes its inventory at the warehouse, or a retailer places an order from the warehouse. Linear inventory holding costs are charged both at the warehouse and at retailers. The holding costs, fixed costs, and demand rates are constant over time and can be different at each facility. We detail our notation below, where for ease of exposition we use i as the index associated with suppliers and j as the index associated with retailers:

- $S = \{1, \cdots, n\}$: the set of suppliers
- $R = \{1, \cdots, m\}$: the set of retailers
- d_{ij}: demand for product from supplier i at retailer j
- k_i^s: fixed cost for delivery from supplier i to the warehouse
- k_j^r: fixed cost for delivery from the warehouse to retailer j
- h^i: holding cost rate of the product of supplier i at the warehouse
- h_{ij}: holding cost rate of the product from supplier i at retailer j
- $h'_{ij} = h_{ij} - h^i$: echelon holding cost rate of product from supplier i at retailer j; We assume that $h^i \leq h_{ij}$, so that $h'_{ij} \geq 0$.

We call the problem of minimizing the long-run average cost under centralized control in the One Warehouse Multi-Retailer Multi-Supplier (OWMRMS) problem, while satisfying all demand, Problem **(P)**. Unfortunately, this problem is $NP-hard$ even with only one supplier (Arkin et al., 1989). As the optimal policy can be extremely complicated, with non-stationary order quantities and order intervals, we focus on quantifying the effectiveness of heuristics. Note that in this deterministic model, replenishments are made only when inventory drops to zero (a *zero-inventory-ordering*, or ZIO policy); otherwise, we can postpone the replenishment or reduce earlier shipping quantity and save inventory cost.

To develop a lower bound on the optimal long run average cost of Problem **(P)**, we characterize an heuristic policy for the problem, find a lower bound on that heuristic, and then show that this lower bound is a lower bound on any feasible solution to the problem.

Among all possible feasible policies, we first consider an easy to implement policy, the integer ratio policy, which is a special case of a periodic policy. A periodic policy is one in which each supplier and retailer have a constant order interval, and all order intervals have a common multiple. We represent the order intervals for suppliers as $\boldsymbol{T^s} = (T_1^s, T_2^s, \cdots, T_n^s)$, where T_i^s is the order interval for supplier i. Similarly, $\boldsymbol{T^r} = (T_1^r, T_2^r, \cdots, T_m^r)$ captures order intervals for retailers.

An integer ratio policy is a periodic policy where $\forall i \in S$ and $\forall j \in R$, $T_i^s/T_j^r \in \mathbb{W}$, where $\mathbb{W}$ is the set of all positive integers and their reciprocals.

We introduce the following problem for OWMRMS under an integer ratio policy:

$$\begin{aligned}
(\mathbf{PI}) \;\; \min \mathbf{C_I}(\mathbf{T^s}, \mathbf{T^r}) \quad & = \sum_{i\in S} \frac{k_i^s}{T_i^s} + \sum_{j\in R} \frac{k_j^r}{T_j^r} + \frac{1}{2} \sum_{i\in S, j\in R} \max{(T_i^s, T_j^r)} d_{ij} h^i \\
& + \frac{1}{2} \sum_{i\in S, j\in R} T_j^r d_{ij} h'_{ij} \\
s.t. \quad & T_i^s, T_j^r > 0, \forall i \in S, j \in R \\
& T_i^s/T_j^r \in \mathbb{W}, \forall i \in S, j \in R \qquad (3.1)
\end{aligned}$$

This cost function $C_I(\boldsymbol{T^s}, \boldsymbol{T^r})$ is similar to the function Roundy developed for the single supplier OMWR. Roundy showed that for the single supplier case, $C_I(T_1^s, \boldsymbol{T^r})$ is the exact total cost for the class of integer ratio policies, and the optimal cost of a relaxed problem (defined later), $C_{IR}(T_1^{s*}, \boldsymbol{T^{r*}})$, is a lower bound on the long run cost of any policy (Roundy, 1985). We first extend these results to our problem with multiple suppliers. We show the cost function is still exact for integer ratio policies, and the optimal solution to its relaxed problem (**PIR**), where integer ratio constraints (3.1) are relaxed in (**PI**), is a lower bound for any arbitrary policy.

In the cost function $C_I(\boldsymbol{T^s}, \boldsymbol{T^r})$, the first two terms are the fixed costs, and the third and fourth terms are the echelon inventory holding costs at the warehouse and retailers. In particular, as in Roundy's paper (Roundy, 1985), we consider two cases. If retailer j orders more frequently than the warehouse does for product from supplier i, $T_i^s > T_j^r$, the echelon inventory at the warehouse follows a standard "sawtooth" pattern with order interval of T_i^s, and the inventory at the retailer has interval T_j^r. If the retailer orders no more frequently than the warehouse does for product from supplier i, $T_i^s \le T_j^r$, then we need only consider inventory cost incurred at the retailer.

Thus, $C_I(\boldsymbol{T^s}, \boldsymbol{T^r})$ is the exact total cost for the OWMRMS when an integer ratio policy is applied, so (**PI**) exactly models integer ratio policies. Later we prove, in Section 3.1.6, that the optimal solution to the integer ratio relaxation of (**PI**), which we denote (**PIR**) and define in Section 3.1.2, is a lower bound on any feasible policy for Problem (**P**).

To simplify notation, it is traditional for this class of problems to substitute $g_{ij} =$

$\frac{1}{2}d_{ij}h'_{ij}$ and $g^{ij} = \frac{1}{2}d_{ij}h^i$. Making this substitution into Problem (**PI**), we get:

$$(\mathbf{PI'}) \ \min \quad C_{I'}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i\in S}\frac{k_i^s}{T_i^s} + \sum_{j\in R}\frac{k_j^r}{T_j^r} + \sum_{i\in S, j\in R}\max\left(T_i^s, T_j^r\right)\cdot g^{ij} + \sum_{i\in S, j\in R} T_j^r \cdot g_{ij}$$
$$s.t. \quad T_i^s, T_j^r > 0, \forall i \in S, j \in R$$
$$T_i^s / T_j^r \in \mathbb{W}, \forall i \in S, j \in R$$

3.1.2 Relaxation of the Problem

We can relax the integer ratio constraints in Problem (**PI′**) to get

$$(\mathbf{PIR}) \ \min \quad C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i\in S}\frac{k_i^s}{T_i^s} + \sum_{j\in R}\frac{k_j^r}{T_j^r} + \sum_{i\in S, j\in R}\max\left(T_i^s, T_j^r\right)\cdot g^{ij} + \sum_{i\in S, j\in R} T_j^r \cdot g_{ij}$$
$$s.t. \quad T_i^s, T_j^r > 0, \forall i \in S, j \in R$$

Observe that the objective function of this problem remains convex.

Given any solution $(\boldsymbol{T^s}, \boldsymbol{T^r})$ to (**PIR**), following Roundy (Roundy, 1985), we can divide the retailers into sets as follows, where each retailer will be in multiple sets, depending on the number of products sold at that retailer:

$$L_i = \{j \in R : T_j^r < T_i^s\}, \ E_i = \{j \in R : T_j^r = T_i^s\}, \ G_i = \{j \in R : T_j^r > T_i^s\}.$$

Of course, if there is only one supplier, there are only three sets, and each retailer is in only one of these sets – indeed, this observation is a critical part of Roundy's solution approach. However, in our case there are multiple suppliers, so that as we observed above, for each distinct supplier i, L_i, G_i, and E_i can be different. Thus, we need to find an alternative approach to partition $R \cup S$.

A natural approach is to group retailers and suppliers with the same order interval together. We denote that partition $P(U_1), P(U_2), \cdots, P(U_k)$, where U_l is the order interval. That is,

$$P(U_l) = \{i \in S : T_i^s = U_l\} \cup \{j \in R : T_j^r = U_l\}.$$

Without loss of generality, we order U_l such that $U_1 < U_2 \cdots < U_k$. Therefore the corresponding optimal cost can be decomposed as follows:

$$\begin{aligned}
C_{IR}(\boldsymbol{T_i^s}, \boldsymbol{T_j^r}) &= \sum_{i\in S} \frac{k_i^s}{T_i^s} + \sum_{j\in R} \frac{k_j^r}{T_j^r} + \sum_{i\in S, j\in R} \max(T_i^s, T_j^r) \cdot g^{ij} + \sum_{i\in S, j\in R} T_j^r \cdot g_{ij} \\
&= \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} \frac{k_j^r}{T_j^r} + \sum_{\substack{i\in S, j\in L_i \\ or\ i\in S, j\in E_i}} T_i^s \cdot g^{ij} + \sum_{j\in R, i:j\in G_i} T_j^r \cdot g^{ij} + \sum_{i\in S, j\in R} T_j^r \cdot g_{ij} \\
&= \sum_{l\in\{1,\cdots,k\}} \Big(\frac{K(U_l)}{U_l} + H(U_l) \cdot U_l \Big) \\
&\triangleq \sum_{l\in\{1,\cdots,k\}} C_{IRS}^{U_l}(\boldsymbol{T^r}, \boldsymbol{T^s}),
\end{aligned}$$

where

$$K(U_l) = \sum_{i\in P(U_l)} k_i^s + \sum_{j\in P(U_l)} k_j^r,$$

and

$$H(U_l) = \sum_{\substack{i\in P(U_l) \\ j\in L_i}} g^{ij} + \sum_{\substack{i\in P(U_l) \\ j\in P(U_l)}} g^{ij} + \sum_{\substack{j\in P(U_l) \\ i:j\in G_i}} g^{ij} + \sum_{\substack{i\in S \\ j\in P(U_l)}} g_{ij}.$$

$K(U_l)$ and $H(U_l)$ can be viewed as aggregate fixed ordering cost and holding cost for $P(U_l)$.

Given these definitions, (**PIR**) can decomposed into a series of convex subproblems, one for each partition $P(U_l)$:

$$\begin{aligned}
(\mathbf{PIR_l})\ \ \min\ & C_{IRS}^{U_l}(\boldsymbol{T^r}, \boldsymbol{T^s}) = \frac{K(U_l)}{U_l} + H(U_l) \cdot U_l \\
s.t.\ & U_l > 0, \forall l
\end{aligned}$$

By the first order necessary condition, we obtain the optimal solution to ($\mathbf{PIR_l}$) as:

$$U_l^* = \sqrt{\frac{K(U_l)}{H(U_l)}}. \tag{3.2}$$

Thus, given any partition of retailers and suppliers, we can calculate aggregate fixed cost and holding cost in each group, and thus find optimal order intervals. Therefore our problem (**PIR**) is equivalent to finding the optimal partition of $R \cup S$. The conditions for optimality are specified in the following theorem:

Theorem 1. *The following conditions are necessary and sufficient for* $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *to be the optimal order intervals in* **(PIR)***:*
$(C1)$ *For the corresponding ordered partition* $P(U_1), P(U_2), \cdots, P(U_k)$ *of* $R \cup S$*, where* $U_1 < U_2 < \cdots < U_k$*, we have* $U_l = \sqrt{\frac{K(U_l)}{H(U_l)}}$*, where* $K(U_l)$ *and* $H(U_l)$ *are the aggregate fixed and holding cost of* $P(U_l)$*.*
$(C2)$ $\forall l, \forall$ *subset* $P \subset P(U_l^*)$*,*

$$\sqrt{\frac{\sum_{i \in P} k_i^s + \sum_{j \in P} k_j^r}{\sum_{i \in S, j \in P} g_{ij} + \sum_{i \in P, j \in L_i} g^{ij} + \sum_{j \in P, i: j \in G_i} g^{ij} + \sum_{j \in P, j \in P} g^{ij}}} \leq U_l^* \leq \sqrt{\frac{\sum_{i \in P} k_i^s + \sum_{j \in P} k_j^r}{\sum_{i \in S, j \in P} g_{ij} + \sum_{j \in P, j \in P} g^{ij}}}$$

.

The proof of Theorem 1 can be found in Appendix A.1. In Theorem 1, $(C1)$ guarantees first order conditions for each group, and $(C2)$ ensures no deviation from the partition could improve cost.

3.1.3 MSIRR Algorithm

From our previous analysis, we know that finding the optimal solution to (**PIR**) is equivalent to finding an optimal partition of $R \cup S$. Therefore our goal is to determine the partition and the order intervals simultaneously while satisfying conditions $(C1)$ and $(C2)$ in Theorem 1.

In this subsection, we introduce an algorithm which we call Multi-Supplier Integer Ratio Relaxation (MSIRR) algorithm, and prove that the solution obtained by the MSIRR algorithm satisfies $(C1)$ and $(C2)$ and is thus optimal to (**PIR**). The MSIRR algorithm determines order intervals from the largest to the smallest iteratively. In each iteration, some suppliers and retailers have already been assigned to some ordered groups in previous iterations, and the remaining are unassigned. Existing groups are ordered by order intervals of the group, where more recently formed groups have smaller order interval. For those suppliers and retailers, we sequentially assume that each supplier, each retailer, and each pair of one supplier and one retailer has the largest order interval among all the unassigned candidates, and calculate the corresponding order interval. We pick the largest such order interval as a candidate to enter the set of assigned groups. If it is smaller than the order intervals of all existing groups (and in particular, smaller than the most recently formed group), the corresponding supplier, retailer, or pair forms a new group, otherwise it is assigned to the most recently formed group, and we recalculate the order interval for the new group according to (3.2). If the new order interval is larger then that of some other existing group, we combine the groups again until $(C1)$ is satisfied.

In the following Algorithm, $\bar{S}$ and $\bar{R}$ are the sets of unassigned suppliers and retailers, τ is the most recently formed group, and $List_G$ is the list of all existing groups.

Algorithm: MSIRR Algorithm

Initialize: $\tau \leftarrow \varnothing$, $list_G \leftarrow \varnothing$, $T_{cur} \leftarrow \infty$, $\bar{R} \leftarrow R$, $\bar{S} \leftarrow S$, $\boldsymbol{T^s} \leftarrow \mathbf{0}$, $\boldsymbol{T^r} \leftarrow \mathbf{0}$;

while $\bar{R} \cup \bar{S} \neq \varnothing$ **do**

  forall $i \in \bar{S}$ **do**

    $T_i^s \leftarrow \sqrt{\frac{k_i^s}{\sum_{j \in \bar{R}} g^{ij}}}$;

  end

  forall $j \in \bar{R}$ **do**

    $T_j^r \leftarrow \sqrt{\frac{k_j^r}{\sum_{i \in \bar{S}} g^{ij} + \sum_{i \in \bar{S}} g_{ij}}}$;

  end

  forall $i \in \bar{S}$, $j \in \bar{R}$ **do**

    $T_{ij} \leftarrow \sqrt{\frac{k_i^s + k_j^r}{\sum_{i \in \bar{S}} g^{ij} + \sum_{i \in \bar{S}} g_{ij} + \sum_{k \in \bar{S}} g^{kj} + \sum_{k \in \bar{R}} g^{ik} - g^{ij}}}$;

  end

  $i_0 \leftarrow \arg\max_{i \in \bar{S}} T_i^s, \quad j_0 \leftarrow \arg\max_{j \in \bar{R}} T_j^s, \quad (\hat{i_0}, \hat{j_0}) \leftarrow \arg\max_{i \in \bar{S}, j \in \bar{R}} T_{ij}$;

  if $\max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}}) < T_{cur}$ **then**

    $T_{cur} \leftarrow \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$, append τ to the end of $list_G$;

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \{i_0\}$, $\bar{S} \leftarrow \bar{S} \setminus \{i_0\}$;

    else if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \{j_0\}$, $\bar{R} \leftarrow \bar{R} \setminus \{j_0\}$;

    else

      $\tau \leftarrow \{\hat{i_0}, \hat{j_0}\}$, $\bar{S} \leftarrow \bar{S} \setminus \{\hat{i_0}\}$, $\bar{R} \leftarrow \bar{R} \setminus \{\hat{j_0}\}$;

    end

  else if $\max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}}) = T_{cur}$ **then**

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{i_0\}$, $\bar{S} \leftarrow \bar{S} \setminus \{i_0\}$;

    else if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{j_0\}$, $\bar{R} \leftarrow \bar{R} \setminus \{j_0\}$;

    else

      $\tau \leftarrow \tau \cup \{\hat{i_0}, \hat{j_0}\}$, $\bar{S} \leftarrow \bar{S} \setminus \{\hat{i_0}\}$, $\bar{R} \leftarrow \bar{R} \setminus \{\hat{j_0}\}$;

    end

  else

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{i_0\}$;

    else if $T_{j_0}^r = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{j_0\}$;

    else

      $\tau \leftarrow \tau \cup \{\hat{i_0}, \hat{j_0}\}$;

    end

    $\bar{S} \leftarrow \bar{S} \cup \tau \cap S$, $\bar{R} \leftarrow \bar{R} \cup \tau \cap R$;

    $K_\tau \leftarrow \sum_{i \in \tau \cap \bar{S}} k_i^s + \sum_{j \in \tau \cap \bar{R}} k_j^r$;

    $H_\tau \leftarrow \sum_{\substack{i \in \tau \\ j \in \bar{R} \setminus \tau}} g^{ij} + \sum_{\substack{i \in \tau \\ j \in \tau}} g^{ij} + \sum_{\substack{j \in \tau \\ i \in \bar{S} \setminus \tau}} g^{ij} + \sum_{\substack{i \in \bar{S} \\ j \in \tau}} g_{ij}$;

    $T_{cur} \leftarrow \sqrt{\frac{K_\tau}{H_\tau}}$;

    forall $i \in \tau$, $j \in \tau$ **do**

      $T_i^s \leftarrow T_{cur}$; $T_j^r \leftarrow T_{cur}$;

    end

    $\bar{S} \leftarrow \bar{S} \setminus \tau$, $\bar{R} \leftarrow \bar{R} \setminus \tau$;

  end

end

In each iteration, it takes $O(nm)$ operations to find i_0, j_0 and $(\hat{i_0}, \hat{j_0})$. In each iteration, either a new group is formed, or two existing groups are combined. Thus in the entire algorithm, grouping happens at most $n+m$ times, because combined groups are never later partitioned, and this leads to an overall complexity of $O(nm \cdot (n+m))$.

Lemma 1. *This MSIRR algorithm finds the optimal solution to problem* (**PIR**).

A proof of Lemma 1 can be found in Appendix A.2. If $\boldsymbol{T^{s*}}$ and $\boldsymbol{T^{r*}}$ denote the optimal solution to (**PIR**) obtained from MSIRR algorithm, the corresponding average cost is

$$C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) = \sum_{l \in \{1, \cdots, k\}} 2\sqrt{K(U_l^*) \cdot H(U_l^*)}. \tag{3.3}$$

where $K(U_l^*)$ and $H(U_l^*)$ are the aggregate fixed and holding cost in optimal partition $P(U_l^*)$, as defined above.

3.1.4 Second Order Cone Approach

As we analyze in previous Section 3.1.2, (**PIR**) is nonlinear but convex. In this section, we propose a different approach to transform the problem into a conic program, which can be solved directly by standard optimization software packages such as CPLEX, Gurobi or Mosek.

By introducing auxiliary variables T_{ij}, t_i and $t_j \geq 0$ to represent nonlinear term in the objective, we can reformulate (**PIR**) as:

$$\begin{aligned}
(\textbf{PIR2}) \ \min \quad & C_{SOC}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i \in S} k_i^s t_i + \sum_{j \in R} k_j^r t_j + \sum_{i \in S, j \in R} T_{ij} g^{ij} + \sum_{i \in S, j \in R} T_j^r g_{ij} \\
s.t. \quad & T_{ij} \geq T_i^s, \forall i \in S, j \in R && (3.4) \\
& T_{ij} \geq T_j^r, \forall i \in S, j \in R && (3.5) \\
& t_i \geq \frac{1}{T_i^s}, \forall i \in S && (3.6) \\
& t_j \geq \frac{1}{T_j^r}, \forall j \in R && (3.7) \\
& T_i^s, T_j^r, T_{ij}, t_i, t_j \geq 0, \forall i \in S, j \in R
\end{aligned}$$

We notice in (**PIR2**), (3.4) and (3.5) guarantees $T_{ij} \geq max(T_i, T_j)$. In a minimization problem with all non-negative coefficients, (**PIR2**) is equivalent to (**PIR**).

Next we transform the fractional constraint (3.6),

$$t_i \geq \frac{1}{T_i^s}, T_i^s > 0 \Leftrightarrow t_i T_i^s \geq 1, T_i^s \geq 0$$
$$\Leftrightarrow \frac{(t_i + T_i^s)^2}{4} \geq \frac{(t_i - T_i^s)^2}{4} + 1, T_i^s \geq 0$$
$$\Leftrightarrow \frac{t_i + T_i^s}{2} \geq \sqrt{\frac{(t_i - T_i^s)^2}{4} + 1}, T_i^s \geq 0 \tag{3.8}$$

(3.8) is a second order cone, and similarly (3.7) can be reformulated as

$$\frac{t_j + T_j^r}{2} \geq \sqrt{\frac{(t_j - T_j^r)^2}{4} + 1}, T_j^r \geq 0 \tag{3.9}$$

Therefore, **(PIR2)** can be reformulated as:

$$\begin{aligned}
(\textbf{PSCOP}) \ \min \quad & C_{SOC}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i\in S} k_i^s t_i + \sum_{j\in R} k_j^r t_j + \sum_{i\in S, j\in R} T_{ij} g^{ij} + \sum_{i\in S, j\in R} T_j^r g_{ij} \\
s.t. \quad & T_{ij} \geq T_i^s, \forall i \in S, j \in R \\
& T_{ij} \geq T_j^r, \forall i \in S, j \in R \\
& \frac{t_i + T_i^s}{2} \geq \sqrt{\frac{(t_i - T_i^s)^2}{4} + 1}, \forall i \in S \\
& \frac{t_j + T_j^r}{2} \geq \sqrt{\frac{(t_j - T_j^r)^2}{4} + 1}, \forall j \in R \\
& T_i^s, T_j^r, T_{ij}, t_i, t_j \geq 0, \forall i \in S, j \in R
\end{aligned}$$

Notice that the objective of **(PSCOP)** is linear, and all constraints are either lnear or quadratic, which fits into Second Order Cone Program (SCOP). Later in Section 3.6, we compare the efficiency of our MSIRR Algorithm and SCOP using CPLEX, and show our algorithm is much faster than SCOP and can handle problems of larger scale.

3.1.5 The Power-of-Two Policy

In this section, we consider a special case of integer ratio policies, the Power-of-Two (PoT) policy. Recall that a PoT policy is a periodic ordering policy such that $T_i^s \in \{m \in \mathcal{N} : 2^m \cdot T_0\}$, where T_0 is a base order interval. A PoT policy is implemented so that two parties with the same order interval always order at the same time.

$$(\textbf{PPOT}) \ \min C_{POT}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i\in S} \frac{k_i^s}{T_i^s} + \sum_{j\in R} \frac{k_j^r}{T_j^r} + \sum_{i\in S, j\in R} \max(T_i^s, T_j^r) g^{ij} + \sum_{i\in S, j\in R} T_j^r g_{ij}$$
$$s.t. T_i^s, T_j^r \in \{2^k \cdot T_0, k \in \mathcal{N}\}, \forall i \in S, j \in R$$

In the following we show that the PoT policy obtained from $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is an easy-to-implement policy with worst case ratio of 94%. In other words, the cost of the PoT policy is at most 6% more than the optimal cost of (**PIR**).

We use $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$, the optimal solution to (**PIR**), to construct a PoT solution as follows:

$$T_{i,P}^{s*} = \min\{2^m T_0 : 2^m T_0 \geq \frac{T_i^{s*}}{\sqrt{2}}\}.$$

That is,

$$\frac{T_i^{s*}}{\sqrt{2}} \leq T_{i,P}^{s*} < \sqrt{2} T_i^{s*}.$$

Recall that in the MSIRR algorithm, we greedily determine order intervals of suppliers and retailers sequentially from largest to smallest. In each iteration, we either create a new group or combine existing groups whenever condition $(C2)$ of Theorem 1 is violated. Recall that we group suppliers and retailers with the same order interval into a partition, so that some members of the set $P(U_l^*)$ (the set of suppliers and retailers with order interval U_l^*) may be combined after rounding, because different U_l^* values could be rounded to the same PoT interval. However, to facilitate our analysis, we continue to consider them separately. Hence, the previous partition for $R \cup S$ still applies, so that $P(U_l^*) = P(U_l^{P*})$. Next, we bound the worst case performance for this feasible PoT policy.

Theorem 2. $C_{POT}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) \leq \frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2}) C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. *That is, the total cost of the PoT policy we obtain from rounding* $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *is no more than about* 1.06 *the optimal centralized cost.*

A proof of Theorem 2 can be found in Appendix A.3. In the computational analysis in Section 3.6, we show that when per truck transportation cost is moderate, PoT rounding works well. In fact, an extremely large or extremely small per truck transportation cost is required to drive the capacitated optimal solution far from the uncapacitated solution, to come close to the worst case bound of 2.

3.1.6 The Lower Bound

For the OWMR system with one supplier, Roundy proved that the optimal integer relaxation objective is a lower bound on the cost of an arbitrary policy (Roundy, 1985). In this subsection we extend the result to the multiple supplier case.

Theorem 3. $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *is a lower bound on the average cost of any policy for Problem* (**P**).

A proof of Theorem 3 is in Appendix A.4. This theorem implies that the Power-of-Two policy finds a solution with a worst case performance ratio of 94% with respect to any feasible policy. Note that this is significantly stronger than previous results in literature that focus on nested and stationary policies, which can be arbitrarily bad.

3.2 Decentralized Zero-Inventory Ordering Policy

In the previous section, we assumed that the system was operated under centralized control, so that shipments from different suppliers to the warehouse, and shipments from the warehouse to retailers could be coordinated to minimize overall system costs. In this section, we consider a decentralized model where suppliers and retailers make their own decisions based on information locally available to them. As discussed in the introduction, we are motivated by current practice, where (at least at the MACC with which we worked), suppliers deliver products to the warehouse, paying transportation costs, unloading costs, and holding costs until goods are shipped to retailers, and retailers order from the warehouse, paying transportation costs.

Specifically, we assume that suppliers pay a fixed transportation cost per shipment as well as holding cost at the warehouse, and must meet retailer demand. Similarly, we assume that retailers must pay a fixed ordering cost for deliveries, as well as the holding cost at their own stores, and must meet customer demand without backorder.

We also assume that retailers first optimize their own strategy, and then suppliers must react to this strategy, in line with what we have observed in practice. In this section, we analyze the optimal retailers' strategy and propose a stationary fixed order interval heuristic for the challenging-to-optimize suppliers' problem, a stationary ZIO policy.

We summarize the new notation for this section below. We use Γ to denote order intervals under decentralized models.

- Γ_j^r (Γ_j^{r*}) : (optimal) order interval for retailer j in the decentralized model
- $\boldsymbol{\Gamma^{r*}} = (\Gamma_1^{r*}, \Gamma_2^{r*}, \cdots, \Gamma_m^{r*})$: vector of optimal order intervals for all retailers in the decentralized model
- Γ_i^s (Γ_i^{s*}) : (optimal) order interval for supplier i in the decentralized model
- $\boldsymbol{\Gamma^{s*}} = (\Gamma_1^{s*}, \Gamma_2^{s*}, \cdots, \Gamma_n^{s*})$: vector of optimal order intervals for all suppliers
- $I_{ij}^w(t)$: the inventory level at time t of product i at the warehouse that is ultimately intended for retailer j
- $I_{ij}(t)$: the inventory at time t of product i at retailer j

- $EI^w_{ij}(t) = I^w_{ij}(t) + I_{ij}(t)$: echelon inventory of product i intended for retailer j (that is, inventory at warehouse intended for retailer j plus the inventory at retailer j)

3.2.1 Retailers' Policy

Note that each retailer will have an optimal ZIO policy (indeed, this can be viewed as an EOQ problem at the retailer). Each order will thus contain products from all suppliers sufficient to cover demand during that cycle. Hence, as in the EOQ, the optimal strategy is a fixed order interval strategy, and the problem for retailer j is:

$$(\mathbf{PDR_j}) \quad \min C^r_j(\Gamma^r_j) = \frac{k^r_j}{\Gamma^r_j} + \sum_{i\in S} \frac{1}{2} d_{ij} h_{ij} \Gamma^r_j = \frac{k^r_j}{\Gamma^r_j} + \sum_{i\in S} (g^{ij} + g_{ij}) \Gamma^r_j.$$

By the first order conditions, we obtain the optimal order interval for retailer j:

$$\Gamma^{r*}_j = \sqrt{\frac{k^r_j}{\sum\limits_{i\in S}(g^{ij} + g_{ij})}}.$$

The optimal decentralized cost per unit time for retailer j is therefore:

$$C^r_j(\Gamma^{r*}_j) = 2\sqrt{k^r_j \Big(\sum_{i\in S} (g^{ij} + g_{ij}) \Big)}.$$

Theorem 4. *The optimal order interval for each retailer is longer in centralized model than in the decentralized model. That is,* $\Gamma^{r*}_j \leq T^{r*}_j$.

Compared with centralized model, it is easily seen that $\Gamma^{r*}_j = \sqrt{\frac{k^r_j}{\sum_{i\in S}(g_{ij}+g^{ij})}} \leq \sqrt{\frac{k^r_j}{\sum_{i\in S} g_{ij} + \sum_{i:j\in L_i\cup E_i} g^{ij}}} \leq T^{r*}_j$. This follows because in the centralized model, the shipping decision is made based on the marginal additional holding cost at the retailer, whereas in the decentralized model the shipping decision accounts for the fact that all of the holding cost is paid by the retailer.

3.2.2 The Zero-Inventory-Ordering Supplier Policy

Each time a supplier makes a delivery to the warehouse, it needs to ensure that there is enough inventory to cover demands from retailers until the next delivery. However, since orders may not line up, some retailers may order during a particular supplier interval to cover demand during the next supplier order interval, so that the demand from a retailer in a particular supplier interval may exceed the demand that the retailer faces during this interval. Thus, the optimal supply policy may be complex and non-stationary, even if the retailer ordering pattern is known and stationary, and we are motivated to consider heuristics for supplier policy.

Recall that although aggregate demand from retailers to a supplier is discrete and likely time variant (since retailers are heterogeneous), it is deterministic. Thus, each time a supplier makes a shipment to the warehouse, it can calculate precisely the amount that will be demanded by retailers until the supplier makes its next shipment. This will result in a ZIO policy at the warehouse.

Specifically, for supplier i, given any order interval Γ_i^s, the replenishment quantity at the start of that interval can be set equal to the total amount of retailer orders for that product during the interval, resulting in a ZIO policy. To calculate the expected cost for each supplier, as before we decompose inventory at the warehouse by supplier and intended retailer, and determine the holding cost associated with each retailer and product.

Inventory Cost at the Warehouse

We first characterize inventory cost at the warehouse for three cases. We introduce more notation in this section for convenience.

- $H_{ij}^e(\Gamma_i^s, \Gamma_j^r)$: the average holding cost per unit time at the warehouse for product i ultimately intended for retailer j if $\Gamma_i^s = \Gamma_j^r$
- $H_{ij}^g(\Gamma_i^s, \Gamma_j^r)$: the average holding cost if $\Gamma_i^s < \Gamma_j^r$
- $H_{ij}^l(\Gamma_i^s, \Gamma_j^r)$ be the average holding cost if $\Gamma_i^s > \Gamma_j^r$
- $\hat{H}_{ij}^g(\Gamma_i^s, \Gamma_j^r)$: an upper bound on $H_{ij}^g(\Gamma_i^s, \Gamma_j^r)$
- $\hat{H}_{ij}^l(\Gamma_i^s, \Gamma_j^r)$: an upper bounds on $H_{ij}^l(\Gamma_i^s, \Gamma_j^r)$
- b_{ij} : integer part of $\frac{\Gamma_i^s}{\Gamma_j^{r*}}$
- a_{ij} : fractional part of $\frac{\Gamma_i^s}{\Gamma_j^{r*}}$, if a_{ij} is rational, we further let $a_{ij} = \frac{p_{ij}}{q_{ij}}$ where integers p_{ij} and q_{ij} are coprime.
- $\tilde{b}_{ij}$, $\tilde{a}_{ij}$: integer part and fractional part of $\frac{\Gamma_j^{r*}}{\Gamma_i^s}$, and we define $\tilde{a}_{ij} = \frac{\tilde{p}_{ij}}{\tilde{q}_{ij}}$ correspondingly if $\tilde{a}_{ij}$ rational.

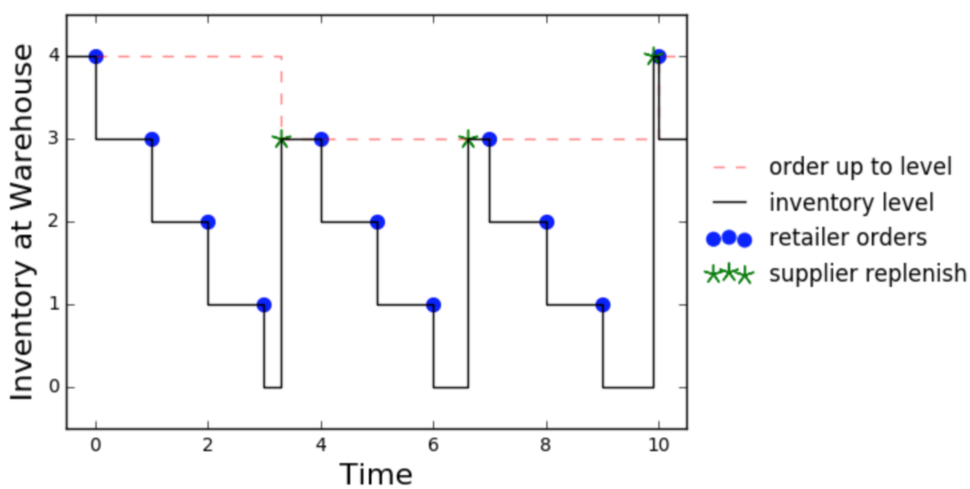
The holding cost associated with I_{ij}^w depends on the relationship between Γ_i^s and Γ_j^{r*}. We consider the three possible cases below:

1. $\Gamma_i^s = \Gamma_j^{r*}$:
   We assume that for each supplier and each retailer, the first order occurs at the start of the horizon. Thus $\Gamma_i^s = \Gamma_j^{r*}$ implies that retailer j and supplier i always order simultaneously, so there is no inventory held at the warehouse for retailer j. That is,

$$H_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) = 0.$$

2. $\Gamma_i^s > \Gamma_j^{r*}$:
   As we see from Figure 3.1, within each order interval, supplier i faces b_{ij} or $b_{ij}+1$ orders from retailer j.

Figure 3.1: $\Gamma_i^s > \Gamma_j^{r*}$, inventory from supplier i to retailer j



This observation allows us to exactly characterize $H_{ij}^l(\Gamma_i^s, \Gamma_j^{r*})$, the holding cost of $I_{ij}^w(t)$ if $\Gamma_i^s < \Gamma_j^{r*}$.

Theorem 5.

$$H_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) = \begin{cases} (\Gamma_i^s - \frac{\Gamma_j^{r*}}{q_{ij}})g^{ij}, & if\ \frac{\Gamma_i^s}{\Gamma_j^{r*}} \in \mathbb{Q} \\ g^{ij}\Gamma_i^s, & otherwise \end{cases}.$$

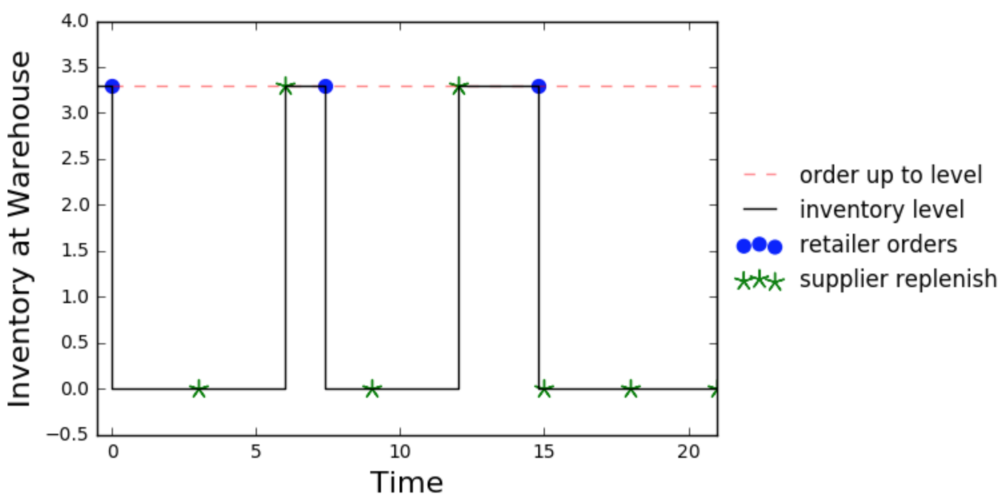
A proof of Theorem 5 is in Appendix A.6. From Theorem 5, it is straightforward to see that

$$\hat{H}ij^l(\Gamma_i^s, \Gamma_j^{r*}) \triangleq \Gamma_i^s g^{ij} \tag{3.10}$$

is an upper bound on $H^l_{ij}(\Gamma^s_i, \Gamma^{r*}_j)$.

3. $\Gamma^s_i < \Gamma^{r*}_j$:
   Supplier i only transfers inventory for retailer j for cycles when retailer j places an order, as illustrated in Figure 3.2.

Figure 3.2: $\Gamma^s_i < \Gamma^{r*}_j$, inventory from supplier i to retailer j



To evaluate $H^g_{ij}(\Gamma^s_i, \Gamma^{r*}_j)$, we first introduce a technical lemma:

Lemma 2. *Let $\Delta_{ij}(k)$ be the time between k^{th} order of retailer j, and supplier i's last replenishment before retailer j's order. If supplier i replenishes inventory simultaneously with k^{th} order from retailer j, we let $\Delta_{ij}(k) = 0$. Then the long run average of $\Delta_{ij}(k)$ is*

$$\frac{1}{N}\sum_{k=1}^{N} \Delta_{ij}(k) \xrightarrow{N\to\infty} \begin{cases} \frac{\Gamma^s_i(\tilde{q}_{ij}-1)}{2\tilde{q}_{ij}} & \text{if } \frac{\Gamma^{r*}_j}{\Gamma_i} \in \mathbb{Q} \\ \frac{\Gamma^s_i}{2}, & \text{otherwise} \end{cases}.$$

Similarly, we let $\overline{\Delta}_{ij}(k)$ be the time between k^{th} order of retailer j, and supplier i's next order after retailer j's order. We assume $\overline{\Delta}_{ij}(k) = \Gamma^s_i$ if supplier i and retailer j replenishes inventory at the same time. Then the long run average of $\overline{\Delta}_{ij}(k)$ is

$$\frac{1}{N}\sum_{k=1}^{N} \overline{\Delta}_{ij}(k) \xrightarrow{N\to\infty} \begin{cases} \frac{\Gamma^s_i(\tilde{q}_{ij}+1)}{2\tilde{q}_{ij}} & \text{if } \frac{\Gamma^{r*}_j}{\Gamma_i} \in \mathbb{Q} \\ \frac{\Gamma^s_i}{2}, & \text{otherwise} \end{cases}.$$

A proof of Lemma 2 is in Appendix A.7. We use this lemma to prove the following results:

Theorem 6. *If $\Gamma_i^s < \Gamma_j^r$, the average holding cost per unit time for the warehouse inventory of product i intended for retailer j is:*

$$H_{ij}^g(\Gamma_i^s, \Gamma_j^{r*}) = \begin{cases} \frac{\tilde{q}_{ij}-1}{\tilde{q}_{ij}} g^{ij}\Gamma_i^s, & if \ \frac{\Gamma_j^{r*}}{\Gamma_i^s} \in \mathbb{Q} \\ g^{ij}\Gamma_i^s, & otherwise \end{cases}. \tag{3.11}$$

A proof of Theorem 6 is in Appendix A.8. A natural lower bound is observed from (3.11),

$$\hat{H}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*}) = g^{ij}\Gamma_i^s.$$

Combining these cases results in an upper bound on the cost faced by supplier i under a ZIO policy. Thus the problem for supplier i is:

$$(\mathbf{PDS_i^{zio}}) \quad \min C_i^{zio}(\Gamma_i^s) = \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i} \hat{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in E_i} H_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in G_i} \hat{H}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*})$$
$$= \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i \cup G_i} g^{ij}\Gamma_i^s.$$

Optimal order intervals for suppliers

To solve $(\mathbf{PDS_i^{zio}})$, we separate the feasible region into two parts, depending on whether or not the supplier i has the same order interval as any retailer. Though $(\mathbf{PDS_i^{zio}})$ is nonconvex, we show the following optimal solution $\Gamma_{i,zio}^{s*}$ can be found in a finite set.

Theorem 7. *The optimal order interval for supplier i in decentralized ZIO policy $\in \{\Gamma_1^{r*}, \Gamma_2^{r*}, \cdots, \Gamma_m^{r*}, \tilde{\Gamma}_i^{s*}\}$.*

A proof of Theorem 7 can be found in Appendix A.9. Hence we can solve $(\mathbf{PDS_i^{zio}})$ by comparing all of $C_i^{zio}(\Gamma_j^{r*})$ and $C_i^{zio}(\tilde{\Gamma}_i^{s*})$, and selecting the interval generating the smallest cost. There are at most m different values for Γ_j^{r*} and $C_i^{zio}(\tilde{\Gamma}_i^{s*})$ can be obtained by (A.13). Thus $(\mathbf{PDS_i^{zio}})$ can be solved efficiently.

3.3 Decentralized Order-up-to Policy

In contrast to the ZIO policy in the previous section, here we analyze the optimal *stationary order-up-to* policy for suppliers. Further, we characterize a bound on the worst-case ratio of the cost of the *order-up-to* policy to the optimal centralized policy of 2.5.

Specifically, we consider a setting in which each supplier utilizes an *order-up-to* policy with a fixed order interval. Since orders from different retailers are independent, in each shipment to the warehouse, supplier i decides quantities to satisfy orders from different retailers separately. The inventory cost is thus naturally decomposed according to which retailer the inventory will be delivered to.

In this policy, each time when supplier i replenishes inventory at warehouse, it raises I_{ij}^w to the same retailer-dependent level. These order-up-to levels are determined by the maximum number of orders each retailer may place in each cycle of supplier i. For example, if $\Gamma_i^s < \Gamma_j^{r*}$, supplier i always raises I_{ij}^w to $\Gamma_j^{r*} d_{ij}$. This is because in each order cycle of supplier i, retailer j will place at most one order. We characterize this stationary *order-up-to* policy below.

3.3.1 Inventory Cost at the Warehouse

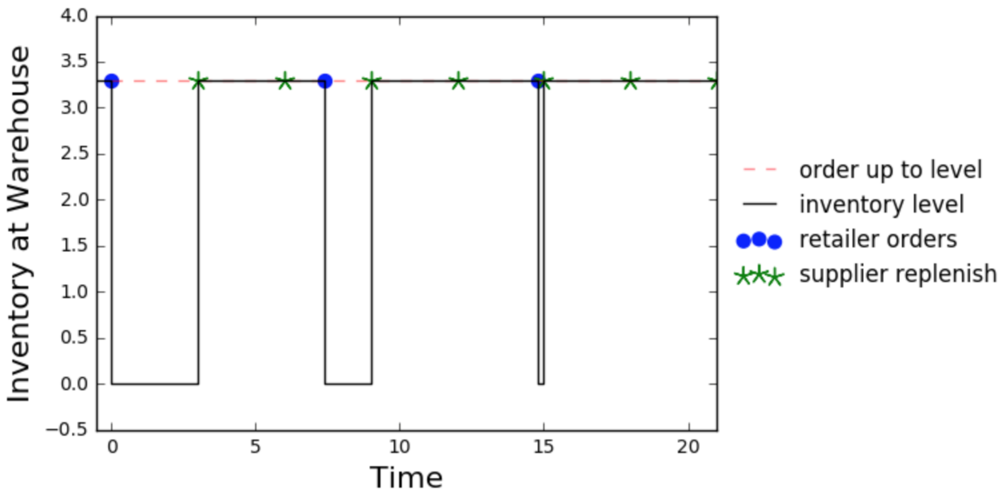
To distinguish notation from analysis in Section 3.2.2, we introduce the following notation for order-up-to policy:

- $\underline{H}_{ij}^e(\Gamma_i^s, \Gamma_j^r)$: the average holding cost per unit time at the warehouse for product i ultimately intended for retailer j if $\Gamma_i^s = \Gamma_j^r$
- $\underline{H}_{ij}^g(\Gamma_i^s, \Gamma_j^r)$: the average holding cost if $\Gamma_i^s < \Gamma_j^r$
- $\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^r)$ be the average holding cost if $\Gamma_i^s > \Gamma_j^r$
- $\hat{\underline{H}}_{ij}^g(\Gamma_i^s, \Gamma_j^r)$: an upper bound on $H_{ij}^g(\Gamma_i^s, \Gamma_j^r)$
- $\hat{\underline{H}}_{ij}^l(\Gamma_i^s, \Gamma_j^r)$: an upper bounds on $H_{ij}^l(\Gamma_i^s, \Gamma_j^r)$

We characterize the holding cost in 3 different scenarios:

1. $\Gamma_i^s = \Gamma_j^{r*}$:
   Similar to Section 3.2.2,
   $$\underline{H}_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) = H_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) = 0.$$
2. $\Gamma_i^s < \Gamma_j^{r*}$:
   In this case, supplier i always raises inventory level for retailer j to $\Gamma_j^{r*} d_{ij}$, which is the quantity retailer j orders each time. Hence, in those order cycles where retailer j does not place any order, the warehouse holds full inventory for retailer j.

Figure 3.3: $\Gamma_i^s < \Gamma_j^{r*}$, inventory from supplier i to retailer j



Theorem 8. *If $\Gamma_i^s < \Gamma_j^r$, the average holding cost per unit time for the inventory of product i intended for retailer j is:*

$$\underline{H}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*}) = \begin{cases} (2\Gamma_j^{r*} - \frac{\tilde{q}_{ij}+1}{\tilde{q}_{ij}}\Gamma_i^s)g^{ij}, & if\ \frac{\Gamma_j^{r*}}{\Gamma_i^s} \in \mathbb{Q} \\ (2\Gamma_j^{r*} - \Gamma_i^s)g^{ij}, & otherwise \end{cases}.$$

A proof of Theorem 8 is in Appendix A.10. Observe from Theorem 8 that $\hat{\underline{H}}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*}) = (2\Gamma_j^{r*} - \Gamma_i^s)g^{ij}$ is an upper bound of $\underline{H}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*})$.

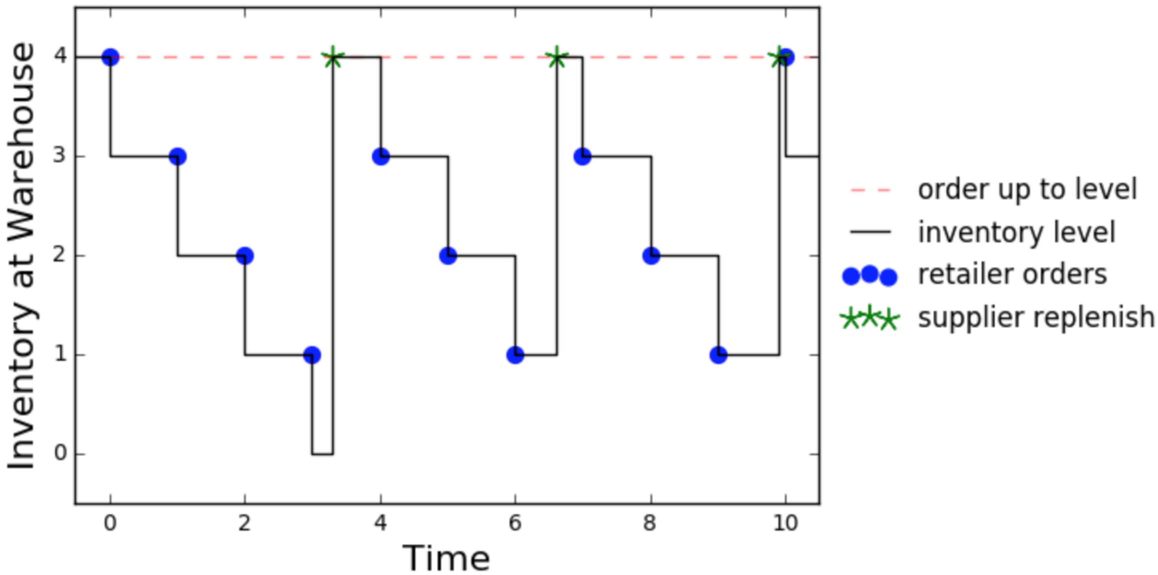
3. $\Gamma_i^s > \Gamma_j^{r*}$:
In this case, although retailer j has a constant order interval, during a replenishment cycle of supplier i, retailer j may place different numbers of orders. Recall that $\frac{\Gamma_i^s}{\Gamma_j^{r*}} = b_{ij} + a_{ij}$, where b_{ij} is the integer part and a_{ij} is the fractional part. Within each supplier order interval, supplier i faces b_{ij} orders from retailer j if $a_{ij} = 0$, otherwise faces either b_{ij} or $b_{ij} + 1$ orders from retailer j. Therefore for supplier i, the minimum order-up-to level that can satisfy all demand is

$$(b_{ij} + \mathbb{1}_{a_{ij}\neq 0})d_{ij}\Gamma_j^{r*}.$$

In appendix, we use the notion of echelon inventory to evaluate $\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*})$. Installation inventory at warehouse can be obtained from the difference of echelon inventory at the warehouse and echelon inventory at retailer j, as the latter two are easier to analyze. We use this notion to show the following result:

Lemma 3. $\hat{\underline{H}}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) = (\Gamma_i^s + 2\Gamma_j^{r*})g^{ij}$ *is an upper bound on* $\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*})$.

Figure 3.4: $\Gamma_i^s > \Gamma_j^{r*}$, inventory from supplier i to retailer j



A proof of Lemma 3 is in Appendix A.11.

Based on these three cases, we have the following upper bound on the the cost faced by supplier i when this *order-up-to* policy is employed:

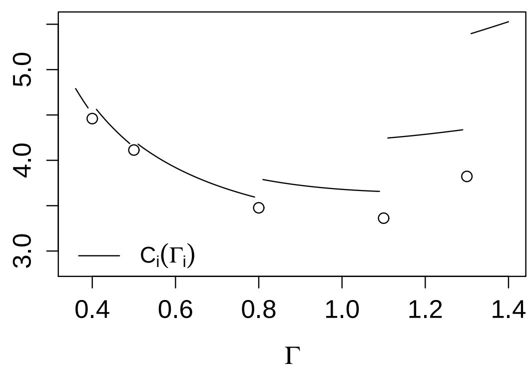
$$(\mathbf{PDS_i^{out}}) \quad \min C_i^{out}(\Gamma_i^s) = \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i} \underline{\hat{H}}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in E_i} \underline{H}_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in G_i} \underline{\hat{H}}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*})$$

$$= \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i} g^{ij}\Gamma_i^s - \sum_{j \in G_i} g^{ij}\Gamma_i^s + 2 \sum_{j \in L_i \cup G_i} g^{ij}\Gamma_j^{r*}.$$

3.3.2 Optimal Order Intervals for Suppliers

For the convenience of analysis, we order retailers such that $\Gamma_1^{r*} \le \Gamma_2^{r*} \le \cdots \le \Gamma_m^{r*}$. $\boldsymbol{\Gamma^{r*}}$ divides $(0, \infty)$ into $\le m+1$ open intervals: $(0, \Gamma_1^{r*}), (\Gamma_1^{r*}, \Gamma_2^{r*}), \cdots, (\Gamma_m^{r*}, \infty)$. In each open interval, $C_i^{out}(\Gamma_i^s)$ is piecewise convex and continuous in Γ_i^s everywhere except points Γ_j^{r*}, as illustrated in Figure 3.5.

Figure 3.5: Decentralized cost for supplier i



To solve ($\mathbf{PDS_i^{out}}$), we find the locally optimal solution, which we denote $\Gamma_{i,k}^{s*}$, in each interval, and then search for the global optimal order interval for supplier i.

In each open interval $(\Gamma_k^{r*}, \Gamma_{k+1}^{r*})$, both $\sum_{j\in L_i\cup G_i} 2g^{ij}\Gamma_j^{r*} = \sum_{j=1}^{k} 2g^{ij}\Gamma_j^{r*}$ and $\sum_{j\in L_i} g^{ij} - \sum_{j\in G_i} g^{ij}$ are constant. Thus finding locally optimal solution to ($\mathbf{PDS_i^{out}}$) in $(\Gamma_k^{r*}, \Gamma_{k+1}^{r*})$ is equivalent to solving

$$(\mathbf{PDS_{i,k}^{out}}) \quad \min C_i^{out}(\Gamma_i^s) = \frac{k_i^s}{\Gamma_i^s} + \left(\sum_{j\in L_i} g^{ij} - \sum_{j\in G_i} g^{ij}\right)\Gamma_i^s$$
$$s.t. \quad \Gamma_i^s \in (\Gamma_k^{r*}, \Gamma_{k+1}^{r*})$$

We consider two cases depending on Γ_i^s:

1. If $\sum_{j\in L_i} g^{ij} - \sum_{j\in G_i} g^{ij} > 0$:
   We take first order condition and obtain:

$$\hat{\Gamma}_{i,k}^{s*} = \sqrt{\frac{k_i^s}{\sum_{j\in L_i} g^{ij} - \sum_{j\in G_i} g^{ij}}}.$$

   Hence the locally optimal order interval is:

$$\Gamma_{i,k}^{s*} = \begin{cases} \hat{\Gamma}_{i,k}^{s*}, & \text{if } \hat{\Gamma}_{i,k}^{s*} \in (\Gamma_k^{r*}, \Gamma_{k+1}^{r*}) \\ \text{does not exist}, & \text{otherwise} \end{cases}.$$

2. If $\sum_{j \in L_i} g^{ij} - \sum_{j \in G_i} g^{ij} \leq 0$:
   $C^{out}_{i,k}(\Gamma^s_i)$ is decreasing on $(\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$, hence the locally optimal solution in the open interval $(\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$ does not exist either.

Next we show that for each open interval $(\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$, if $\Gamma^{s*}_{i,k}$ does not exist, $C^{out}_i(\Gamma^s_i)$ in the interval is lower bounded by the minimum of the two endpoints, $C^{out}_i(\Gamma^{r*}_k)$ and $C^{out}_i(\Gamma^{r*}_{k+1})$. Otherwise, we claim that there exists at most one index k such that $\hat{\Gamma}^{s*}_{i,k} \in (\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$, which we denote k'. Then as we state in the main paper:

Theorem 9. *The optimal solution to minimizing $C^{out}_i(\Gamma^s_i)$ is either one of Γ^{r*}_i or $\hat{\Gamma}^{s*}_{i,k'}$.*

A proof of Theorem 9 is in Appendix A.12.

We proved in Theorem 4 that retailers order less frequently in centralized model compared to decentralized model. Unfortunately, we have not found such results for suppliers. As a result, each supplier need to search all $m+1$ candidates to determine the optimal one.

Alhough this *order-up-to* policy may generate higher inventory cost, it is easy to implement, and has a bounded worst case ratio of 2.5 (Theorem 12).

3.4 Semi-Decentralized Model with PoT Control

In Section 3.2, we proposed two retailer-driven decentralized policies, where both suppliers and retailers have full flexibility in choosing reorder intervals.

In this section, we consider a semi-decentralized model where suppliers and retailers still make their own plans, but under certain restriction. In particular, they are required to use stationary PoT policies with given base cycle T_0. That is, their order interval must be belong to $\{2^m T_0 : m \in \mathcal{N}\}$.

Below, we characterize supplier and retailer policies in this setting, and then bound the ratio of the cost of this semi-decentralized policy to the centralized policy. We see that this bound is *almost as good* as that of the ZIO policy. In Section 3.6, we show that in practice, this policy performs better than both decentralized policies.

We use τ to denote order intervals under this semi-decentralized model with PoT requirements.

- τ^r_j (τ^{r*}_j) : (optimal) order interval for retailer j in semi-decentralized model with PoT restriction
- $\boldsymbol{\tau^{r*}} = (\tau^{r*}_1, \tau^{r*}_2, \cdots, \tau^{r*}_m)$: vector of optimal order intervals for all retailers
- τ^s_i (τ^{s*}_i) : (optimal) order interval for supplier i in semi-decentralized model with PoT restriction

- $\boldsymbol{\tau^{s*}} = (\tau_1^{s*}, \tau_2^{s*}, \cdots, \tau_n^{s*})$: vector of optimal order intervals for all suppliers

3.4.1 Retailers' Policy

Under Power-of-Two restrictions, the subproblem for retailer j is:

$$(\mathbf{PDPR_j}) \quad \min C_{j,p}^r(\tau_j^r) = \frac{k_j^r}{\tau_j^r} + \sum_{i \in S} \frac{1}{2} d_{ij} h_{ij} \tau_j^r = \frac{k_j^r}{\tau_j^r} + \sum_{i \in S} (g^{ij} + g_{ij}) \tau_j^r$$
$$s.t. \quad \tau_j^r \in \{2^m T_0 : m \in \mathcal{N}\}$$

For $(\mathbf{PDPR_j})$, we first solve the relaxed problem:

$$\Gamma_j^{r*} = \sqrt{\frac{k_j^r}{\sum_{i \in S} (g^{ij} + g_{ij})}}.$$

Then apply Power-of-Two rounding:

$$\tau_j^{r*} = \min \left\{ 2^m T_0 : 2^m T_0 \geq \frac{\Gamma_j^{r*}}{\sqrt{2}} \right\}.$$

Theorem 10. $C_{j,p}^r(\tau_j^{r*}) \leq \frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2}) C_j^r(\Gamma_j^{r*})$

This follows from standard Power-of-Two rounding analysis. Under the Power-of-Two policy, the cost for each retailer is no more than 1.06 $C_j^r(\Gamma_j^{r*})$, the decentralized (full flexibility) optimal cost. That is, the PoT restriction increases retailer costs by at most 6%.

3.4.2 Suppliers' Policy

Under the Power-of-Two policy requirement, orders from retailers are more likely to "line up." To characterize the holding cost at the warehouse, we consider two cases:

1. $\tau_i^s > \tau_j^{r*}$:
   Since $\tau_i^s,\ \tau_j^{r*} \in \{2^m T_0 : m \in \mathcal{N}\}$, τ_i^s is a multiple of τ_j^{r*}. Hence the average inventory cost of $I_{ij}^w(t)$ is:

$$(\tau_i^s - \tau_j^{r*}) \frac{1}{2} d_{ij} h^i = (\tau_i^s - \tau_j^{r*}) g^{ij}.$$

2. $\tau_i^s \leq \tau_j^{r*}$:
   In this case, supplier i places an order each time retailer j orders, thus supplier i does not need to store inventory at the warehouse for retailer j. That means $I_{ij}^w(t) = 0$, and thus the corresponding inventory cost is also 0.

Thus, the average inventory cost associated with $I^w_{ij}(t)$ is

$$(\max(\tau^s_i, \tau^{r*}_j) - \tau^{r*}_j) \cdot g^{ij}. \tag{3.12}$$

The problem faced by the supplier is therefore:

$$(\mathbf{PDPS_i}) \quad \min C^s_{i,p}(\tau^s_i) = \frac{k^s_i}{\tau^s_i} + \sum_{j \in L_i} g^{ij}(\tau^s_i - \tau^{r*}_j)$$
$$s.t. \quad \tau^s_i \in \{2^m T_0 : m \in \mathcal{N}\}$$

Optimal Supplier Order Interval

We consider the following relaxation problem of ($\mathbf{PDPS_i}$) where PoT constraints are relaxed:

$$(\mathbf{PDPSR_i}) \quad \min C^s_{i,P}(\tau^s_i) = \frac{k^s_i}{\tau^s_i} + \sum_{j \in L_i} g^{ij}(\tau^s_i - \tau^{r*}_j)$$
$$s.t. \quad \tau^s_i > 0$$

$C^s_{i,P}(\tau^s_i)$ is continuous in τ^s_i and its gradient is:

$$\frac{dC^s_{i,P}(\tau^s_i)}{d\tau^s_i} = \frac{-k^s_i}{(\tau^s_i)^2} + \sum_{j:\tau \geq \tau^{r*}_j} g^{ij}. \tag{3.13}$$

(3.13) is increasing, which implies $C^s_{i,P}(\tau^s_i)$ is convex. Therefore, $\exists \hat{\tau}^s_i$ such that the gradient vanishes at $\hat{\tau}^s_i$, so $\hat{\tau}^s_i$ is the optimal solution to ($\mathbf{PDPSR_i}$).

Next we round $\hat{\tau}^s_i$ to the two nearest PoT solutions $\underline{\tau}^s_i$ and $\overline{\tau}^s_i$. One of these must be the optimal solution to ($\mathbf{PDPR_i}$). That is,

$$\tau^{s*}_i = argmin\{C_{i,P}(\underline{\tau}^s_i),\ C_{i,P}(\overline{\tau}^s_i)\}, \tag{3.14}$$

where

$$\underline{\tau}^s_i = \max_{\tau} \left\{ \frac{k^s_i}{\tau^2} \geq \sum_{j:\tau \geq \tau^{r*}_j} g^{ij},\ \tau \in \{2^m T_0 : m \in \mathcal{N}\} \right\}$$
$$\overline{\tau}^s_i = \min_{\tau} \left\{ \frac{k^s_i}{\tau^2} \leq \sum_{j:\tau \geq \tau^{r*}_j} g^{ij},\ \tau \in \{2^m T_0 : m \in \mathcal{N}\} \right\}.$$

3.5 Cost of Decentralization

In this section, we consider the worst case performance (with respect to optimal centralized performance) of the two decentralized policies proposed in Section 3.2, Section 3.3 and a semi-decentralized policy proposed in Section 3.4 using the bound $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ developed in Section 3.1.6 as a benchmark – in effect, we bound the cost increase due to decentralization. The worst case performance is bounded in both *zero-inventory-ordering* and *order-up-to* supplier policies.

3.5.1 The Cost of Decentralization Using the Decentralized ZIO Policy

In our decentralized model, the total cost suppliers and retailers pay in the optimal solution is:

$$C_{zio}(\boldsymbol{\Gamma^{s*}}, \boldsymbol{\Gamma^{r*}}) = \sum_{i \in S} C_i^{zio}(\Gamma_i^{s*}) + \sum_{j \in R} C_j^r(\Gamma_j^{r*}). \tag{3.15}$$

In this subsection, we analyze how well our decentralized model performs relative to the optimal centralized policy. In Section 3.1.6, we saw that $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on optimal centralized cost. Therefore, the performance of the decentralized ZIO policy (and thus the optimal decentralized policy) is bounded by:

$$\frac{C_{zio}(\boldsymbol{\Gamma^{s*}_{zio}}, \boldsymbol{\Gamma^{r*}})}{\text{Optimal Centralized Cost}} \leq \frac{C_{zio}(\boldsymbol{\Gamma^{s*}_{zio}}, \boldsymbol{\Gamma^{r*}})}{C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})}.$$

Building on this development, we bound the ratio of total decentralized cost under the ZIO policy to optimal centralized cost:

Theorem 11. $C_{zio}(\boldsymbol{\Gamma^{s*}_{zio}}, \boldsymbol{\Gamma^{r*}}) \leq \frac{3}{2} C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. *That is, the total cost of the ZIO decentralized system is no more than* $\frac{3}{2}$ *times the optimal cost in the centralized system.*

A proof of Theorem 11 is provided in Appendix A.13.

3.5.2 The Cost of Decentralization Using the Decentralized OUT Policy

Similarly, we use $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ to bound the performance of the decentralized OUT policy analyzed in Section 3.3.

Theorem 12. $C_{zio}(\boldsymbol{\Gamma^{s*}_{out}}, \boldsymbol{\Gamma^{r*}}) \leq \frac{5}{2} C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. *That is, the total cost of the OUT decentralized system is no more than* $\frac{5}{2}$ *times the optimal cost in the centralized system.*

A proof of Theorem 12 is in Appendix A.14.

3.5.3 The Cost of Semi-Decentralization Using POT Policy

In this subsection, we show that the ratio of the cost of this semi-decentralized model to the optimal centralized cost is bounded. First, we characterize a bound on the ratio of $C^s_{i,p}(\tau^{s*}_i)$ to $C'_i(\tilde{\Gamma}^{s*}_i)$:

Theorem 13. *Let* $C^{s*}_{i,p} = C^s_{i,P}(\tau^{s*}_i)$ *be the optimal cost for Problem* ($\mathbf{PDPS_i}$)*, and recall that* $C'_i(\tilde{\Gamma}^{s*}_i) = \min_{\Gamma^s_i} \left(\frac{k^s_i}{\Gamma^s_i} + \sum_{j \in R} g^{ij} \Gamma^s_i \right)$*. We have* $C^{s*}_{i,p} \leq (\sqrt{2} + \frac{\sqrt{2}}{2}) \cdot C'_i(\tilde{\Gamma}^{s*}_i)$*.*

A proof of Theorem 13 is in Appendix A.15. We use this result to prove that:

Theorem 14. *The total cost of semi-decentralized system is no more than* $\frac{3}{4}(\sqrt{2} + \frac{\sqrt{2}}{2}) C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$*.*

A proof of Theorem 14 is in Appendix A.16. Note that this bound is slightly worse than the bound for the ZIO policy, although in computational testing, this algorithm generally performs better.

3.6 Computational Study

We use a computational study to better understand the relative performance of our centralized and decentralized polices in various settings. In particular, we explore the following questions:

- How does changing the number of suppliers relative to number of retailers affect the system performance?
- How do variations in costs affect performance of different policies?
- What is the impact of reducing warehouse cost?
- How do correlated costs affect centralization and decentralization?

Before we dig into all these questions, we first compare the efficiency of our MSIRR Algorithm and SCOP approach in Section 3.6.1.

All computational study is run on a MacBook Pro with 2.7 GHz Intel Core i5 processor and 8 GB 1867 MHz DDR3 memory.

3.6.1 MSIRR vs SCOP

In this subsection, we focus more on scale of the problem and speed of the approaches, so we randomly generate all parameters h'_{ij}, h^i, d_{ij}, k^s_i, $k^r_j \sim Unif(1,2)$. We implemented MSIRR in R 3.0.3 and use CPLEX 12.6 to solve SCOP. All results are summarized in Table 3.1.

Table 3.1: Comparison of MSIRR and CPLEX

		MSIRR		**CPLEX**	
n	m	Time (s)	Cost	Time (s)	Cost
50	50	0.479	1329.69	5.83	1329.69
100	100	3.58	3595.23	22.66	3595.23
200	200	24.12	10903.95	336.39	10903.95
500	500	396.18	43028.00	—	—
1000	1000	3612.00	118833.15	—	—

We compare the two methods in following three aspects:

- Optimality:
  From Table 3.1 we can see that, on relatively small scale problems, both MSIRR and CPLEX obtains the optimal solution (the optimality of MSIRR is proved in Lemma 1).

- Efficiency:
  MSIRR runs about 10 times faster than CPLEX on those problems solved by both.

- Problem Scale:
  When we transform (**PIR**) to (**PSCOP**), variable number increases from $n+m$ to $n \cdot m$ due to linearization. For our case with 500/1000 suppliers and retailers, CPLEX cannot solve the corresponding SCOP on the platform. But our MSIRR algorithm is able to handle large scale problem with complexity of $O(mn(m+n))$.

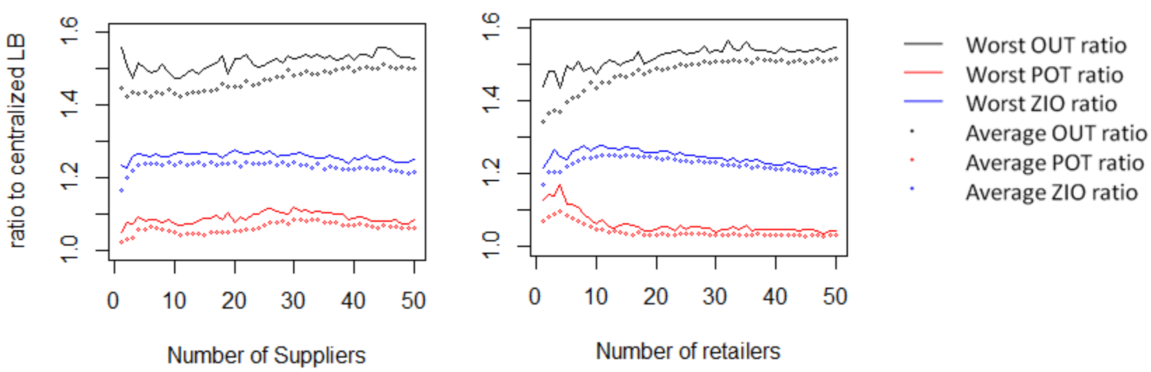
Now we show MSIRR is both faster and can handle problems of larger scales, so throughout following computational study, we apply our MSIRR Algorithm to compute $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ as a benchmark for optimal centralized cost. Then we search for the optimal ZIO, *order-up-to*, and Power-of-Two policies based on Theorem 7, Theorem 9, and (3.14). Given these, we calculate ratios of (semi-) decentralized costs to centralized cost. All parameters are randomly generated, and for each set of parameter generating distributions, we generate 20 cases and record the average cost ratio as well as the largest (worst) cost ratio. We also calculate coefficient of variation (ratio of standard deviation to mean) for each set of experiments, but it was very small (for most cases $< 5\%$), so for clarity of presentation we omitted it. We ran over $100,000$ cases with different parameter settings, and we summarize our results below.

3.6.2 Number of Suppliers/Retailers

In this subsection, we analyze cost ratios for different policies while varying the number of suppliers and retailers. All suppliers and retailers are assumed to be identical.

We tested a variety of distributions, and observed similar results given similarly scaled parameters, so we report our results for parameters generated using uniform distributions for illustration: $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,1)$, $d_{ij} \sim Unif(1,2)$, $k_i^s \sim Unif(1,2)$, $k_j^r \sim Unif(1,2)$. We initially let $m = 8, n \in \{1, 2, \cdots, 50\}$, and explored how the number of suppliers affects cost ratio. We generated 20 sets of parameters for each m, and calculated the worst case ratio and average ratio for the three decentralized policies. Next, we let $n = 8, m \in \{1, 2, \cdots, 50\}$ and reran the experiments.

Figure 3.6: Decentralized to centralized ratio with differing supplier number n / retailer number m



As can be observed in Figure 3.6, as the number of retailers or suppliers increases, the *order-up-to* policy becomes increasingly bad, while ZIO and PoT policy performs better. In general, when all costs are of the same order of magnitude, the PoT policy performs best among the three decentralized policies, in both average and worst case performance. This is because in the PoT policy, suppliers and retailers are better coordinated by "lining up" their orders.

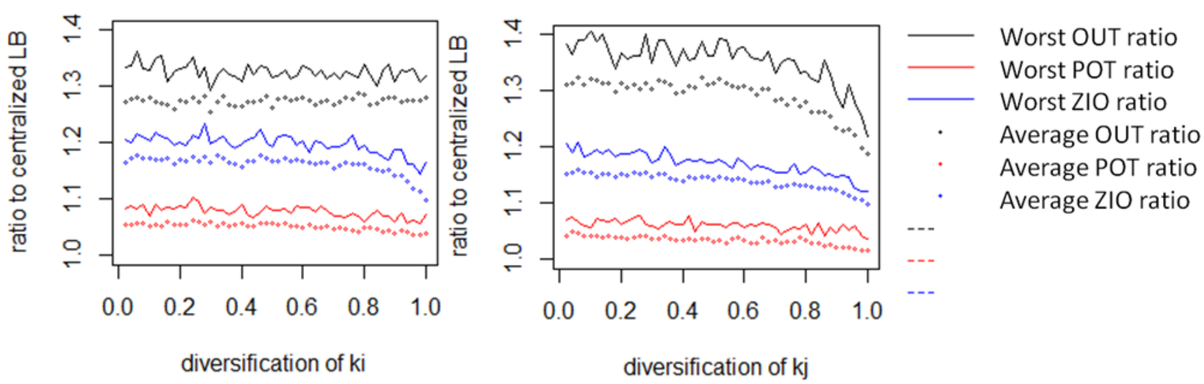
3.6.3 Cost Variation

In this subsection, we explore how diversity in costs affects the performance of the different policies.

Diversity in Fixed Cost

We use the same strategy to generate costs: $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,1)$, $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8$. To analyze the impact of diversity in supplier fixed cost, we keep $k_j^r \sim Unif(1,2)$ and differentiate suppliers by letting $k_i^s \sim \begin{cases} Unif(0, 2-\frac{2k}{51}), & \text{if } i \leq 4 \\ Unif(\frac{2k}{51}, 2), & \text{if } i > 4 \end{cases}$, where $k \in \{1, 2, \cdots, 50\}$. We generate k_i in this way to guarantee that changes in cost ratio purely comes from diversity of k_i^s, but not the scale of k_i^s. Similarly, we run experiments with $k_j^r \sim \begin{cases} Unif(0, 2-\frac{2k}{51}), & \text{if } j \leq 4 \\ Unif(\frac{2k}{51}, 2), & \text{if } j > 4 \end{cases}$, where $k \in \{1, 2, \cdots, 50\}$.

Figure 3.7: Decentralized to centralized ratio with diversity of supplier/retailer fixed costs

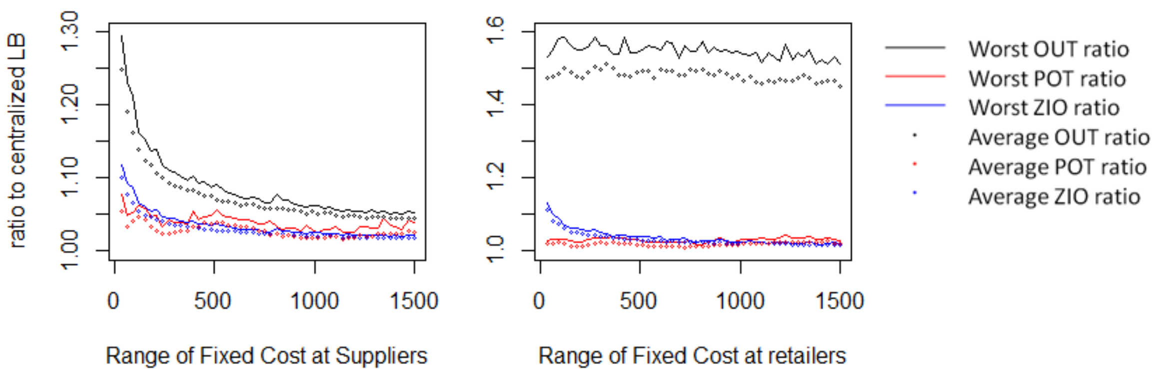


We observe in Figure 3.7 that if costs are of the same order of magnitude, decentralization is most effective when k_j^r are extremely diverse.

Scaling in Fixed Cost

We again use a similar strategy to generate cost parameters: $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,1)$, $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8$. We keep $k_j^r \sim Unif(1,2)$ and differentiate suppliers by letting $k_i^s \sim \begin{cases} Unif(1, k), & \text{if } i \leq 4 \\ Unif(1, 2), & \text{if } i > 4 \end{cases}$, where $k \in \{30, 60, \cdots, 1500\}$. We generate 20 sets of parameters for each k, apply different policies and record the average and worst case cost ratios. Similarly, we run experiments with $k_i^s \sim Unif(1,2)$ and retailers' fixed cost $k_j^r \sim \begin{cases} Unif(1,k), & \text{if } j \leq 4 \\ Unif(1,2), & \text{if } j > 4 \end{cases}$, on same set of k.

Figure 3.8: Decentralized to centralized ratio with diversity of supplier/retailer fixed cost scaling

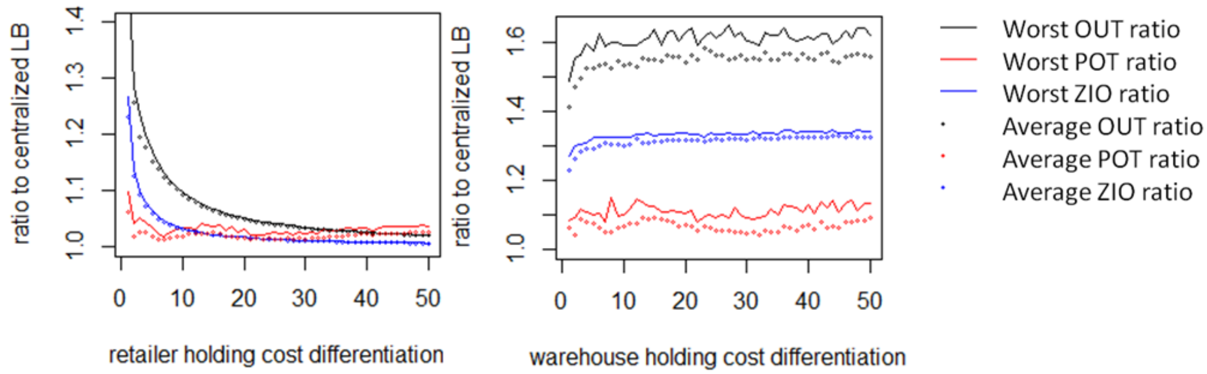


Observe in Figure 3.8 that the semi-decentralized PoT policy is relatively stable under fixed cost variations. The decentralized ZIO policy works well when there is very large differentiation, whether among suppliers or retailers. The *order-up-to* policy also performs well with differentiated supplier fixed costs, but not with diverse retailer fixed costs. In general, this is true because more diverse suppliers or retailers are, the less benefits result from centralization. On the other hand, since our decentralized policies are all retailer-driven, diversification of suppliers does not affect decentralized policies significantly. However, suppliers might pay higher holding cost under the *order-up-to* policy when retailers are more diverse.

Variation in Holding Cost

We next explore how holding cost variations affect system performance. As before, we generate costs as follows: $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8, k_i^s \sim Unif(1,2)$, $k_j^r \sim Unif(1,2)$. To explore the effect of holding cost at the warehouse, we let $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,k)$ where $k \in \{1,2,\cdots,50\}$. Similarly, we let $h'_{ij} \sim Unif(0,k)$, $h^i \sim Unif(0,1)$ to analyze effect of echelon holding costs at retailers.

Figure 3.9: Decentralized to centralized ratio with holding cost variation at the warehouse/retailers



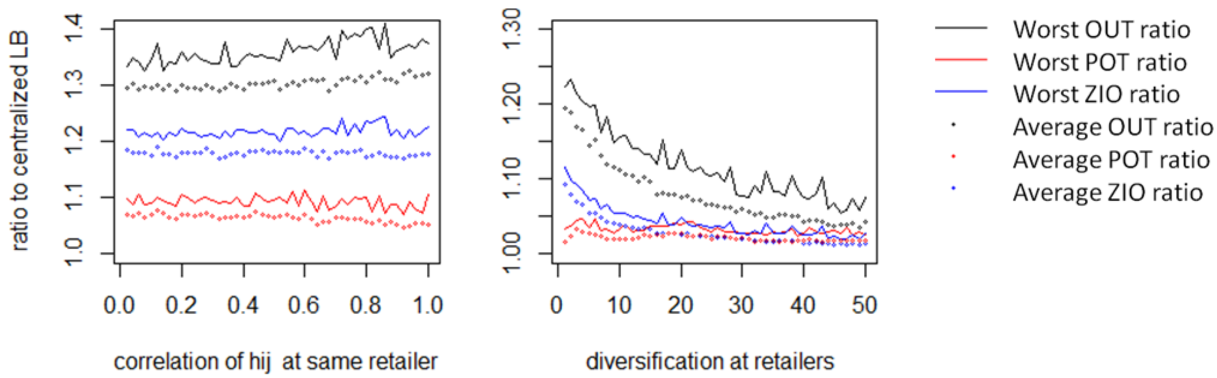
As illustrated in Figure 3.9, differentiation of echelon holding cost at warehouse does not affect cost ratios significantly. This is because echelon holding cost for the same product is the same across retailers, so we still benefit from centralization even if echelon holding costs at the warehouse are diversified. However, if holding costs at retailers are differentiated, retailers and suppliers are less likely to group in the centralized model, and therefore the difference between decentralized and centralized models decreases.

Diversified Retailers with Correlated Costs

In the previous experiments, we generated costs at different locations for different products from independent distributions. In reality, holding costs for different products at the same location are typically correlated. This is particularly relevant for the grocery market, where holding costs depend on the location of the retailer. To generate correlated holding cost, we first generate a base holding cost associated with each retailer as $\hat{h_j} \sim Unif(0,1)$, and i.i.d. $\hat{h}'_{ij} \sim Unif(0,1)$. Then we let $h'_{ij} = \rho\hat{h_j} + (1-\rho)\hat{h}'_{ij}$. It is easy to check that $Cor(h'_{i_1 j}, h'_{i_2 j}) = \rho$.

Observe in Figure 3.10 that correlation of holding costs for products at the same retailer has little effect on cost ratios, as long as holding costs at different retailers are of same order of magnitude. Next, we fix $\rho = 0.8$ and explore how diversity among retailers affects cost ratios. We let $\hat{h_j} \sim Unif(0,k)$ where $k \in \{1,2,\cdots,50\}$, and let $h'_{ij} = 0.8\hat{h_j} + 0.2\hat{h}'_{ij}$ with $\hat{h}'_{ij} \sim Unif(0,1)$.

Figure 3.10: Decentralized to centralized ratio with correlated holding cost at diversified retailers



The right half of Figure 3.10 also presents cost ratios with this differentiation among retailers. The more differentiated holding costs among retailers are, the less we lose due to decentralization.

3.6.4 Summary

Based on all of our experiments, centralization does not lead to dramatic benefits. Indeed, our decentralized policies, especially the semi-decentralized and ZIO policies work well, especially when holding costs at different retailers are well differentiated, when holding costs at warehouse are relatively lower than at the retailers, or when fixed costs at suppliers are diverse.

Chapter 4

Per-truck Transportation Cost

In this chapter, we consider a generalized cost structure with per truck transportation cost instead of fixed ordering cost for each retailer. This cost structure is also motivated by the actual policy of ES3: they charge retailers per truck transportation cost whenever retailers order at ES3's warehouse. Because suppliers that serve a large number of retailers typically face large aggregate demand, they often ship by TL. Hence in this paper, we do not worry about transportation cost for suppliers.

In Chapter 3, we built models for simple settings that we used to analyze consolidation between multiple suppliers and retailers in a both centralized and decentralized settings. In this chapter, we extend the cost structure to incorporate truck transportation. Similar to the way that ES3 operatess their warehouse centers, each retailer pays a per truck transportation cost, which is independent of how much of the truck's capacity is utilized. We modify our previous models to incorporate transportation cost, and consider centralized control of orders, so that suppliers can benefit from combining deliveries ultimately intended for different retailers in order to save on transportation, and retailers can save by ordering products from different suppliers simultaneously. Later we show that each retailer will order only one truck each time because they pay per truck transportation cost.

In the following, we first generalize our previous results to the new setting with transportation cost. We then argue that the problem is equivalent to a capacitated One Warehouse Multi-retailer Multi-supplier problem, since each retailer only orders one truck in each order. We provide a modified MSIRR algorithm to solve the integer relaxation problem, and prove it to be a lower bound on the cost of an arbitrary policy when per truck transportation cost is considered. We then implement the order intervals obtained in relaxation problem to both a feasible ZIO policy and a PoT policy, with worst case cost ratio of $\frac{3}{2}$ and 2 respectively.

4.1 The Centralized Lower Bound

4.1.1 The Model

We consider a similar setting to that in Chapter 3, where the warehouse serves as both the outbound warehouse for suppliers and the inbound warehouse for retailers. Each supplier manufactures a unique product and supplies all the retailers. At each retailer, customer demand for each product occurs at constant rate. Demand must be met without backlogging. A fixed ordering cost is incurred whenever a supplier replenishes its inventory at the warehouse. As for transportation cost, suppliers usually have a large amount to ship, thus their transportation cost to the warehouse is linear and can be omitted. However for retailers, based on observation from reality, their cost is incurred by each truckload they use. The per truck cost is assumed independent of the truckload. Linear holding costs are charged both at the warehouse and at retailers for inventory. The holding costs, fixed ordering and transportation costs, and demand rates are constant over time and different at each facility. We follow our previous notation and introduce more for the transportation cost:

- c_j: per truck shipping cost for delivery from the warehouse to retailer j
- q: truck capacity

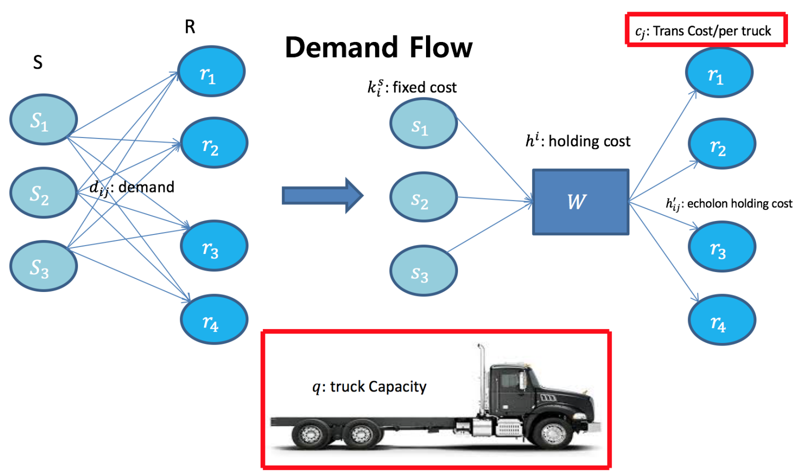


Figure 4.1: Collaboration with Truck Transportation Cost

Our goal is find a policy with minimum long-run average cost while satisfying all the demand, and we denote this Problem **(PT)**. Unfortunately, this problem is $NP-hard$ even in special case of only one supplier and unlimited capacity Arkin et al. (1989).

One degenerate case is discussed in Chapter 3, where truck capacity is set to infinity. The optimal policy to this problem can be extremely complicated, with non-stationary order quantities and order intervals. Thus we focus on quantifying the effectiveness of heuristics.

Among all the feasible policies, we first consider an easy to implement policy, the integer ratio policy. Thus the problem can be formulated as:

$$
\begin{aligned}
(\mathbf{PTI})\ \min \quad & C_{TI}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i \in S} \frac{k_i}{T_i^s} + \sum_{j \in R} \frac{c_j}{T_j^r} \lceil \frac{\sum_{i \in S} d_{ij} T_j^r}{q} \rceil + \sum_{i \in S, j \in R} \max{(T_i^s, T_j^r)} g^{ij} \\
& \qquad + \sum_{i \in S, j \in R} T_j^r g_{ij} \\
s.t. \quad & T_i^s, T_j^r > 0, \forall i \in S, j \in R \\
& T_i^s / T_j^r \in \mathbb{W}, \forall i \in S, j \in R \qquad (4.1)
\end{aligned}
$$

where $\lceil x \rceil$ is the smallest integer $\geq x$.

This cost function $C_{TI}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is generalized from the function Roundy developed for OWMR and we developed in Chapter 3 without transportation cost. When capacity $q \to \infty$, (**PTI**) degenerates to OWMRMS without transportation. In Chapter 3, we showed that the optimal cost of OWMRMS relaxed problem is a lower bound on the long run cost of any policy. In this chapter, we first extend these results to our problem with multiple suppliers and a per truck transportation cost structure. That is, each retailer pays a fixed cost per truck, rather than per delivery (where a delivery could be more than one truck). We show $C_{TI}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is still exact for integer ratio policies, and the optimal solution to a version of the problem where the integer ratio constrinats are relaxed, Problem (**PTR**), is also a lower bound for any arbitrary policy for Problem (**P**). Later in Section 4.5 and Section 4.6.2, we use this lower bound to characterize the cost of decentralized operation in this system.

In the cost function $C_{TI}(\boldsymbol{T^s}, \boldsymbol{T^r})$, the first two terms are the fixed costs for suppliers and transportation costs for retailers, while the latter two terms are the inventory holding costs at the warehouse and retailers. In particular, $\lceil \frac{\sum_{i \in S} d_{ij} T_j^r}{q} \rceil$ is the number of trucks retailer j requires in each replenishment. As for holding cost, we consider two cases following analysis in Chapter 3. If retailer j orders no more frequently than the warehouse does for product from supplier i, then under an integer ratio policy, we only need to consider inventory holding cost incurred at retailers. If the retailer orders less frequently than warehouse does for product from supplier i, the echelon holding cost at the warehouse follows a "sawtooth" pattern with order interval of T_i^s, and inventory at the retailer has interval T_j^r.

Thus, $C_{TI}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is the exact total cost for the centralized OWMRMS with per truck transportation cost under an integer ratio policy, so (**PTI**) exactly models (**PT**) under an integer ratio policy. Later we prove in Section 4.1.4 that the optimal solution to a relaxed version of (**PTI**), which we denote (**PTR**), is a lower bound on any feasible policy for Problem (**PT**).

4.1.2 Equivalent Capacitated OWMRMS

In this section, we transform the relaxation problem into an equivalent capacitated problem, and solve it by modified MSIRR algorithm proposed later. We first relax the integer ratio constraints in Problem **(PTI')** to get

$$
\begin{aligned}
(\mathbf{PTR})\ \min C_{TIR}(\boldsymbol{T^s},\boldsymbol{T^r}) &= \sum_{i\in S}\frac{k_i}{T_i^s}+\sum_{j\in R}\frac{c_j}{T_j^r}\lceil\frac{\sum_{i\in S}d_{ij}T_j^r}{q}\rceil+\sum_{i\in S,j\in R}\max{(T_i^s,T_j^r)}\cdot g^{ij}\\
&+\sum_{i\in S,j\in R}T_j^r\cdot g_{ij}\\
s.t\ T_i^s,T_j^r > 0,\forall i\in S, j\in R
\end{aligned}
$$

The objective function of this problem contains a ceiling function, making the problem challenging to analyze. However, the following lemma enables us to simplify the formulation:

Lemma 4. *In the optimal solution to* **(PTR)***, exactly one truck will be sent to each retailer at a time.*

A proof of Lemma 4 is in Appendix A.17.It follows from Lemma 4 that Problem (**PIR**) is equivalent to the following relaxation of a capacitated OWMRMS problem:

$$
\begin{aligned}
(\mathbf{PCR})\ \min \mathbf{C_C}(\mathbf{T^s},\mathbf{T^r}) = \sum_{i\in S}\frac{k_i}{T_i^s}+\sum_{j\in R}\frac{c_j}{T_j^r}&+\sum_{i\in S,j\in R}\max{(T_i^s,T_j^r)}\cdot g^{ij}+\sum_{i\in S,j\in R}T_j^r\cdot g_{ij}\\
s.t\quad & T_i^s,T_j^r>0,\forall i\in S, j\in R\\
& T_j^r\le T_j^{r0}=\frac{q}{\sum_{i\in S}d_{ij}}
\end{aligned}
$$

The objective of (**PCR**) is the maximum of several convex functions, thus it is still convex. Given any solution ($\boldsymbol{T^s},\boldsymbol{T^r}$) to (**PCR**), following our approach in Chapter 3, we partition suppliers and retailers based on their order intervals, and group retailers and supplier with the same order interval together. We denote the partition $P(U_1)$, $P(U_2)$, $\cdots$, $P(U_k)$, where U_l is the order interval. That is,

$$P(U_l)=\{i\in S: T_i^s=U_l\}\cup\{j\in R: T_j^r=U_l\}.$$

Without loss of generality, we order U_l such that $U_1 < U_2 \cdots < U_k$. Therefore the corresponding optimal cost can be decomposed as follows:

$$\begin{aligned}
C_{TIR}(\boldsymbol{T_i^s}, \boldsymbol{T_j^r}) &= \sum_{i \in S} \frac{k_i}{T_i^s} + \sum_{j \in R} \frac{c_j}{T_j^r} + \sum_{i \in S, j \in R} \max(T_i^s, T_j^r) \cdot g^{ij} + \sum_{i \in S, j \in R} T_j^r \cdot g_{ij} \\
&= \sum_{i \in S} \frac{k_i}{T_i^{s*}} + \sum_{j \in R} \frac{c_j}{T_j^r} + \sum_{\substack{i \in S, j \in L_i \\ or \ i \in S, j \in E_i}} T_i^s \cdot g^{ij} + \sum_{j \in R, i: j \in G_i} T_j^r \cdot g^{ij} + \sum_{i \in S, j \in R} T_j^r \cdot g_{ij} \\
&= \sum_{l \in \{1, \cdots, k\}} \Big(\frac{K(U_l)}{U_l} + H(U_l) \cdot U_l \Big) \\
&\triangleq \sum_{l \in \{1, \cdots, k\}} C_{TIRS}^{U_l}(\boldsymbol{T^r}, \boldsymbol{T^s})
\end{aligned}$$

where

$$K(U_l) = \sum_{i \in P(U_l)} k_i + \sum_{j \in P(U_l)} c_j$$

and

$$H(U_l) = \sum_{\substack{i \in P(U_l) \\ j \in L_i}} g^{ij} + \sum_{\substack{i \in P(U_l) \\ j \in P(U_l)}} g^{ij} + \sum_{\substack{j \in P(U_l) \\ i: j \in G_i}} g^{ij} + \sum_{\substack{i \in S \\ j \in P(U_l)}} g_{ij}$$

$K(U_l)$ and $H(U_l)$ can be viewed as aggregate transportation cost and holding cost for $P(U_l)$.

Given these definitions, (**PTR**) can be decomposed into a series of convex subproblems, one for each partition $P(U_l)$:

$$\begin{aligned}
(\mathbf{PC_l}) \ \min \ & C_{TIRS}^{U_l}(\boldsymbol{T^r}, \boldsymbol{T^s}) = \frac{K(U_l)}{U_l} + H(U_l) \cdot U_l \\
s.t. \ & U_l \geq 0 \\
& U_l \leq T_j^{r_0}, \forall l, j \in P(U_l)
\end{aligned}$$

Lemma 5. *The optimal solution to decomposed problem* ($\mathbf{PC_l}$) *is*

$$U_l^* = \min_{j \in G(U_l)} \left(\sqrt{\frac{K(U_l)}{H(U_l)}}, T_j^{r0} \right).$$

A proof of Lemma 5 is in Appendex A.18.

Thus, given any partition of retailers and suppliers, we can calculate aggregate fixed costs and holding cost in each group, and thus find optimal order intervals. Therefore our problem (**PCR**) is equivalent to finding the optimal partition of $R \cup S$. The conditions for optimality are specified in the following theorem:

Theorem 15. *The following conditions are necessary and sufficient for* $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *to be the optimal order intervals in* **(PCR)**:
$(C1)$ *For the corresponding ordered partition* $P(U_1^*), P(U_2^*), \cdots, P(U_k^*)$ *of* $R \cup S$, *where* $U_1^* < U_2^* < \cdots < U_k^*$, *we have* $U_l^* = \min_{j \in G(U_l)} \left(\sqrt{\frac{K(U_l)}{H(U_l)}}, T_j^{r0} \right)$, *where* $K(U_l)$ *and* $H(U_l)$ *are the aggregate fixed and holding cost of* $P(U_l)$.
$(C2)$ $\forall l, \forall$ *subset* $P \subset P(U_l^*)$,

$$U_l^* \geq \min_{j \in P} \left\{ T_j^{r0}, \sqrt{\frac{\sum_{i \in P} k_i + \sum_{j \in P} c_j}{\sum_{i \in S, j \in P} g_{ij} + \sum_{i \in P, j \in L_i} g^{ij} + \sum_{j \in P, i: j \in G_i} g^{ij} + \sum_{j \in P, j \in P} g^{ij}}} \right\}$$

$$U_l^* \leq \min_{j \in P} \left\{ T_j^{r0}, \sqrt{\frac{\sum_{i \in P} k_i + \sum_{j \in P} c_j}{\sum_{i \in S, j \in P} g_{ij} + \sum_{j \in P, j \in P} g^{ij}}} \right\}$$

A proof of Theorem 15 is in Appendix A.19. In Theorem 15, $(C1)$ guarantees KKT conditions are met for each group, and $(C2)$ ensures no deviation from the partition could improve cost.

4.1.3 Modified MSIRR Algorithm

From our previous analysis, we know that finding the optimal solution to (**PCR**) is equivalent to finding an optimal partition of $R \cup S$. Therefore our goal is to determine the partition and the order intervals simultaneously, while satisfying conditions $(C1)$ and $(C2)$ in Theorem 15.

In Chapter 3, we proposed the MSIRR Algorithm to solve the uncapacitated OWMSMR relaxation problem. In this subsection, we introduce a modified MSIRR algorithm and prove that the solution obtained by the MSIRR algorithm satisfies $(C1)$ and $(C2)$ and is thus optimal to (**PCR**). The modified MSIRR algorithm also determines order intervals from the largest to the smallest iteratively. A major difference is in this capacitated problem, order cycles of retailers are upper bounded because of truck capacity. In each iteration, some suppliers and retailers have already been assigned to some ordered groups in previous iterations, and the remaining are unassigned. Existing groups are ordered by order intervals of the group, where more recently formed groups have smaller order interval. For those suppliers and retailers, we sequentially assume that each supplier, each retailer, and each pair of one supplier and one retailer has the largest order interval among all the unassigned candidates, and calculate the corresponding order interval according to Lemma 5, so that order interval is upper bounded. We then pick the largest such order interval as a candidate to enter the set of assigned groups. If it is smaller than the order intervals of all existing groups (and in particular, smaller than the most recently formed group), the corresponding supplier, retailer, or pair forms a new group, otherwise it is assigned to the most recently formed group, and we recalculate the order interval for the new group according to Lemma 5. If the new order interval is larger then that of some other existing group, we combine the groups again until $(C1)$ is satisfied.

Algorithm: Modified MSIRR Algorithm

Initialize: $\tau \leftarrow \varnothing$, $list_G \leftarrow \varnothing$, $T_{cur} \leftarrow \infty$, $\bar{R} \leftarrow R$, $\bar{S} \leftarrow S$, $\boldsymbol{T^s} \leftarrow \mathbf{0}$, $\boldsymbol{T^r} \leftarrow \mathbf{0}$;

while $\bar{R} \cup \bar{S} \neq \varnothing$ **do**

  forall $i \in \bar{S}$ **do**

    $T_i^s \leftarrow \sqrt{\frac{k_i^s}{\sum_{j \in \bar{R}} g^{ij}}}$;

  end

  forall $j \in \bar{R}$ **do**

    $T_j^r \leftarrow \min\left(T_j^{r0}, \sqrt{\frac{k_j^r}{\sum_{i \in \bar{S}} g^{ij} + \sum_{i \in \bar{S}} g_{ij}}}\right)$;

  end

  forall $i \in \bar{S}$, $j \in \bar{R}$ **do**

    $T_{ij} \leftarrow \min\left(T_j^{r0}, \sqrt{\frac{k_i + k_j^r}{\sum_{i \in \bar{S}} g^{ij} + \sum_{i \in \bar{S}} g_{ij} + \sum_{k \in \bar{S}} g^{kj} + \sum_{k \in \bar{R}} g^{ik} - g^{ij}}}\right)$;

  end

  $i_0 \leftarrow \arg\max_{i \in \bar{S}} T_i^s$, $\quad j_0 \leftarrow \arg\max_{j \in \bar{R}} T_j^s$, $\quad (\hat{i_0}, \hat{j_0}) \leftarrow \arg\max_{i \in \bar{S}, j \in \bar{R}} T_{ij}$;

  if $\max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}}) < T_{cur}$ **then**

    $T_{cur} \leftarrow \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$, add τ to the end of $list_G$;

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \{i_0\}$, $\bar{S} \leftarrow \bar{S} \setminus \{i_0\}$;

    else if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \{j_0\}$, $\bar{R} \leftarrow \bar{R} \setminus \{j_0\}$;

    else

      $\tau \leftarrow \{\hat{i_0}, \hat{j_0}\}$, $\bar{S} \leftarrow \bar{S} \setminus \{\hat{i_0}\}$, $\bar{R} \leftarrow \bar{R} \setminus \{\hat{j_0}\}$;

    end

  else if $\max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}}) = T_{cur}$ **then**

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{i_0\}$, $\bar{S} \leftarrow \bar{S} \setminus \{i_0\}$;

    else if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{j_0\}$, $\bar{R} \leftarrow \bar{R} \setminus \{j_0\}$;

    else

      $\tau \leftarrow \tau \cup \{\hat{i_0}, \hat{j_0}\}$, $\bar{S} \leftarrow \bar{S} \setminus \{\hat{i_0}\}$, $\bar{R} \leftarrow \bar{R} \setminus \{\hat{j_0}\}$;

    end

  else

    if $T_{i_0}^s = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{i_0\}$;

    else if $T_{j_0}^r = \max(T_{i_0}^s, T_{j_0}^r, T_{\hat{i_0}\hat{j_0}})$ **then**

      $\tau \leftarrow \tau \cup \{j_0\}$;

    else

      $\tau \leftarrow \tau \cup \{\hat{i_0}, \hat{j_0}\}$;

    end

    $\bar{S} \leftarrow \bar{S} \cup \tau \cap S$, $\bar{R} \leftarrow \bar{R} \cup \tau \cap R$;

    $K_\tau \leftarrow \sum_{i \in \tau \cap \bar{S}} k_i + \sum_{j \in \tau \cap \bar{R}} c_j$;

    $H_\tau \leftarrow \sum_{\substack{i \in \tau \\ j \in \bar{R} \setminus \tau}} g^{ij} + \sum_{\substack{i \in \tau \\ j \in \tau}} g^{ij} + \sum_{\substack{j \in \tau \\ i \in \bar{S} \setminus \tau}} g^{ij} + \sum_{\substack{i \in \bar{S} \\ j \in \tau}} g_{ij}$;

    $T_{cur} \leftarrow \min_{j \in \bar{R}} \{\sqrt{\frac{K_\tau}{H_\tau}}, T_j^{r0}\}$;

    forall $i \in \tau$, $j \in \tau$ **do**

      $T_i^s \leftarrow T_{cur}$; $T_j^r \leftarrow T_{cur}$;

    end

    $\bar{S} \leftarrow \bar{S} \setminus \tau$, $\bar{R} \leftarrow \bar{R} \setminus \tau$;

  end

end

In each iteration, it takes $O(nm)$ to find i_0, j_0 and $(\hat{i_0}, \hat{j_0})$. In each iteration, either a new group is formed, or two existing groups are combined. Thus in the entire algorithm, grouping occurs at most $m+n$ times, because combined groups are never later partitioned, and this lead to total complexity of $O((n+m)nm)$.

Lemma 6. *The Modified MSIRR algorithm obtains the optimal solution to Problem* (**PCR**)*, and thus to Problem* (**PTR**)*.*

A proof of Lemma 6 is in Appendix A.20. If $\boldsymbol{T^{s*}}$ and $\boldsymbol{T^{r*}}$ denote the optimal solution to (**PCR**) obtained from Modified MSIRR algorithm, the corresponding average cost

$$C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) = C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) = \sum_{l \in \{1, \cdots, k\}} \frac{K(U_l^*)}{U_l^*} + H(U_l^*) \cdot U_l^*, \tag{4.2}$$

where $K(U_l^*)$ and $H(U_l^*)$ are the aggregate fixed and holding cost in optimal partition $P(U_l^*)$, as defined above.

4.1.4 The Lower Bound

For OWMRMS without truck capacity, we proved in the previous chapter that the optimal integer relaxation objective is a lower bound on the cost of an arbitrary policy. In this subsection we extend the result to model with general truck transportation cost. In Theorem 16, we show that $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on the long run cost of any policy.

Theorem 16. $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *(thus* $C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$*) is a lower bound on the average cost of any policy for Problem* (**PT**)*.*

A proof of Theorem 16 is in Appendix A.21. Later we use this lower bound as a benchmark to evaluate the performance of decentralized policies, as well as the performance of heuristics for this NP-hard centralized problem.

4.2 Centralized Heuristics

Before we propose decentralized policies, we introduce two kinds of centralized heuristics for comparison. Later in the computational study, performance of these policies is compared with that of decentralized policies. Note that the centralized problem is NP-hard Arkin et al. (1989).

4.2.1 The Power-of-Two Policy I

In this subsection, we consider a special case of integer ratio policies, the Power-of-Two (PoT) policy. Recall that a PoT policy is a periodic ordering policy such that $T_i^s \in \{m \in \mathcal{N} : 2^m \cdot T_0\}$, where T_0 is the base order interval. A PoT policy is implemented so that two parties with the same reorder interval always order at the same time. It is typical in related literature to build on a lower bound of the type we developed in Section 4.1.4 to develop and analyze a PoT heuristic.

$$
\begin{aligned}
(\mathbf{PPoT})\ \min C_{PoT}(\boldsymbol{T^s}, \boldsymbol{T^r}) = \quad & \sum_{i \in S} \frac{k_i}{T_i^s} + \sum_{j \in R} \frac{c_j}{T_j^r} \lceil \frac{\sum_{i \in S} d_{ij} T_j^s}{q} \rceil + \sum_{i \in S, j \in R} \max (T_i^s, T_j^r) g^{ij} \\
& + \sum_{i \in S, j \in R} T_j^r g_{ij} \\
s.t. \quad & T_i^s, T_j^r \in \{2^m \cdot T_0, m \in \mathcal{N}\}, \forall i \in S, j \in R
\end{aligned}
$$

In the following we show that the PoT policy obtained from $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is an easy-to-implement policy with a worst case ratio of 2. Note that this is significantly worse than the PoT rounding ratio in the uncapacitated problem, where PoT rounding can be done in several EOQ style subproblems and generates a worst case ratio of 1.06. In (**PCR**), there are two factors affecting the effectiveness in the rounding. One is capacity constraints may drive the optimal solution far from the optimal solution to uncapacited problem, thus we might not benefit from robustness near EOQ optimal solution. The other reason is rounding up to the nearest PoT solution may be infeasible because of capacity constraints.

We use $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$, the optimal solution to (**PCR**), to construct a PoT solution as:

$$
T_{i,P}^{s*} = \max \{2^m T_0 : 2^m T_0 \leq T_i^{s*}\} \tag{4.3}
$$

In other words, we round $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ down to the nearest PoT solution. Because we are rounding down the order intervals, capacity constraints are still satisfied. With this rounding rule, we preserve the order of intervals in the solution. That is:

$$
T_i^{s*} \geq T_j^{r*} \Rightarrow T_{i,P}^{s*} \geq T_{j,P}^{r*}.
$$

PoT rounding leads to a new partition:

$$
\frac{U_l^*}{2} < U_l^{P*} \leq U_l^*
$$

Recall that in the MSIRR algorithm, we greedily determine order intervals of suppliers and retailers sequentially from largest to smallest. In each iteration, we either create a new group or combine existing groups whenever condition ($C2$) of Theorem 15 is violated. Recall that we group suppliers and retailers with the same order interval into a partition, so that some members of the set $P(U_l^*)$ (the set of suppliers and retailers with order interval U_l^*) may be combined after rounding, because different U_l^* values could be rounded to the same PoT interval. However, to facilitate our analysis, we continue to consider them separately. Hence, the previous partition for $R \cup S$ still applies, so that $P(U_l^*) = P(U_l^{P*})$. Next, we bound the worst case performance for this feasible PoT policy.

Theorem 17. $C_{POT}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) \leq 2C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. *That is, the total cost of the PoT policy we obtain from rounding* $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ *is no more than twice the optimal centralized cost.*

A proof of Theorem 17 is in Appendix A.22. In the computational analysis in Section 4.7, we show that when per truck transportation cost is moderate, PoT rounding works well. In fact, an extremely large or extremely small per truck transportation cost is required to drive the capacitated optimal solution far from the uncapacitated solution, to come close to the worst case bound of 2.

4.2.2 The Power-of-Two Policy II

In this section, we consider another rounding approach to obtain a feasible PoT policy. In Section 4.2.1, we round down $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ to satisfy capacity constraints. However, such a rounding rule may lead to a low number of truckloads in shipment, thus we consider a different method in this section. We apply the normal PoT rounding rule to come to the nearest PoT solution, and we send two trucks if the capacity constraint is violated. the optimal solution to (**PCR**), to construct a PoT solution as:

$$T_{i,\hat{P}}^{s*} = \min\{2^m T_0 : 2^m T_0 \geq \frac{T_i^{s*}}{\sqrt{2}}\}. \tag{4.4}$$

That is,

$$\frac{T_i^{s*}}{\sqrt{2}} \leq T_{i,\hat{P}}^{s*} < \sqrt{2} T_i^{s*}.$$

With this rounding rule, we preserve the order of intervals in the solution. That is

$$T_i^{s*} \geq T_j^{r*} \Rightarrow T_{i,\hat{P}}^{s*} \geq T_{j,\hat{P}}^{r*}$$

PoT rounding leads to a new partition:

$$\frac{U_l^*}{\sqrt{2}} < U_l^{\hat{P}*} \leq \sqrt{2} U_l^*$$

While some of $P(U_l^*)$ may be combined after rounding, for the convenience of comparison, we continue to consider them separately. Hence the previous partition for $R \cup S$ still applies so that is $P(U_l^*) = P(U_l^{\hat{P}*})$. Therefore

$$C_{PoT}(\boldsymbol{T^{s*}_{\hat{P}}}, \boldsymbol{T^{r*}_{\hat{P}}}) = \sum_{l \in \{1,\cdots,k\}} \Big(\frac{K(U_l^{\hat{P}*})}{U_l^{\hat{P}*}} + H(U_l^{\hat{P}*}) \cdot U_l^{\hat{P}*} \Big)$$

After rounding, capacity constraints for retailers may be violated, thus a second truck may be assigned in some cases.

Theorem 18. *The worst case performance for this PoT rounding is* 2.

A proof of Theorem 18 is in Appendix A.23.

4.2.3 Centralized Zero-Inventory-Ordering Policy

In this subsection, we consider a *Zero-Inventory-Ordering* (ZIO) policy. Instead of doing rounding to get a feasible integer ratio solution, we use the exact solution $\boldsymbol{T^{s*}}$ and $\boldsymbol{T^{r*}}$ obtained in modified MSIRR Algorithm, but changes the quantity to ship in each replenishment of suppliers, to construct a ZIO policy.

The policy is stationary in that each supplier and each retailer has a fixed order interval. Since the policy might not be in the class of integer ratio policies, the same supplier may face different number of orders from same retailers in different order cycles. To help evaluate long run costs of each supplier, we recall previous notation:

- $I_{ij}^w(t)$: the inventory level at time t of product i at the warehouse that is ultimately intended for retailer j
- $I_{ij}(t)$: the inventory at time t of product i at retailer j
- $EI_{ij}^w(t) = I_{ij}^w(t) + I_{ij}(t)$: echelon inventory of product i intended for retailer j (that is, inventory at warehouse intended for retailer j plus the inventory at retailer j

Cost at Retailers

The transportation and installation inventory cost associated with retailer j is:

$$C_j^r(T_j^r) = \frac{c_j^r}{T_j^r} \lceil \frac{\sum_{i \in S} d_{ij} T_j^r}{q} \rceil + \sum_{i \in S} \frac{1}{2} d_{ij} h_{ij} T_j^r = \frac{c_j^r}{T_j^r} \lceil \frac{\sum_{i \in S} d_{ij} T_j^r}{q} \rceil + \sum_{i \in S} (g^{ij} + g_{ij}) T_j^r$$

Cost at Suppliers

Each time a supplier makes a delivery to the warehouse, it needs to ensure that there is enough inventory to cover demands from retailers until the next delivery. However, since orders may not line up, some retailers may order during a particular supplier interval to cover demand during the next supplier order interval, so that the demand from a retailer in a particular supplier interval may exceed the demand that the retailer faces during this interval.

Although the total demand from all the retailers for a single supplier is time-variant and discrete, it is deterministic and known before suppliers make decisions. In particular, each time when a supplier replenishes inventory at warehouse, it knows exact how many orders will occur in the this replenishment cycle, and it can ship the exact amount of product to cover them all.

For supplier i, given any order interval T_i^s, the replenishment quantity can be set equal to the total amount of retailer orders for that product during the interval, resulting in a ZIO policy. To calculate the expected cost for each supplier, as before, we decompose inventory at the warehouse by supplier and intended retailer, and then determine the holding cost associated with each retailer and product.

From derivations in the previous chapter, the cost at supplier i is:

$$C_i^{zio}(T_i^s) = \frac{k_i^s}{T_i^s} + \sum_{j \in L_i \cup G_i} g^{ij} T_i^s$$

Policy Effectiveness

The total cost to suppliers of the ZIO policy is:

$$\begin{aligned} C_{zio}(\boldsymbol{T^s}, \boldsymbol{T^r}) &= \sum_{i \in S} C_i^{zio}(T_i^s) + \sum_{j \in R} C_j^r(T_j^r) \\ &= \sum_{i \in S} \left\{ \frac{k_i^s}{T_i^s} + \sum_{j \in L_i \cup G_i} g^{ij} T_i^s \right\} + \sum_{j \in R} \left\{ \frac{c_j^r}{T_j^r} \lceil \frac{\sum_{i \in S} d_{ij} T_j^r}{q} \rceil + \sum_{i \in S} (g^{ij} + g_{ij}) T_j^r \right\}. \end{aligned} \tag{4.5}$$

In Section 4.1.4 we showed that $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on optimal cost. Hence it suffices to evaluate the worst case ratio of $C_{zio}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ to $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$.

In the next theorem, we prove the effectiveness of this ZIO policy with $\boldsymbol{T^{s*}}$ and $\boldsymbol{T^{r*}}$ is at least $\frac{2}{3}$.

Theorem 19. $C_{zio}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) \le \frac{3}{2} C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$

A proof of Theorem 19 is in Appendix A.24.

4.3 Comparison with the Solution from the Uncapacitated Model

In this section, we compare our truck cost models with our simple fixed cost models from Chapter 3. If we apply our simple model with fixed transporation cost here, the optimal solution of uncapacitated problem may exceed a full truckload for some retailers. In that case, we ship multiple trucks in the same replenishment and pay more for transportation cost. We show that the worst case ratio is $\frac{8}{5}$, and the bound is tight.

If we denote $\tilde{\boldsymbol{T}}^{\boldsymbol{s}*}$ $\tilde{\boldsymbol{T}}^{\boldsymbol{r}*}$ as the optimal solution from MSIRR in Section 3.1.3, then the corresponding cost is

$$C_{sim}(\tilde{\boldsymbol{T}}^{\boldsymbol{s}*}, \tilde{\boldsymbol{T}}^{\boldsymbol{r}*}) = \sum_{i \in S} \frac{k_i^s}{\tilde{T}_i^{s*}} + \sum_{j \in R} \frac{c_j^r}{\tilde{T}_j^{r*}} \lceil \frac{\sum_{i \in S} d_{ij} \tilde{T}_j^{r*}}{q} \rceil + \sum_{i \in S, j \in R} max(\tilde{T}_i^{s*}, \tilde{T}_j^{r*}) g^{ij} + \sum_{i \in S, j \in R} \tilde{T}_j^{r*} g_{ij}$$

Theorem 20. $\sup \left\{ \frac{C_{sim}(\tilde{\boldsymbol{T}}^{\boldsymbol{s}*}, \tilde{\boldsymbol{T}}^{\boldsymbol{r}*})}{C_C(\boldsymbol{T}^{\boldsymbol{s}*}, \boldsymbol{T}^{\boldsymbol{r}*})} \right\} \geq \frac{8}{5}$. *That is, the worst case performance using a simple model incurs at least 60% cost increase.*

A proof of Theorem 20 is in Appendix A.25. Thus we know if we only use simple model without differentiating the number of trucks each retailer use, the worst case performance can incure at least 60% cost increase, and this bad case is easily obtained when the optimal number of trucks is slightly more than two in the simple model.

4.4 A Retailer-Driven Decentralized ZIO Policy

In the previous section, we assumed that the system was operated under centralized control, so that shipments from different suppliers to the warehouse, and shipments from the warehouse to retailers could be coordinated to minimize overall system costs. In this section, we consider a decentralized model where suppliers and retailers make their own decisions based on their information. As discussed in the introduction, we are motivated by current practice, where (at least at the MACC with which we worked), suppliers deliver products to the warehouse, paying transportation costs, unloading costs, and holding costs until goods are shipped to retailers, and retailers order from the warehouse, paying transportation costs.

Specifically, we assume that suppliers pay a fixed transportation cost per shipment as well as holding cost at the warehouse, and must meet retailer demand. Similarly, we assume that retailers must pay a per truck transportation cost for deliveries, as well as the holding cost at their own stores, and must meet customer demand without backorder.

We assume that retailers first optimize their own strategy, and then suppliers must react to this strategy, in line with what we have observed in practice. In this section, we analyze the optimal retailers' strategy and propose one stationary fixed order interval heuristics for the challenging-to-optimize suppliers' problem, which is a ZIO policy.

We summarize the new notation for this section below. We use Γ to denote order intervals under decentralized models.

- Γ_j^r (Γ_j^{r*}) : (optimal) order interval for retailer j in the decentralized model
- $\boldsymbol{\Gamma^{r*}} = (\Gamma_1^{r*}, \Gamma_2^{r*}, \cdots, \Gamma_m^{r*})$: vector of optimal order intervals for all retailers in the decentralized model
- Γ_i^s (Γ_i^{s*}) : (optimal) order interval for supplier i in the decentralized model
- $\boldsymbol{\Gamma^{s*}} = (\Gamma_1^{s*}, \Gamma_2^{s*}, \cdots, \Gamma_n^{s*})$: vector of optimal order intervals for all suppliers

4.4.1 Retailers' Policy

Note that each retailer will have an optimal ZIO policy (indeed, this can be viewed as a capacitated EOQ problem at the retailer). Each order will thus contain products from all suppliers sufficient to cover demand during that cycle. Hence, the optimal strategy is a fixed order interval strategy, and the problem for retailer j is:

$$(\mathbf{PTDR_j}) \quad \min C_j^r(\Gamma_j^r) = \frac{c_j^r}{\Gamma_j^r}\lceil \frac{\sum_{i\in S} d_{ij}}{q} \rceil + \sum_{i\in S} \frac{1}{2} d_{ij} h_{ij} \Gamma_j^r = \frac{c_j^r}{\Gamma_j^r}\lceil \frac{\sum_{i\in S} d_{ij}}{q} \rceil + \sum_{i\in S} (g^{ij} + g_{ij})\Gamma_j^r$$

Follow the same logic, each time retailer orders exact one truckload. Hence ($\mathbf{PTDR_j}$) is the same as a capacitated EOQ problem:

$$(\mathbf{PDRC_j}) \quad \min C_{C,j}^r(\Gamma_j^r) = \frac{c_j^r}{\Gamma_j^r} + \sum_{i\in S} (g^{ij} + g_{ij})\Gamma_j^r$$
$$s.t. \quad \Gamma_j^r \leq T_j^{r0}$$

By the KKT condition, we obtain the optimal order interval for retailer j:

$$\Gamma_j^{r*} = min\left\{ T_j^{r0}, \sqrt{\frac{k_j^r}{\sum\limits_{i\in S}(g^{ij} + g_{ij})}} \right\}.$$

The optimal decentralized cost per unit time for retailer j is therefore:

$$C_{C,j}^r(\Gamma_j^{r*}) = \frac{c_j}{\Gamma_j^{r*}} + \Gamma_j^{r*}(g^{ij} + g_{ij}). \tag{4.6}$$

Theorem 21. *The optimal order interval for each retailer is longer in centralized model than in the decentralized model. That is,* $\Gamma_j^{r*} \leq T_j^{r*}$.

A proof of Theorem 21 is in Appendix A.26.

This follows because in the centralized model, the shipping decision is made based on the marginal additional holding cost at the retailer, whereas in the decentralized model the shipping decision accounts for the fact that all of the holding cost is paid by the retailer.

4.4.2 Suppliers' Policy: The Zero-Inventory-Ordering Policy

Similar to what we derive in Section 3.2, under a ZIO policy, the problem for supplier i is:

$$(\mathbf{PDS_i^{zio}}) \quad \min C_i^{zio}(\Gamma_i^s) = \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i} \hat{\underline{H}}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in E_i} \underline{H}_{ij}^e(\Gamma_i^s, \Gamma_j^{r*}) + \sum_{j \in G_i} \hat{\underline{H}}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*})$$
$$= \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i \cup G_i} g^{ij} \Gamma_i^s$$

And the optimal solution is selected from the same limited set:

Theorem 22. *The optimal order interval in the decentralized ZIO policy*

$$\Gamma_i^{s*} \in \{\Gamma_1^{r*}, \Gamma_2^{r*}, \cdots, \Gamma_m^{r*}, \tilde{\Gamma}_i^{s*}\}, \text{ where } \tilde{\Gamma}_i^{s*} = \sqrt{\frac{k_i^s}{\sum_{j \in R} g^{ij}}}.$$

A proof of Theorem 22 can be found in Appendix A.9.

4.5 An Easily Implementable Retailer-Driven Decentralized Policy

4.5.1 Suppliers' Policy: The Order-Up-To Policy

As in Section 3.3, we have the following upper bound on the the cost faced by supplier i when this *order-up-to* policy is employed:

$$(\mathbf{PDS_i^{out}}) \quad \min C_i^{out}(\Gamma_i^s) = \frac{k_i^s}{\Gamma_i^s} + \sum_{j \in L_i} g^{ij} \Gamma_i^s - \sum_{j \in G_i} g^{ij} \Gamma_i^s + 2 \sum_{j \in L_i \cup G_i} g^{ij} \Gamma_j^{r*}$$

Theorem 23. *The optimal order interval for supplier* $i \in \{\Gamma_1^{r*}, \Gamma_2^{r*}, \cdots, \Gamma_m^{r*}, \Gamma_{i,k'}^{s*}\}$, *where* k' *is the only index such that* $\hat{\Gamma}_{i,k}^{s*} \in (\Gamma_k^{r*}, \Gamma_{k+1}^{r*})$. *That is,*

$$\Gamma_i^{s*} = argmin\left\{ C_i^{out}(\Gamma_i^s) : \Gamma_i^s \in \left\{ \Gamma_j^{r*}, \forall j, \ or\ \Gamma_{i,k'}^{s*} \right\} \right\},$$

where $\Gamma_{i,k'}^{s*}$ *is the only solution that satisfies local optimality.*

Though this *order-up-to* policy may generate higher inventory cost, it is easy to implement in reality with a bounded worst case ratio, as we show in next subsection.

4.6 Cost of Decentralization

In this section, we consider the worst case performance (with respect to optimal centralized performance) of the two decentralized policies proposed in Section 4.4 and Section 4.5, using the bound $C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ developed in Section 4.1.4 as a benchmark – in effect, we bound the cost increase due to decentralization. The worst case performance is bounded in both *zero-inventory-ordering* and *order-up-to* supplier policies.

4.6.1 The Cost of Decentralization Using the Decentralized ZIO Policy

Building on this development, inn our decentralized ZIO model, the total cost suppliers and retailers pay in the optimal solution is:

$$C_{zio}(\boldsymbol{\Gamma^{s*}}, \boldsymbol{\Gamma^{r*}}) = \sum_{i \in S} C_i^{zio}(\Gamma_i^{s*}) + \sum_{j \in R} C_j^r(\Gamma_j^{r*}). \tag{4.7}$$

In Theorem 16 we also showed $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on the cost of centralized model. Therefore performance of decentralized ZIO policy is bounded by:

$$\frac{C_{zio}(\boldsymbol{\Gamma^{s*}}, \boldsymbol{\Gamma^{r*}})}{\text{Optimal Centralized Cost}} \leq \frac{C_{zio}(\boldsymbol{\Gamma^{s*}}, \boldsymbol{\Gamma^{r*}})}{C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})}$$

bound the ratio of total decentralized ZIO cost to optimal centralized cost:

Theorem 24. *The total cost of the ZIO decentralized system is no more than* $\frac{3}{2}$ *times the optimal cost in the centralized system.*

A proof of Theorem 24 is provided in Appendix A.27. Note that we proved in Theorem 21 that retailers order less frequently in centralized model compared to decentralized model. Unfortunately, we have not found such results for suppliers. As a result, each supplier need to search all $m + 1$ candidates to determine the optimal one.

4.6.2 The Cost of Decentralization Using the OUT Policy

In our decentralized *order-up-to* model, the total cost suppliers and retailers pay in the optimal solution is:

$$C_{out}(\mathbf{\Gamma^{s*}}, \mathbf{\Gamma^{r*}}) = \sum_{i \in S} C_i^{out}(\Gamma_i^{s*}) + \sum_{j \in R} C_j^r(\Gamma_j^{r*}). \tag{4.8}$$

In this subsection, we analyze how well our decentralized model is performed compared to the optimal policy in centralized setting. In Section 4.1.4 we have shown that $C_{TIR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on optimal centralized cost. Hence it suffices to evaluate the worst case ratio of $C_C(\mathbf{\Gamma^{s*}}, \mathbf{\Gamma^{r*}})$ to $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. The ratio of optimal decentralized cost to the optimal centralized cost is called "price of anarchy".

In the next theorem, we prove price of anarchy for this *order-up-to* policy is at most 2.5.

Theorem 25. *The total cost of decentralized system is no more than $\frac{5}{2}$ times the cost of centralized system.*

A proof of Theorem 25 is in Appendix A.28.

4.6.3 Decentralized Policies are Effective

As we see from Theorem 24 and Theorem 25, both the decentralized *zero-inventory-ordering* and *order-up-to* policies are relatively effective; indeed, the decentralized *zero-inventory-ordering* policy is at most 1.5 times a lower bound on the optimal centralized policy. In Section 4.7, we show using a computational study that the loss due to decentralization is typically much smaller than its theoretical maximum. This serves to motivate and justify current practice, where information sharing and centralized control is strictly limited, and to suggest that it the additional effort and expensive to centralize control (if it is possible) is likely not worth the trouble.

4.7 Computational Study

We use a computational study to better understand the cost of decentralization in various settings. We have shown in Theorem 19, Theorem 24, Theorem 25 and Theorem 18 that all our centralized and decentralized heuristics have bounded worst case performance. Here, we explore how tight those bounds are, how these heurisitcs perform in different scenarios, and how centralization benefits the system. In particular, we run this simulation to explore the following questions:

- How does changing the number of suppliers relative to number of retailers affect the system performance?
- How do variations in costs affect performance of different policies?
- How does truck capacity affect centralized decentralized operation?

Throughout computational study, we apply our MSIRR Algorithm to compute $C_{TIR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ as a benchmark for optimal centralized cost. Then we search for the optimal decentralized ZIO and OUT policies based on Theorem 22 and Theorem 23. Given these, we calculate ratios of decentralized costs to centralized cost. For comparison, we also calculate centralized ZIO and PoT heuristics according to MSIRR Algorithm and (4.3). All parameters are randomly generated, and for each set of parameter generating distributions, we generate 20 cases and record the average cost ratio as well as the largest (worst) cost ratio. We also calculate coefficient of variation (ratio of standard deviation to mean) for each set of experiments, but it was very small (for most cases $< 5\%$), so for clarity of presentation we omitted it. We ran over $100,000$ cases with different parameter settings, and we summarize our results below.

4.7.1 General Observations

Overall, our computational study suggests that the cost of decentralization is relatively insignificant, particularly, compared to the benefits of easy implementation and information privacy. In the computational study, we use $C_{TIR}(T^{s*}, T^{r*})$, obtained in Section 4.1.3, as a benchmark, and evaluate the ratio of cost of each policy to this lower bound. In Figure 4.2, we show the simulation results from $20,000$ runs, and plot cost ratio histogram for a variety of different policies, where $n, m \sim unif\{3, \cdots, 20\}, h_{ij} \sim Unif(0,1), h_i \sim Unif(0,1), d_{ij} \sim Unif(1,2), k_i^s \sim Unif(1,2), k_j^r \sim Unif(1,2)$. We observe that the decentralized ZIO policy is quite stable, and its cost ratio is usually less than 120%. The decentralized OUT policy incurs a higher holding cost, but even then cost ratio to the lower bound is still less than 140% in most cases. Performance of the centralized ZIO policy is almost the same as the decentralized version of the same policy, but slightly more variable. The ratio of centralized PoT cost to the cost lower bound is usually below 120%, but it fluctuates in certain scenario and performance can be extremely bad.

In general, performance of all policies is much better than the theoretical worst case, and the decentralized policies typically perform almost as well as centralized policies.

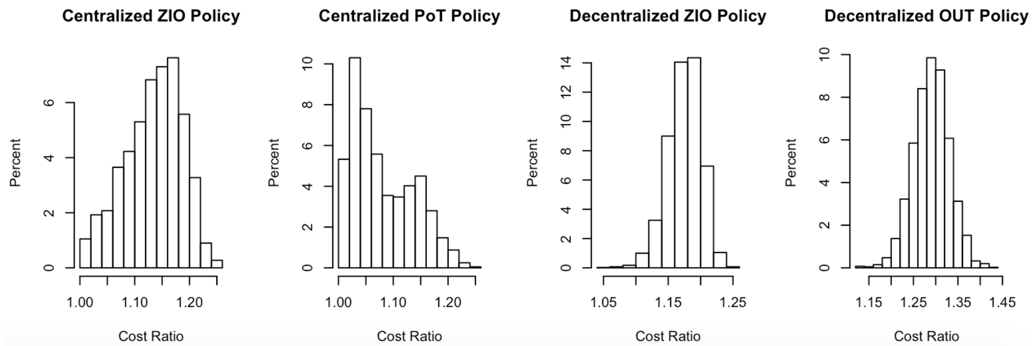


Figure 4.2: Cost Ratios of Different Policies

4.7.2 Number of Suppliers/Retailers

In this subsection, we analyze cost ratios for different policies with varying the number of suppliers and retailers. All suppliers and retailers are assumed to be identical. We tested a variety of distributions, and observed similar results given similarly scaled parameters, so we report our results for parameters generated using uniform distributions for illustration: $h_{ij} \sim Unif(0,1), h_i \sim Unif(0,1), d_{ij} \sim Unif(1,2), k_i^s \sim Unif(1,2), k_j^r \sim Unif(1,2)$. We ran simulation in two sets of scenarios, with small truck capacity $q = 5$ and large truck capacity $q = 20$, and initially let $m = 8, n \in \{1, 2, \cdots, 50\}$, and explored how the number of suppliers affects cost ratio. For each m, we randomly generated 20 sets of parameters, and calculated the worst case ratio and average ratio for the two centralized policy and two decentralized policies. Next, we let $n = 8, m \in \{1, 2, \cdots, 50\}$ and reran the experiments.

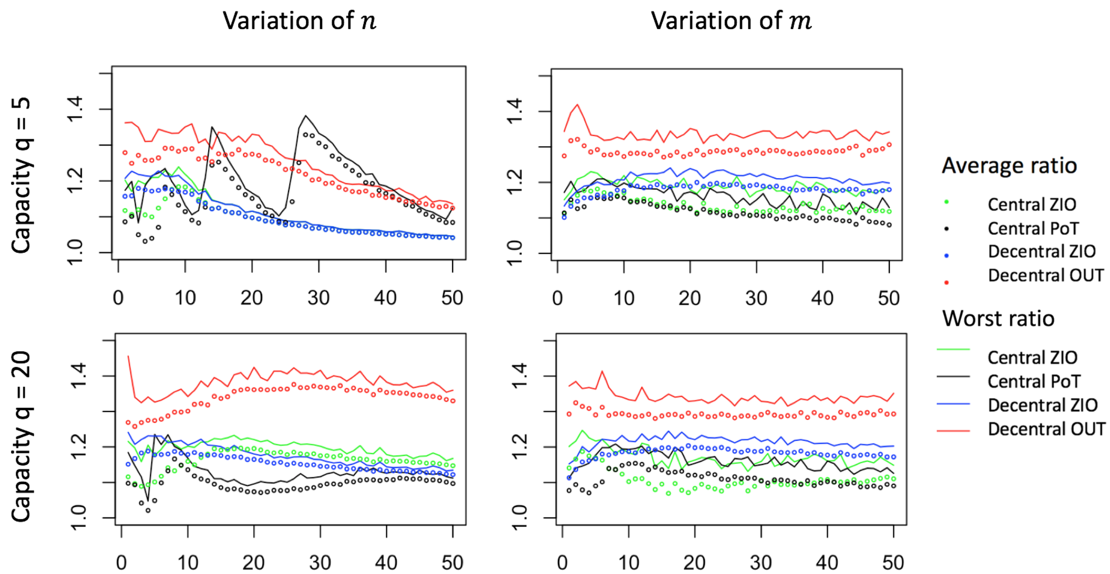


Figure 4.3: Number of participants

As can be observed in Figure 4.3, as the number of retailers or suppliers increases, the decentralized OUT policy becomes increasingly bad, while other three policies perform better. In general when all costs are of the same order of magnitude, the centralized PoT policy and centralized ZIO policy perform best when truck capacity is large and small, respectively. However, the performance of decentralized ZIO policy is stable and very close to the best centralized policy in both cases, with cost increase below 20% compared with the lower bound, demonstrating the effectiveness of decentralization.

4.7.3 Cost Variation

In this subsection, we analyze how cost variation affects performance of different policies.

Scaling in Fixed Cost

We again use a similar strategy to generate cost parameters: $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,1)$, $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8$. Different from last subsection, we keep $k^r_j \sim Unif(1,2)$ and differentiate suppliers by letting $k^s_i \sim \begin{cases} Unif(1,2k), & \text{if } i \leq 4 \\ Unif(1,2), & \text{if } i > 4 \end{cases}$, where $k \in \{1, 2, \cdots, 50\}$. We test policy performance with large truck capacity ($q = 30$) and small truck capacity ($q = 5$). We generate 20 sets of parameters for each k, apply different policies and record the average and worst case cost ratios. Similarly, we run experiments with $k_i^s \sim Unif(1,2)$ and retailers' fixed cost $k_j^r \sim \begin{cases} Unif(1,k), & \text{if } j \leq 4 \\ Unif(1,2), & \text{if } j > 4 \end{cases}$, on same set of k.

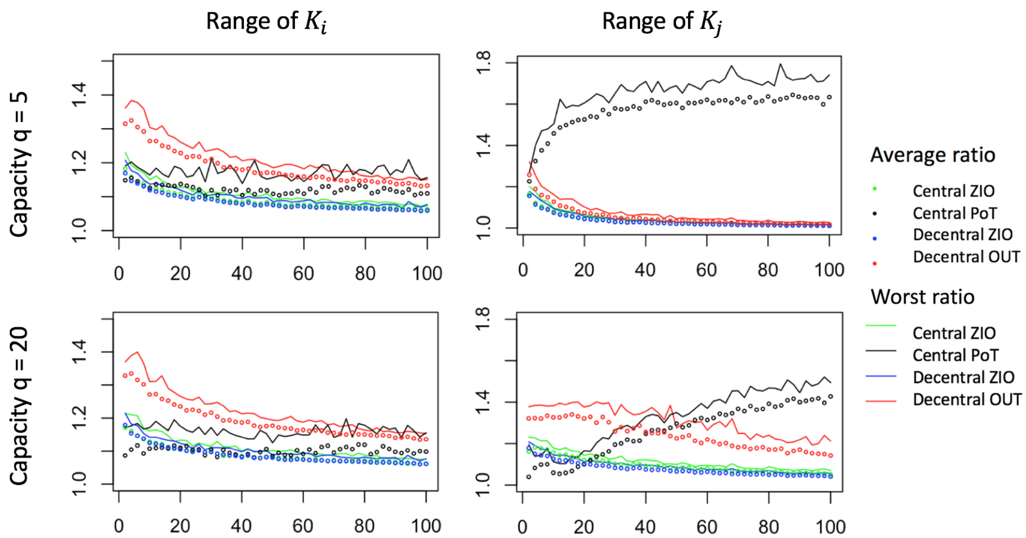


Figure 4.4: Fixed cost scaling with different truck capacity

From Figure 4.4 we observe that the decentralized ZIO policy performs best in almost all cases when transportation costs at retailers are well differentiated. When truck capacity is small, even the performance of the decentralized OUT policy is close to the optimal. On the contrary, the performance of centralized PoT policy is relatively unstable, and the performance of centralized ZIO heuristic is slightly worse than the decentralized policy. In general, this is true because the more diversified suppliers or retailers are, the less valuable centralization is, and thus, the less we lose because of decentralization.

Variation in Holding Cost

We next explore how holding cost variations affect system performance. As before, we analyze policy performance with relatively large (q=5) and small (q=1) truck capacity and generate costs as follows: $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8, k_i^s \sim Unif(1,2)$, $k_j^r \sim Unif(1,2)$. To explore the effect of holding cost at the warehouse, we let $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,k)$ where $k \in \{1,2,\cdots,50\}$. Similarly, we let $h'_{ij} \sim Unif(0,k)$, $h^i \sim Unif(0,1)$ to analyze effect of echelon holding costs at retailers.

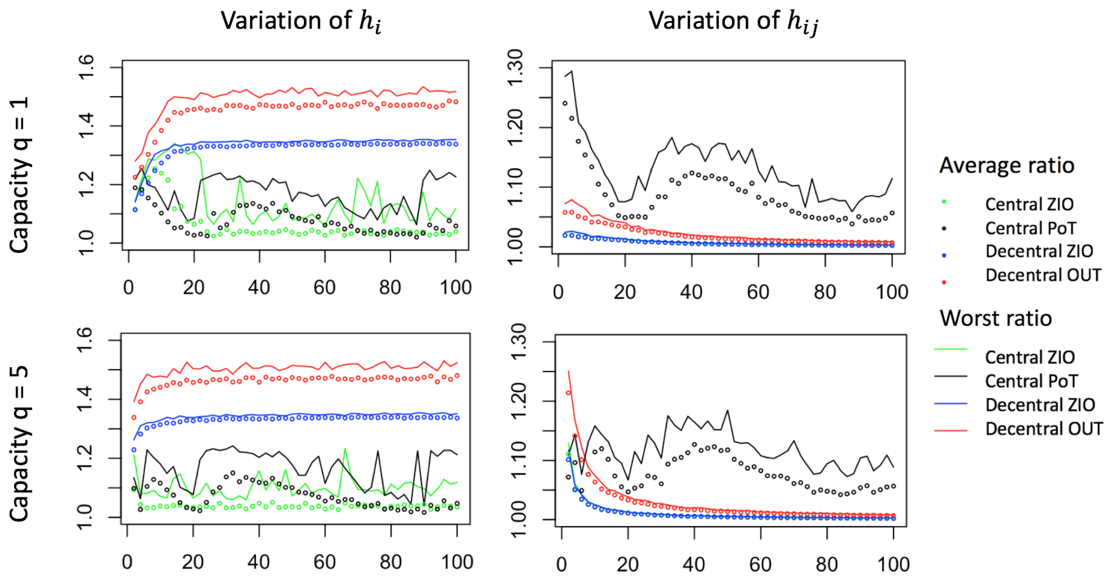


Figure 4.5: Holding cost scaling with different truck capacity

As we observe in Figure 4.5, decentralized policies outperform centralized policies when holding costs at different retailers are differentiated, with no more than 10% cost increase compared to the cost lower bound. On the other hand, when the holding cost at warehouse is differentiated, the performance of decentralized policies is slightly worse, but more stable than the performance of centralized policies. Even in the worst case, the cost of decentralization is about 20%, but with the benefit of information privacy.

Truck Capacity

Last but not least, we directly explore the influence of truck capacity. We let $h'_{ij} \sim Unif(0,1)$, $h^i \sim Unif(0,1)$, $d_{ij} \sim Unif(1,2)$, $m = 8, n = 8, k^r_j \sim Unif(1,2), k^s_i \sim Unif(1,2)$, and we differentiate truck capacity by letting $q \in \{\frac{1}{5}, \frac{2}{5}, \cdots, 10\}$.

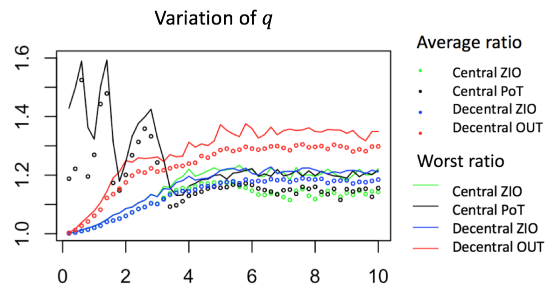


Figure 4.6: Diversity in truck capacity

From Figure 4.6 we see that ZIO policies work best among all policies, and there is no significant difference between the centralized and decentralized versions. The decentralized OUT policy works slightly worse than ZIO policies because of additional holding cost. The centralized PoT policy works almost as well as ZIO when truck capacity is large enough, but performs worse than other policies otherwise. Thus we do not save much by centralization, despite gathering the private information of all retailers.

Comparison to the Simple Model

Another important question we would like to explore involves how much we lose if we just use our simple model without truck capacity constraints, as in the previous chapter.

It is natural to see that if the shipment is less that a truckload (LTL), then our capacitated models actually degenerates to our simple model. However, for the case when a full truck load is not enough for a single shipment, we need to pay more transportation cost if we implement the solution obtained from previous our uncapacitated model.

In the following graph we explore the difference between the capacitated models proposed in this paper, and our simple uncapacitated model from previous paper. We ran 20 cases and recorded the average cost ratio for each set of parameter distributions. For the simple model, we use PoT policy which performs best among all policies in uncapacitated case.

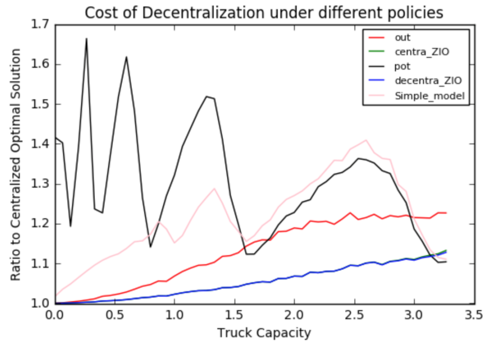


Figure 4.7: Comparison with Uncapacitated Model

As we observe from Figure 4.7, both ZIO and *order-up-to* policies outperform simple uncapacitated policies when truck capacity is an issue. Centralized and deventralized ZIO policy has similar results for most of the cases, since they all use a full truck to deliver products for retailers. The *order-up-to* policy generates higher cost compared to ZIO for the higher inventory level it holds, but it is still better than the uncapacitated PoT policy.

4.7.4 Summary

Based on all of our experiments, centralization does not lead to dramatic benefits. Indeed, our decentralized policies, the ZIO and OUT policies, work well, especially when holding costs at different retailers are well differentiated, when holding costs at warehouse are relatively lower than at the retailers, or when fixed costs at suppliers are diverse. The cost ratio relative to the lower bound associated with the decentralized ZIO policy is usually no more than 10% relative to the cost lower bound, while the cost increase of OUT policy is less than 30%. In comparison, the centralized ZIO performs almost the same as the decentralized version, while the traditional centralized PoT policy becomes unstable when we are approaching full truck load. Overall, decentralization is a good choice for collaboration – its experimental performance dominates theoretical bounds.

Chapter 5

A Stochastic OWMRMS Model

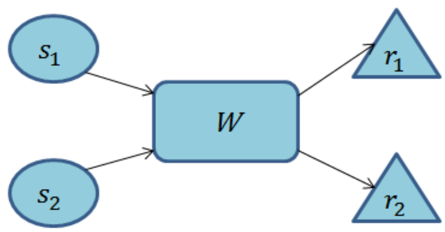
In Chapter 3 and Chapter 4, we considered a deterministic version of this setting, and showed that effective decentralized policies are easy to find, and perform almost as well as challenging (or impossible) to implement centralized policies.

In this chapter, reflecting MACC's that we have observed in practice, we assume a stochastic decentralized setting, in which each supplier and each retailer operates via an independent inventory policy, without sharing private information – suppliers do not share the demand they face from each retailer with other retailers or other suppliers, and retailers do not share their demand information between suppliers, or with other retailers. Our goal is determine an effective operating strategy in this setting. To do this, we consider a setting where several suppliers ship to several retailers through a shared warehouse, so that outbound trucks from the warehouse contain the products of multiple suppliers. We propose an aggregate (Q, S) policy for retailers where an order is placed whenever total demand for all products accumulate to Q since last order. Similarly, we use a typical (s, S) policy for suppliers. We further show that under independent Poisson demand processes, optimal parameter settings are found by our algorithm. All of these policies are also easy to monitor and implement in this setting.

5.1 Model Setting

We consider a model in which multiple suppliers and multiple retailers share a common warehouse. Each supplier produces a unique product and supplies all retailers (thus we refer to products and suppliers interchangeably in what follows). By using a central warehouser, suppliers can transport all goods to the central warehouse together, regardless of their ultimate retailer destinations. In the same way, outbound warehouse trucks heading to a single retailer can carry multiple products. For simplicity of notation, in this paper we use i as the index associated with suppliers and j as the index associated with retailers. We assume independent Poisson demand with rate λ_{ij} at retailer j for product i. By the merging property of Poisson processes, total demand for product i is $\lambda_i = \sum_j \lambda_{ij}$, and total demand at retailer j is $\lambda_j = \sum_i \lambda_{ij}$. We assume constant lead times both from the warehouse to retailer (L_j) and from the supplier to the warehouse (L_i).

Figure 5.1: Multi-echelon inventory systen



Fixed cost k_i is incurred whenever supplier i replenishes its inventory at the warehouse, and similarly k_j is charged when retailer j places an order from the warehouse. Linear holding costs h_i and h_{ij} per unit of inventory per unit time are charged both at the warehouse and at retailer. Unmet demand is fully backlogged at retailer with penalty rate p_{ij} per unit per unit time. For tractability and stability of the system, and following a fairly common convention in related literature, we assume that if inventory of some product at the warehouse is insufficient to cover orders from retailers, the supplier must obtain supplemental product from other sources and pay an additional fee p_i^0 for each unmet unit of demand (e.g., expediting higher-cost supply from other sources).

For this model, the optimal policy is likely to be complicated and difficult to obtain. Thus, we propose an intuitive (Q, S)-style decentralized policy. At aach retailer, an order is placed whenever the total demand it experiences for all products since its last replenishment reaches Q_j, to raise the inventory position to S_{ij}. Similarly each supplier raises its inventory position at warehouse to S_i whenever it drops to $S_i - Q_i$.

The decision flow in this setting is retailer-driven. More specifically, each retailer makes an individual replenishment plan based on private information, and suppliers observe retailer orders and then make ordering decisions. Each party focuses on minimizing its own long run cost.

We summarize notation in the following table:

Notation	Description
λ_{ij}	Poisson demand rate of product i at retailer j
λ_i, λ_j	total Poisson demand rate for product i, and at retailer j
$D^i_{[t_1,t_2)}$, $D^{ij}_{[t_1,t_2)}$	total demand occured in time $[t_1, t_2)$ of product i , and at retailer j
k_i, k_j	fixed cost from supplier i to warehouse, and from warehouse to retailer j
h^i, h_{ij}	holding cost rate for product i at warehouse and at retailer j
$h'_{ij} = h_{ij} - h^i$	echelon holding cost rate of product from supplier i at retailer j
p_{ij}	backorder penalty cost for product i at retailer j
p^0_i	cost to replenish one unit of product i from outsource
L_i, L_j	lead time from supplier i to warehouse, and from warehouse to retailer j
Q_i, Q_j	demand level that triggers replenishment from supplier and for retailer
S_i, S_{ij}	Order-up-to levels at warehouse, and at retailer

5.2 Retailer Policy

In this section, we analyze an easy-to-implement decentralized policy for retailers. To coordinate replenishments for all products, we propose a continuous review $(Q, \boldsymbol{S})$ policy for all retailers. That is, at any retailer j, we monitor the aggregate demand for all products since that retailer's last order, and whenever the aggregate demand reaches Q_j the retailer places an order at the warehouse to replenish inventory positions for all products i to their order-up-to levels S_{ij}.

Thus, the aggregate inventory position at each retailer follows a regenerative process with renewal points at all replenishment times. For simplicity of notation, we define an aggregate order-up-to level $S_j = \sum_i S_{ij}$. Under the assumption of independent Poisson demand processes, it is well known by renewal theory that aggregate inventory position is uniformly distributed in $\{S_j, S_{j-1}, \cdots, S_j - Q_j + 1\}$. Next, we introduce additional notation:

- X_{ij} : aggregate demand for product i at retailer j since last replenishment
- $X_j = \sum_{i \in S} X_{ij}$: aggregate demand at retailer j after the last replenishment
- V_{ij} : difference between the inventory level for product i at retailer j and the order-up-to level S_{ij}
- $\theta_{ij} = \frac{\lambda_{ij}}{\lambda_j}$: the probability that the next demand at retailer j is for product i

5.2.1 Cost Evaluation

Since the demand for each products is independent, we know $X_{ij}|X_j = x_0 \sim Bin(x_j, \theta_{ij})$, where $Bin(n, p)$ is the binomial distribution with n trials and success probability p. Therefore, the aggregate demand for product i at retailer j since last replenishment is as follows:

$$\begin{aligned}
P(X_{ij} = x) &\triangleq u_i(x) \\
&= \sum_{x_j=x}^{Q_j-1} P(X_j = x_j) \cdot P(X_{ij} = x|X_j = x_j) \\
&= \frac{1}{Q_j} \sum_{x_j=x}^{Q_j-1} \binom{x_j}{x} \theta_{ij}^{x} (1-\theta_{ij})^{x_j - x}
\end{aligned}$$

Therefore

$$\begin{aligned}
u_i(x)\theta_{ij}Q_j &= \sum_{x_j=x}^{Q_j-1} \binom{x_j}{x} \theta_{i}^{x+1} (1-\theta_{ij})^{x_j - x} \\
&= \sum_{x_j=x}^{Q_j-1} P((x+1)^{th} \text{ success occurs at } x_j + 1) \\
&= P((x+1)^{th} \text{ success occurs at or before } Q_j) \\
&= P(\text{at least } (x+1) \text{ successes occurs at } Q_j) \\
&= 1 - B(x; Q_j, \theta_{ij}), \qquad (5.1)
\end{aligned}$$

where $B(x; n, p)$ is the cumulative density function of the Binomial distribution $Bin(n, p)$. In addition, we know that distribution of the inventory level at time $t + L_j$ is determined by inventory position at time t, because replenishment orders placed after t will not arrive by time $t + L_j$ and that will have no effect on inventory level at that time. Hence

$$\begin{aligned}
P(V_{ij} = v) &\triangleq m_i(v) \\
&= \sum_{x=0}^{\min\{Q_j, v\}} u_i(x) \cdot P(D^{ij}_{[t, t+L_j)} = v - x) \\
&= \frac{1}{\theta_{ij}Q_j} \sum_{x=0}^{\min\{Q_j, v\}} (1 - B(x; Q_j, \theta_{ij})) \cdot \frac{e^{-\lambda_{ij}L_j}(\lambda_{ij}L_j)^{v-x}}{(v-x)!} \qquad (5.2)
\end{aligned}$$

Therefore the long run average cost at retailer j is:

$$C_j(Q_j, S_{1j}, \cdots, S_{nj}) = \frac{k_j}{Q_j}\lambda_j + \sum_{i\in S} h_{ij}\Big(S_{ij} - \frac{\lambda_{ij}(Q_j-1)}{2} - \lambda_{ij}L_j\Big) + \sum_{i\in S}(p_{ij}+h_{ij})\sum_{v\geq S_{ij}}(v-S_{ij})m_i(v). \tag{5.3}$$

In (5.3), $S_{ij} - \frac{\lambda_{ij}(Q_j-1)}{2} - \lambda_{ij}L_j$ is the average inventory level of product i at retailer j, and $\sum_{v\geq S_{ij}}(v-S_{ij})m_i(v)$ is the average backorder.

5.2.2 Convexity in S_{ij} and Q_i

$C_j(Q_j, S_{1j}, \cdots, S_{nj})$ involves complicated calculations of Q_j and S_{ij}. Thus instead of optimizing (5.3) directly, we first fix aggregate quantity Q_j and analyze the optimal order-up-to level accordingly.

Theorem 26. *When aggregate order quantity Q_j is fixed, the total cost $C_j(Q_j, S_{1j}, \cdots, S_{nj})$ is convex in the order-up-to level S_{ij}. The optimal $S^*_{ij}(Q_j) = M_i^{-1}\Big(\frac{p_{ij}}{p_{ij}+h_{ij}}\Big)$ is a newsvendor-like quantity, where $M_i(v) = \sum_{t\leq v} m_i(t)$ is the cumulative density function of $V_i(t)$.*

A proof of Theorem 26 is in Appendix A.29. Notice that $M_i(v)$ is a complicated function of Q_j from (5.2), but we can show the that $S^*_{ij}(Q_j)$ is monotonic.

Theorem 27. *$S^*_{ij}(Q_j)$ is monotonically increasing in Q_j.*

A proof of Theorem 27 is in Appendix A.30. From (5.2) we see that $M_i(v)$ is a complicated function of Q_j; however, we can further show that $S^*_{ij}(Q_j)$ is also monotonic, which can potentially simplify computation.

Theorem 28. *When order-up-to levels S_{ij} are fixed, the total cost $C_j(Q_j, S_{1j}, \cdots, S_{nj})$ is convex in aggregate order quantity Q_j. Thus the optimal aggregate level*

$$Q^* = argmin_{Q\in\mathbb{Z}}\{Q : C_j(Q+1, S_{1j}, \cdots, S_{nj}) \geq C_j(Q, S_{1j}, \cdots, S_{nj})\}.$$

A proof of Theorem 28 is in Appendix A.31. Thus, optimal order-up-to levels can be effectively found by using convexity and monotonicity. In the next theorem, we evaluate the optimal Q_j when order-up-to levels are fixed.

5.2.3 Finding the Optimal Aggregate (Q, S) Policy

In Theorem 26 and Theorem 28 we show that cost for each retailer is convex in either order-up-to levels S_{ij} or the aggregate quantity Q_j, although it might not be jointly convex. However, since we show in Theorem 27 that the optimal order-up-to level $S^*_{ij}(Q_j)$ is non-decreasing, an iterative search algorithm can be used to find the optimal policy for each retailer with complexity $O(Q_0)$, where Q_0 is the maximum quantity each truck can ship based on its capacity constraint.

Algorithm: Decentralized Retailer Algorithm

Initialize: $S_{ij} \leftarrow 0$, $Q \leftarrow 1$, $C^* \leftarrow \infty$, $Q^* \leftarrow 1$, $S^*_{ij} \leftarrow 0$, $\underline{S_{ij}} \leftarrow 0$;

while $Q \leq Q_0$ **do**
- **forall** $i \in S$ **do**
  - $S_{ij} \leftarrow \underline{S_{ij}}$;
  - $M_{cur} \leftarrow M_i(S_{ij})$;
  - **while** $M_{cur} < \frac{p_{ij}}{p_{ij}+h_{ij}}$ **do**
    - $S_{ij} \leftarrow S_{ij} + 1$;
    - $M_{cur} \leftarrow M_{cur} + m_i(S_{ij})$;
  - **end**
  - $\underline{S_{ij}} \leftarrow S_{ij}$;
- **end**
- $C_{tmp} = C_j(Q_j, S_{i1}, \cdots, S_{nj})$;
- **if** $C_{tmp} < C^*$ **then**
  - $C^* \leftarrow C_{tmp}$;
  - $S^*_{ij} \leftarrow S_{ij}$;
  - $Q^* \leftarrow Q$;
- **end**
- $Q \leftarrow Q + 1$;

end

We iteratively find the optimal $S^*_{ij}(Q_j)$ for each Q_j. In the process, we dynamically update $\underline{S_{ij}}$, the lower bound for $S^*_{ij}(Q_j)$, to reduce the search space. By monotonicity in Theorem 27, we need to evaluate cost $C_j(Q_j, S_{1j}, \cdots, S_{nj})$ at most Q_0 times.

5.3 Supplier Policy

In this section, we characterize a decentralized policy for suppliers. As with the retailer policy, we propose a (Q, S) policy for suppliers. At the warehouse, however, suppliers replenish their products individually, and thus the (Q, S) policy is equivalent to a typical (s, S) policy. More specifically, whenever the installation inventory position of product i falls to or below $(S_i - Q_i)$ at the warehouse, we place an order to raise the inventory position to S_i.

From the analysis in Section 5.2, we know orders from each retailer follow a binomial distribution with Erlang time intervals. However, suppliers can only observe historical orders, and in this setting, combine orders from different retailers, ignoring which retailer places each order. Hence, in making replenishment decisions, each supplier only needs to estimate two parameters: the interarrival times of retailers' orders, and the size of each order.

5.3.1 Demand Approximation

We first characterize the distribution of size of each order that each supplier faces at warehouse. The order process from each retailer follows an independent Erlang process, and hence the long run arrival rate of aggregate orders from all retailers at warehouse is:

$$\lambda_0 \triangleq \sum_{j \in R} \frac{\lambda_j}{Q_j}.$$

Thus, the long run the probability that an order at the warehouse comes from retailer j is:

$$P_j(\boldsymbol{Q_j}) = \frac{\lambda_j / Q_j}{\lambda_0}.$$

Then, we obtain the distribution of order size:

$$OS(x) = Prob(order\ size\ = x) = \sum_{j \in R} P_j(\boldsymbol{Q_j}) \cdot \binom{Q_j}{x} \theta_{ij}^x (1 - \theta_{ij})^{Q_j - x}. \quad (5.4)$$

The expected of order size is:

$$\mathbb{E}(OS) = \sum_{j \in R} \frac{\lambda_j / Q_j}{\lambda_0} \cdot Q_j \cdot \frac{\lambda_{ij}}{\lambda_j} = \sum_{j \in R} \frac{\lambda_{ij}}{\lambda_0}. \quad (5.5)$$

The variance of order size is:

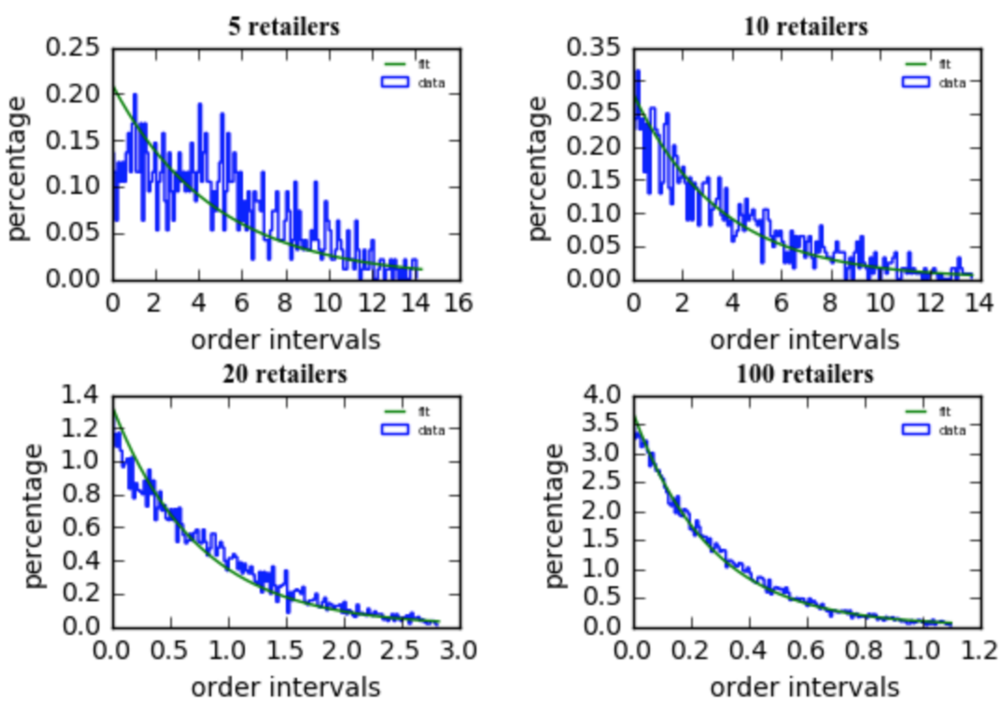
$$Var(OS) = \sum_{j \in R} \left(\frac{\lambda_j / Q_j}{\lambda_0} \right)^2 \cdot Q_j \cdot \frac{\lambda_{ij}}{\lambda_j} (1 - \frac{\lambda_{ij}}{\lambda_j}) = \sum_{j \in R} \frac{\lambda_{ij}(\lambda_j - \lambda_{ij})}{\lambda_0^2 Q_j}. \quad (5.6)$$

The demand size at the warehouse for each supplier actually follows a compound Binomial distribution, but it is complicated to estimate $\boldsymbol{(Q, S, \lambda)}$ individually since these are based on the private information of each retailer. In addition, it is not necessary to do so since the supplier does not care which specific retailer places an order. Typically, the number of retailers in this setting is large enough that suppliers can use the sample distribution from historical orders as a good approximation.

Next we analyze the interarrival times of orders from all retailers. We have calculated the rate of aggregate orders from all retailers at the warehouse: λ_0. We also know that under a $(Q, \boldsymbol{S})$ policy, the order interval from each retailer follows an Erlang distribution, and orders from different retailers are independent.

From the literature, we know that the superposition of independent renewal processes can be asymptotically well-approximated by a Poisson process Teresalam and Lehoczky (1991). As we observe in Figure 5.2, when the number of retailers is relatively large (hundreds of retailers are common in the settings that motivated this work), the arrivals of orders at the warehouse can in fact be well-approximated by a Poisson process. Hence, each supplier can estimate the demand arrival rate using the average interarrival time.

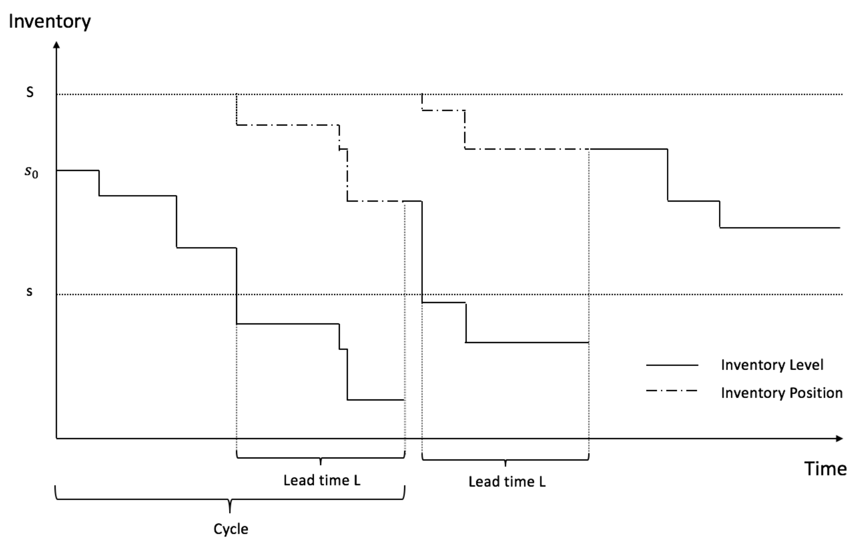
Figure 5.2: Distribution of Aggregate Orders at Warehouse



As we calculated above, λ_0 is the rate for the approximate Poisson demand process. For each supplier, this rate can be obtained from historical ordering data, without requiring detailed information from each retailer. Thus, in the following derivation, we use the compound Poisson process to approximate orders from all retailers at warehouse.

To evaluate the long run cost for each supplier at the warehouse, we introduce a Markov chain. We define a cycle as the time between two consecutive arrivals of replenishments at the warehouse, as illustrated in Figure 5.3. We first evaluate average cycle cost conditioning on cycle starting inventory s_0, we then characterize the stationary distribution of cycles with different starting inventory, and finally we take the weighted average to calculate the long run average cost.

Figure 5.3: Installation (s, S) policy for suppliers



5.3.2 Demand During Lead Time

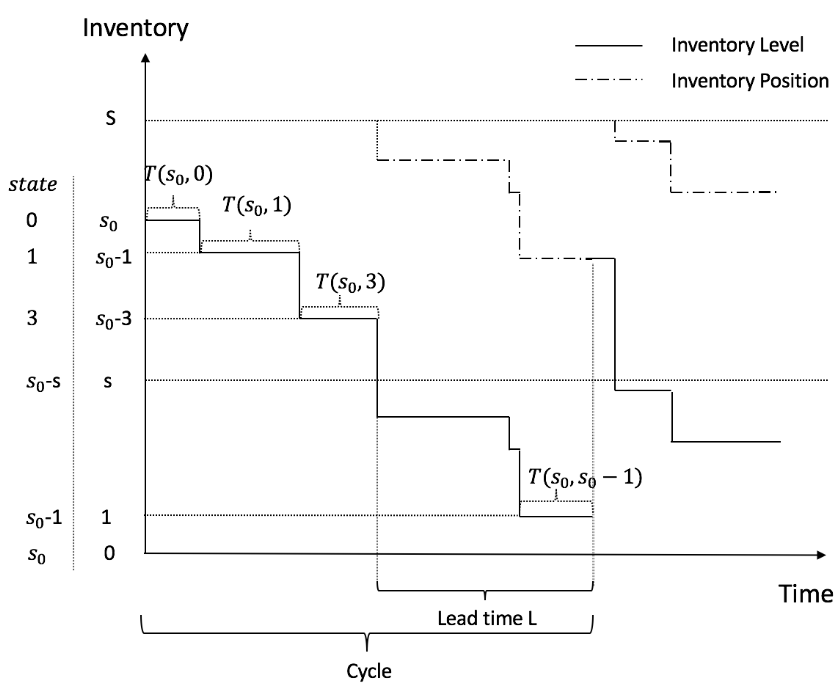
As we discuss above, aggregate orders from different retailers can be viewed as a compound Poisson process at the warehouse, where the rate and distribution of demand size can be obtained from historical data. Beckman proved that an (s, S) policy is optimal for such a system if unmet demand is backordered Beckmann (1961). However, in our model, suppliers must meet retailers' orders and we assume there is a penalty cost associated with each unmet demand (e.g., a cost to expedite delivery from other sources), so the model is actually equivalent to a lost sales model.

In our model, each supplier follows an installation (s, S) policy. That is, each supplier places an order to raise warehouse inventory position to S whenever its warehouse inventory position drops to or below s. For mathematical convenience, we also assume there is at most one replenishment on delivery, e.g., $s < S - s$.. This is the same simplifying assumption as in Archibald and Silver (1978), which only slightly changes the result but significantly simplifies analysis.

We introduce more notation in this cost evaluation, as illustrated in Figure 5.4. Since we evaluate cost for each supplier separately, subscript i for supplier is omitted from now on.

- p_i: probability of demand size i, $i = 0, 1, \cdots, D$, where D is the largest demand size (notice that demand size can be 0 when the retailer only orders other products)
- s_0: the starting inventory position (also inventory level) of the cycle
- $T(s_0, m)$: the long run average time spent in state m when the cycle starting inventory level is s_0. We label inventory s_0 state 0 (so the same inventory level can be different states in different cycles when s_0 is different)
- $P(s_0, m)$: the probability that state m is ever visited in a cycle with starting inventory s_0
- $L(s_0)$: the long run average cycle length with starting inventory s_0
- $D(k)$: the probability that the total units of the product required by all retailers during lead time equals k

Figure 5.4: Transition states of inventory levels



Since we define cycle starting inventory to be s_0, we know

$$P(s_0, 0) = 1. \tag{5.7}$$

When a replenishment has not been triggered, so $k = 1, 2, \cdots, s_0 - s + 1$, and the next order from a retailer occurs in the same cycle, we have the following transition function:

$$\begin{aligned} P(s_0, k) &= \frac{p_1}{1-p_0} \cdot P(s_0, k-1) + \cdots + \frac{p_k}{1-p_0} \cdot P(s_0, 0) \\ &= \frac{1}{1-p_0} \sum_{j=1}^{min(D,k)} p_j \cdot P(s_0, k-j) \end{aligned} \tag{5.8}$$

where $\frac{p_i}{1-p_0}$ is the conditional probability when demand is strictly positive. Furthermore, once a state is visited, the number of periods the system stays in the same state follows a geometric distribution, and therefore

$$T(s_0, k) = \frac{P(s_0, k)}{\lambda_0 \cdot (1 - p_0)}. \tag{5.9}$$

Combining (5.8) and (5.9), we have

$$T(s_0, k) \cdot \lambda_0 \cdot (1 - p_0) = \sum_{i=0}^{k-1} p_i \cdot T(s_0, i) \cdot \lambda_0.$$

That is,

$$\begin{aligned} T(s_0, 0) &= \frac{1}{\lambda_0(1-p_0)} \\ T(s_0, k) &= \sum_{i=0}^{k-1} \frac{p_i}{1-p_0} T(s_0, i). \end{aligned} \tag{5.10}$$

Therefore the average cycle length is the sum of lead time and the time before a replenishment is triggered. That is,

$$L(s_0) = L + \sum_{i=0}^{s_0-s-1} T(s_0, i). \tag{5.11}$$

Next we characterize the distribution of total demand during leadtime. As a compound Poisson process, the total demand during a given time can be calculated by probability generating function (PGF). By extending Adelson's result for PGF of compound

Poisson process Adelson (1966), we know that the PGF for the lead time demand is:

$$PGF(\text{lead time demand}) = exp\Big(- \sum_{i=0}^{D} \lambda_0 L p_i \Big) \cdot exp\Big(\sum_{i=0}^{D} \lambda_0 L p_i x^i \Big). \tag{5.12}$$

Thus

$$\begin{aligned} D(k+1) &= \frac{\lambda_0 L}{k+1}\Big(p_1 D(k) + 2p_2 D(k-1) + \cdots + (k+1)p_{k+1} D(0)\Big) \\ &= \frac{\lambda_0 L}{k+1}\Big(\sum_{j=1}^{min(D,k+1)} j \cdot p_j \cdot D(k+1-j)\Big) \end{aligned} \tag{5.13}$$

where

$$D(0) = exp\Big(- \sum_{i=1}^{D} \lambda_0 L p_i \Big) = exp(- \lambda_0 L(1 - p_0)). \tag{5.14}$$

Then, according to (5.13) and (5.14), we can recursively calculate the distribution of lead time demand.

5.3.3 Inventory Holding Cost at the Warehouse

In this subsection, we calculate the average inventory holding in a cycle with starting inventory s_0. Before a replenishment is triggered (that is when $k > s_0 - s$), $T(s_0, k)$ characterizes the average time in state k in the current cycle. Thus, the average inventory cost incurred before replenishment is:

$$\sum_{i=0}^{s_0-s-1} T(s_0, i) \cdot (s_0 - i). \tag{5.15}$$

If the lead time is infinity, the average inventory holding cost incurred during lead time is:

$$\sum_{i=s_0-s}^{s_0} T(s_0, i) \cdot (s_0 - i). \tag{5.16}$$

However, in our case with finite lead time L, (5.16) over-calculates the inventory holding cost when cycle ending inventory is positive. If we denote average on hand inventory after a lead time L bys:

$$W(k) = \sum_{i=0}^{k-1} (k-i) D(i) \tag{5.17}$$

the average inventory holding cost after L is:

$$\begin{aligned}
&\sum_{i=s_0-s}^{s_0} \frac{P(s_0,i)\sum_{j=0}^{s_0-i}(s_0-i-j)\cdot D(j)}{(1-p_0)\lambda_0} \\
&= \sum_{i=s_0-s}^{s_0} T(s_0,i)\sum_{j=0}^{s_0-i}(s_0-i-j)\cdot D(j) \\
&= \sum_{i=s_0-s}^{s_0} T(s_0,i)\cdot W(s_0-i). \qquad (5.18)
\end{aligned}$$

Combining (5.15), (5.16) and (5.18), we obtain the inventory holding cost for the cycle:

$$H(s_0) \triangleq \sum_{i=0}^{s_0} T(s_0,i)\cdot(s_0-i) - \sum_{i=s_0-s}^{s_0} T(s_0,i)\sum_{j=0}^{s_0-i}(s_0-i-j)\cdot D(j). \qquad (5.19)$$

5.3.4 Lost Sales Cost at the Warehouse

In this subsection, we calculate the penalty cost for suppliers when their inventory at the warehouse is insufficient to cover orders from retailers. Since suppliers are required to fulfill the unmet demand from other sources, the penalty cost can be viewed equivalently as a lost sales cost. To simply calculation, we denote the average demand size of orders from retailers:

$$\overline{d} = \sum_{i=1}^{D} i\cdot p_i.$$

Thus the average inventory level by the end of the cycle is:

$$s_0 - \lambda_0 L\overline{d} - \Big(\sum_{i=0}^{s_0-s-1} T(s_0,i)\Big)\lambda_0\overline{d}. \qquad (5.20)$$

Since suppliers must satisfy all orders from retailers, so warehouse inventory cannot be negative, (5.20) is actually the sum of two components: average inventory on hand (always nonnegative) and average lost sales. Inventory on hand can be calculated by the convolution of the last inventory position before the replenishment trigger point and the demand during the lead time:

$$\sum_{i=0}^{s_0-s-1} P(s_0,i)\sum_{j=s_0-s-i}^{s_0-i} p_j\cdot W(s_0-i-j) = \sum_{i=0}^{s_0-s-1} P(s_0,i)\sum_{j=s_0-s-i}^{min(D,s_0-i)} p_j\cdot W(s_0-i-j). \qquad (5.21)$$

Therefore the average lost sales in a cycle is:

$$LS(s_0) \triangleq -s_0 + \lambda_0 L\bar{d} + \Big(\sum_{i=0}^{s_0-s-1} T(s_0,i) \Big) \lambda_0 \bar{d} + \sum_{i=0}^{s_0-s-1} P(s_0,i) \sum_{j=s_0-s-i}^{min(D,s_0-i)} p_j \cdot W(s_0-i-j). \tag{5.22}$$

5.3.5 Long Run Average Cost

In each cycle with starting inventory s_0, we know by renewal theory that the average cost can be obtained as follows:

$$c(s_0,s) = \frac{k_i + p_i^0 \cdot LS(s_0) + h_i \cdot H(s_0)}{L(s_0)}. \tag{5.23}$$

Thus, we only need to determine the probability of different cycle starting inventories. The weighted average of cycle cost is the long run average cost. The starting inventory of the next cycle only depends on the starting inventory of current cycle. That is, the cycle starting inventory is Markovian. It is easy to see that this system if positive recurrent and thus this stationary distribution exists. Next, we characterize the transition probabilities. We introduce more notation to simplify derivation:

- $P_{ij}(s,S)$: the transition probability of starting inventory state i to starting inventory state j when an (s,S) policy is implemented (where states defined as before)
- $f_k(s_0)$: the probability that state k is the first state immediately after a replenishment is made, in a cycle of starting inventory s_0

$f_k(s_0)$ can be calculated by conditioning on the last state before a replenishment is triggered:

$$f_k(s_0) = \begin{cases} \sum_{m=0}^{s_0-s-1} P(s_0,m)\frac{p_{k-m}}{1-p_0}, & \forall k = s_0-s,\cdots,s_0-1 \\ 1 - \sum_{m=s_0-s}^{s_0-1} f_m(s_0), & k = s_0. \end{cases} \tag{5.24}$$

The starting inventory of the next cycle only depends on $f_k(s_0)$ and the demand during the lead time. Then, we can calculate the transition matrix as follows. If $j = S$, then either no demand occurs during lead time, or the inventory has droped to 0 when the replenishment is made (so that under a lost sales model the inventory level stays the same until the next cycle);if $s < j < S$, we also consider two cases depending on whether the inventory levels drops to 0 during lead time. Therefore

$$P_{ij}(s,S) = \begin{cases} \sum_{k=i-s}^{i-(S-j)-1} f_k(i) \cdot D(S-j) + f_{i-(S-j)}(i) \cdot (1 - \sum_{k=0}^{S-j-1} D(k)), & j \in \{S-s, \cdots, S-1\} \\ & \textit{and } i \in \{s+1, \cdots, S\}, \\ 1 - \sum_{k=i-s}^{i-1} f_k(i) + D(0) \cdot (1 - f_i(i)) = f_i(i) + D(0) \cdot (1 - f_i(i)), & i \in \{s+1, \cdots, S\}, j = S \\ 0, & \textit{otherwise.} \end{cases} \tag{5.25}$$

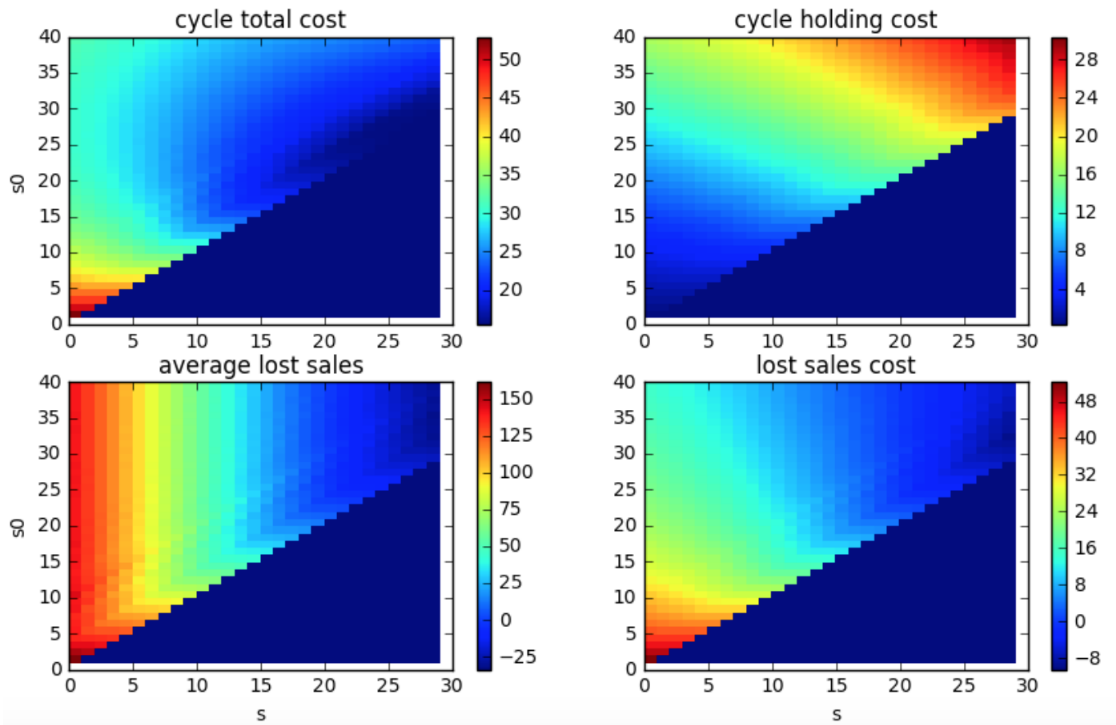
Finally, using transition matrix $\{P_{ij}(s,S)\}$ in the continuous Markov Chain, we obtain the stationary distribution $\pi_i(s,S)$. The long run average cost is therefore

$$C(s,S) = \frac{\sum_{i=s+1}^{S} \pi_i(s,S) \cdot c(i,s,S) \cdot L(s_0)}{\sum_{i=s+1}^{S} \pi_i(s,S) \cdot L(s_0)}. \tag{5.26}$$

5.3.6 Analysis of $c(s_0, s)$

To evaluate long run average cost for the system, we need to calculate $c(s_0, s)$ and the corresponding transition matrix, which can be extremely time consuming. Thus we first analyze $c(s_0, s)$. By the definition of an (s,S) policy, we know that $s < s_0 \leq S$. In Figure 5.5 we plot the holding cost, inventory-deficiency cost, and total cost, when the replenishment trigger point s and cycle start inventory s_0 vary.

Figure 5.5: Cost for $c(s_0, s)$

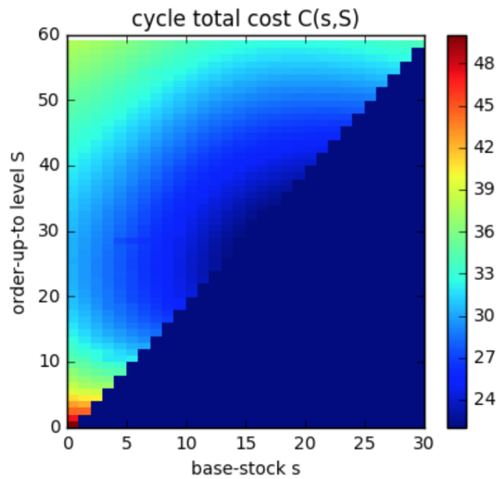


As we see from Figure 5.5, the total cost $c(s_0, s)$ appears to be monotonic and convex in (s_0, s), although of course this might not hold true for the general case, depending on cost parameters. When the order-up-to level S is fixed, s_0 almost determines the cycle length as well as the average lost sales, because lead time is fixed after the replenishment is triggered. Thus, when average cost is taken as the ratio of the two, the lost sales cost appears (numerically) to be monotonic in s and s_0. On the other hand, the holding cost depends on the average inventory on hand as well as the cycle length. The larger $s_0 - s$ is, the longer the cycle length will be. Thus, we can search to find the optimal (s_0^*, s^*). Notice that cycle cost $c(s_0, s)$ does not depend on S, so s_0^* and s^* are optimal for all S.

5.3.7 Optimizing Exact $C(s, S)$

Now that we have completed the evaluation of each $c(s_0, s)$, we can combine them to obtain $C(s, S)$. An exact solution can be found by calculating all $c(s_0, s)$ and then applying (5.26). We assume that there exists at most one outside order, to that $s < S - s$, and conduct a numerical study to evaluate $C(s, S)$ for different combinations of basestock s and order-up-to level S.

Figure 5.6: Exact cost for $C(s, S)$



As we observe in Figure 5.6, the total cost for (s, S) policy appears numerically to be convex. Thus, we can efficiently search to find the optimal s^* and S^*.

5.3.8 An Approximation of $C(s, S)$

In the exact method of $C(s, S)$ evaluation detailed in the previous subsection, it is very time consuming to calculate the stationary distribution of s_0 using (5.25).

Thus, we also propose the following heuristic to evaluate $C(s, S)$: After an order is placed, the time to next cycle is fixed to the lead time L, and the current inventory position is S. Thus, the starting inventory s_0 of the next cycle depends on how many units of demand are consumed from the current inventory. Notice that this is not equivalent to the demand that occurs during the lead time, because when inventory drops to 0, all subsequent demand is satisfied from an outside source, and thus does not affect inventory position at the warehouse. Furthermore, by our assumption that $s < S - s$, the starting inventory position is at least $S - s$. Hence, we use the truncated lead time demand to replace actual inventory level change. That is,

$$C(s, S) \approx \tilde{C}(s, S) = \frac{\sum_{i=S-s}^{S} D(S-i) \cdot c(i, s) \cdot L(i)}{\sum_{i=S-s}^{S} D(S-i) \cdot L(i)}. \tag{5.27}$$

The loss due to this approximation is minor. As we observe from Figure 5.7, the total cost approximation is almost the same as the exact solution, but it is much faster and easier to calculate. We can also search a to find s^* and S^* based on this approximation.

Figure 5.7: Cost for $C(s,S)$ and $\tilde{C}(s,S)$

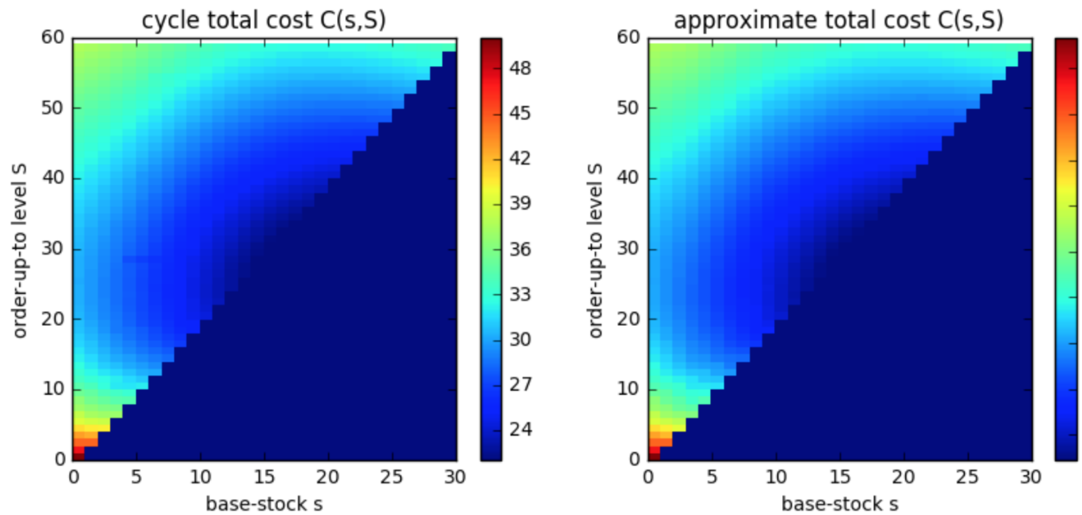
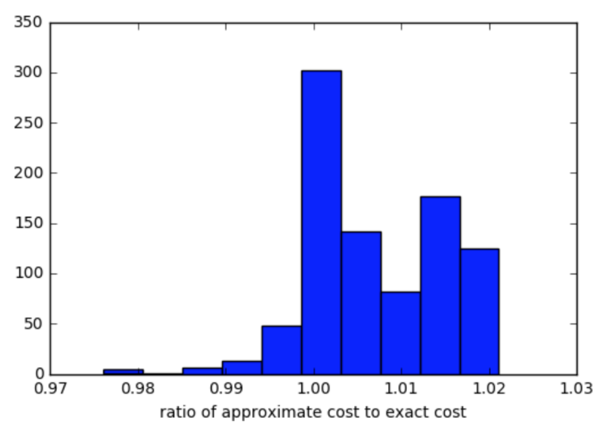


Figure 5.8: Histogram of cost ratio $\frac{\tilde{C}(s,S)}{C(s,S)}$



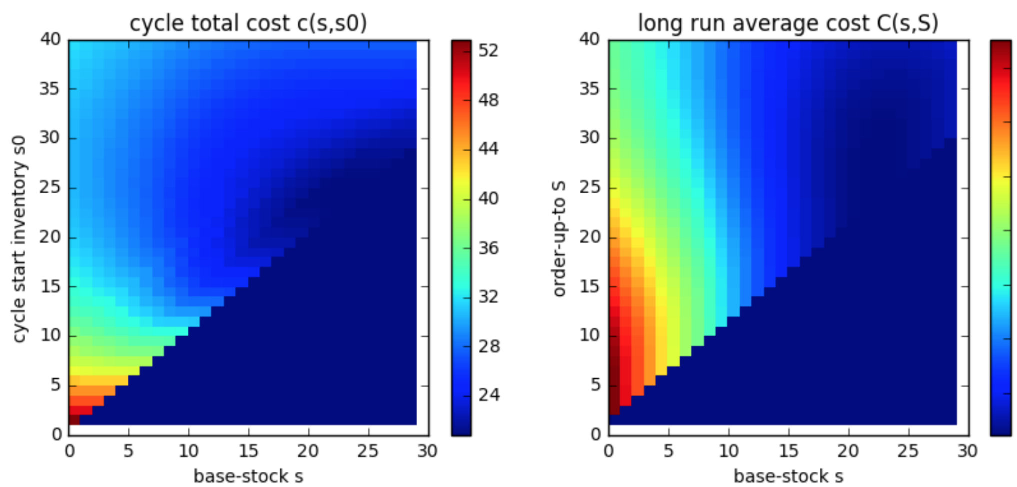
5.3.9 Approximation of s^* and S^*

Alternatively, we can adopt a heuristic heuristic from Archibald (1981) to solve for s^* and S^*. Recall that we have already solved for $s*$ and s_0^* in the last subsection. If we compare $c(s_0, s)$ and $C(s,S)$ in Figure 5.9, $C(s,S)$ is a linear combination of all feasible $c(s_0, s)$ over different s_0, where the coefficients are determined by S. $c(s_0, s)$ appears to be relatively robust near (s_0^*, s^*), and thus we can approximate S^* as follows:

$$\tilde{S}^* = \lceil s_0^* + \bar{d}\lambda L - LS(s_0^*) \rceil, \tag{5.28}$$

where $\bar{d}\lambda L$ is the average demand occured during lead time, and $\bar{d}\lambda L - LS(s_0^*)$ is the average inventory decrease during lead time.

Figure 5.9: Comparison of cycle cost and long run cost



5.4 Conclusion

In this chapter, we consider a stochastic one warehouse multi-supplier multi-retailer problem operated in a decentralized fashion. We assume independent Poisson demand for each product occurs at each retailer. To coordinate replenishment, each retailer follows an aggregate (Q, S) policy, i.e., an order is placed to raise inventory position to S whenever total demand since the last order at that retailer reaches Q. In this setting, demand at the warehouse can be well-approximated by a compound Poisson process, and thus inventory at the warehouse is managed via an (s, S) policy. We develop optimal and heuristic algorithms to optimize parameters for these policies in this model.

Our ultimate goal is to better understand how to effectively operate collaborative logistics systems in settings similar to the one analyzed in this paper. Thus, we are currently working to extend these models to more general operating policies.

Chapter 6

Conclusions

In this dissertation, we analyze decentralized collaboration among multiple suppliers and multiple retailers. A central warehouse serves as both the outbound warehouse for suppliers and inbound warehouse for retailers to motivate collaboration. Suppliers send their products for different retailers to the warehouse first, and thge warehouse will send products from different suppliers to the same retailer in the same delivery. In Chapter 3 and Chapter 4, we modeled setting models in both uncapacitated and truck-capacitated deterministic versions, where suppliers and retailers each make their own decision. By comparing centralized lower bounds and decentralized policies, we showed that decentralized policies are easy to implement in practice, and can protect information privacy, while the loss due to decentralization is minor. Thus it is unnecessary to make a huge effort to centralize these system, and this decentralized mode of operation is consistent with our observation of many 3PL companies.

We also considered a stochastic version of the problem in Chapter 5, where independent Poisson demand occurs at different retailers for different products. To coordinate replenishment, each retailer follows an aggregate (Q,S) policy, i.e., an order is placed to raise inventory position to S whenever total demand since the last order at that retailer reaches Q. In this setting, demand at the warehouse can be well-approximated by a compound Poisson process, and thus inventory at the warehouse is managed via an (s,S) policy. We develop optimal and heuristic algorithms to optimize parameter settings in this model.

Bibliography

Abdul-Jalbar, B., Gutierrez, J., Puerto, J., and Sicilia, J. (2003). Policies for inventory/distribution systems: the effect of centralization vs. decentralization. *International Journal of Production Economics*, 81:281–293.

Adelson, R. (1966). Compound poisson distributions. *OR*, 17(1):73–75.

Archibald, B. C. (1981). Continuous review (s, s) policies with lost sales. *Management Science*, 27(10):1171–1177.

Archibald, B. C. and Silver, E. A. (1978). (s, s) policies under continuous review and discrete compound poisson demand. *Management Science*, 24(9):899–909.

Arkin, E., Joneja, D., and Roundy, R. (1989). Computational complexity of uncapacitated multi-echelon production planning problems. *Operations Research Letters*, 8(2):61–66.

Armstrong Associates, I. (2016). The business of warehousing in north america ? 2015 market size, major 3pls, benchmarking costs, prices and practices.

Aucamp, D. C. (1982). Nonlinear freight costs in the eoq problem. *European Journal of Operational Research*, 9(1):61–63.

Axsäter, S. (1990). Simple solution procedures for a class of two-echelon inventory problems. *Operations Research*, 38(1):64–69.

Axsäter, S. (2000). Exact analysis of continuous review (r, q) policies in two-echelon inventory systems with compound poisson demand. *Operations research*, 48(5):686–696.

Baboli, A., Neghab, M. P., and Haji, R. (2008). An algorithm for the determination of the economic order quantity in a two-level supply chain with transportation costs: Comparison of decentralized with centralized decision. *Journal of Systems Science and Systems Engineering*, 17(3):353–366.

Ballot, E. and Fontane, F. (2010). Reducing transportation co2 emissions through pooling of supply networks: perspectives from a case study in french retail chains. *Production Planning & Control*, 21(6):640–650.

Beckmann, M. (1961). An inventory model for arbitrary interval and quantity distributions of demand. *Management Science*, 8(1):35–57.

Benavides, Luis, V. D. E. and Swan, D. (2012). Six steps to successful supply chain collaboration. `http://www.supplychainquarterly.com/topics/Strategy/20120622-six-steps-to-successful-supply-chain-collaboration/`.

Bertsekas, D. P. (1999). *Nonlinear programming*.

Bijvank, M. and Vis, I. F. (2011). Lost-sales inventory theory: A review. *European Journal of Operational Research*, 215(1):1–13.

Burns, L. D., Hall, R. W., Blumenfeld, D. E., and Daganzo, C. F. (1985). Distribution strategies that minimize transportation and inventory costs. *Operations Research*, 33(3):469–490.

Cassidy, W. B. (2011). Hershey, Ferrero Sign Supply Chain Pact. joc.com. `http://www.joc.com/international-logistics/distribution-centers/hershey-ferrero-sign-supply-chain-pact_20111005.html`.

Chan, L. M. A., Muriel, A., Shen, Z.-J. M., Simchi-Levi, D., and Teo, C.-P. (2002). Effective zero-inventory-ordering policies for the single-warehouse multiretailer problem with piecewise linear cost structures. *Management Science*, 48(11):1446–1460.

Chen, F. (2000). Optimal policies for multi-echelon inventory problems with batch ordering. *Operations research*, 48(3):376–389.

Chen, F., Federgruen, A., and Zheng, Y.-S. (2001a). Coordination mechanisms for a distribution system with one supplier and multiple retailers. *Management Science*, 47(5):693–708.

Chen, F., Federgruen, A., and Zheng, Y.-S. (2001b). Near-optimal pricing and replenishment strategies for a retail/distribution system. *Operations Research*, 49(6):839–853.

Chen, J.-M. and Chen, T.-H. (2005). The multi-item replenishment problem in a two-echelon supply chain: the effect of centralization versus decentralization. *Computers & Operations Research*, 32(12):3191–3207.

Chu, C.-L. and Leon, V. J. (2008). Power-of-two single-warehouse multi-buyer inventory coordination with private information. *International Journal of Production Economics*, 111(2):562–574.

CO3 (2011). JSP and HF-Czechforge to bundle plastics and steel shipments from czech republic to germany. `http://www.co3-project.eu/wo3/wp-content/uploads/2011/12/JSP-Hammerwerk-CO3-Case-study.pdf`.

Daganzo, C. F. (1988a). A comparison of in-vehicle and out-of-vehicle freight consolidation strategies. *Transportation Research Part B: Methodological*, 22(3):173–180.

Daganzo, C. F. (1988b). Shipment composition enhancement at a consolidation center. *Transportation Research Part B: Methodological*, 22(2):103–124.

De Bodt, M. A. and Graves, S. C. (1985). Continuous-review policies for a multi-echelon inventory problem with stochastic demand. *Management Science*, 31(10):1286–1299.

Federgruen, A. and Zheng, Y.-S. (1992). An efficient algorithm for computing an optimal (r, q) policy in continuous review stochastic inventory systems. *Operations research*, 40(4):808–813.

Federgruen, A. and Zheng, Y.-S. (1993). Optimal power-of-two replenishment strategies in capacitated general production/distribution networks. *Management Science*, 39(6):710–727.

Federgruen, A. and Zipkin, P. (1984). An efficient algorithm for computing optimal (s, s) policies. *Operations research*, 32(6):1268–1285.

Gallego, G. and Simchi-Levi, D. (1990). On the effectiveness of direct shipping strategy for the one-warehouse multi-retailer r-systems. *Management Science*, 36(2):240–243.

Goyal, S. and Belton, A. (1979). On" a simple method of determining order quantities in joint replenishments under deterministic demand. *Management Science*, pages 604–604.

Gue, K., E. A. A. E. W. F. and Forger., G. (2014). Material handling and logistics us roadmap. *MHI*.

Hadley, G. and Whitin, T. M. (1963). *Analysis of inventory systems*. Prentice Hall.

Hong, X., Chunyuan, W., Xu, L., and Diabat, A. (2016). Multiple-vendor, multiple-retailer based vendor-managed inventory. *Annals of Operations Research*, 238(1-2):277–297.

Jackson, P., Maxwell, W., and Muckstadt, J. (1985). The joint replenishment problem with a powers-of-two restriction. *IIE transactions*, 17(1):25–32.

Jackson, P. L., Maxwell, W. L., and Muckstadt, J. A. (1988). Determining optimal reorder intervals in capacitated production-distribution systems. *Management Science*, 34(8):938–958.

Jin, Y. and Muriel, A. (2009). Single-warehouse multi-retailer inventory systems with full truckload shipments. *Naval Research Logistics (NRL)*, 56(5):450–464.

KANE Is Able, Inc. (2011). Sun-Maid Uses Consolidation to Drive a 62% Reduction in Outbound Freight Costs. http://www.kaneisable.com/sites/default/files/Sunmaid-Case-Study-2011.pdf.

Kaspi, M. and Rosenblatt, M. J. (1991). On the economic ordering quantity for jointly replenished items. *The International Journal of Production Research*, 29(1):107–114.

Khouja, M. and Goyal, S. (2008). A review of the joint replenishment problem literature: 1989–2005. *European Journal of Operational Research*, 186(1):1–16.

Konur, D. (2014). Carbon constrained integrated inventory control and truckload transportation with heterogeneous freight trucks. *International Journal of Production Economics*, 153:268–279.

Maxwell, W. L. and Muckstadt, J. A. (1985). Establishing consistent and realistic reorder intervals in production-distribution systems. *Operations Research*, 33(6):1316–1341.

McKinnon, A. (2010). European Freight Transport Statistics: Limitations, Misinterpretations and Aspirations. `http://www.acea.be/uploads/publications/SAG_15_European_Freight_Transport_Statistics.pdf`.

Meall, L. (2010). The road to sustainability. `http://www.fsn.co.uk/channel_kpi_environment/the_road_to_sustainability`.

Mitchell, J. S. (1987). 98%-effective lot-sizing for one-warehouse, multi-retailer inventory systems with backlogging. *Operations Research*, 35(3):399–404.

Muckstadt, J. A. and Roundy, R. O. (1987). Multi-item, one-warehouse, multi-retailer distribution systems. *Management Science*, 33(12):1613–1621.

Özer, Ö. and Xiong, H. (2008). Stock positioning and performance estimation for distribution systems with service constraints. *IIE Transactions*, 40(12):1141–1157.

Rieksts, B. Q. and Ventura, J. A. (2008). Optimal inventory policies with two modes of freight transportation. *European Journal of Operational Research*, 186(2):576–585.

Rong, Y., Atan, Z., and Snyder, L. V. (2011). Heuristics for base-stock levels in multi-echelon distribution networks. Technical report, Working Paper. Available at ssrn:¡ http://ssrn. com/abstract= 1475469¿. Last checked on February 10.

Roundy, R. (1985). 98%-effective integer-ratio lot-sizing for one-warehouse multi-retailer systems. *Management science*, 31(11):1416–1430.

Roundy, R. (1986). A 98%-effective lot-sizing rule for a multi-product, multi-stage production/inventory system. *Mathematics of operations research*, 11(4):699–727.

Ryder System, I. (2014). Ryder's Warehouse Network: Mixing and Consolidation Centers (MACC). `http://www.ryder.com/supply-chain/solutions-by-capability/warehousing-and-distribution/macc.aspx/`.

Silver, E. A. (1976). A simple method of determining order quantities in joint replenishments under deterministic demand. *Management Science*, 22(12):1351–1361.

Snyder, L. V. and Shen, Z.-J. M. (2011). *Fundamentals of supply chain theory*. John Wiley & Sons.

Taleizadeh, A. A., Niaki, S. T. A., and Barzinpour, F. (2011). Multiple-buyer multiple-vendor multi-product multi-constraint supply chain problem with stochastic demand and variable lead-time: a harmony search algorithm. *Applied Mathematics and Computation*, 217(22):9234–9253.

Teresalam, C. and Lehoczky, J. P. (1991). Superposition of renewal processes. *Advances in Applied Probability*, 23(01):64–85.

Trunick, P. A. (2011). Colgate logistics delivers smiles. *Inbound Logistics*, 31(5).

US Department of Transportation (2014). National transportation statistics 2013.

Viswanathan, S. and Mathur, K. (1997). Integrating routing and inventory decisions in one-warehouse multiretailer multiproduct distribution systems. *Management Science*, 43(3):294–312.

Wagelmans, A., Van Hoesel, S., and Kolen, A. (1992). Economic lot sizing: an o (n log n) algorithm that runs in linear time in the wagner-whitin case. *Operations Research*, 40(1-supplement-1):S145–S156.

Weyl, H. (1910). Über die gibbs? sche erscheinung und verwandte konvergenzphänomene. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 30(1):377–407.

Zheng, Y.-S. (1991). A simple proof for optimality of (s, s) policies in infinite-horizon inventory systems. *Journal of Applied Probability*, 28(04):802–810.

Zheng, Y.-S. and Federgruen, A. (1991). Finding optimal (s, s) policies is about as simple as evaluating a single policy. *Operations research*, 39(4):654–665.

Appendix A

A.1 Proof of Theorem 1

Proof. Proof. Notice $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is the sum of several maximum functions which are strictly convex, thus $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is strictly convex. Therefore the local minimum is also global minimum.

Let $\boldsymbol{T} = (\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ and $\forall$ direction $\boldsymbol{v} = (\boldsymbol{v^s}, \boldsymbol{v^r}) = (v_1^s, \cdots, v_n^s, v_1^r, \cdots, v_m^r)$. $\boldsymbol{T}$ is a local minimum if and only if the directional derivatives are nonnegative, so that:

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} = \lim_{t \to 0^+} \frac{C_{IR}(\boldsymbol{T} + t\boldsymbol{v}) - C_{IR}(\boldsymbol{T})}{t} \geq 0.$$

We do not need to consider the left derivative because:

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}^-} = \frac{\partial C_{IR}}{\partial (-\boldsymbol{v})^+}.$$

$$C_{IR}(\boldsymbol{T}) = \sum_{i \in S} \frac{k_i}{T_i} + \sum_{j \in R} \frac{k_j}{T_j} + \sum_{i \in S} \sum_{j \in R} g_{ij} T_j + \sum_{i \in S} \sum_{j \in L_i \cup G_i} \max(T_i, T_j) g^{ij},$$

so

$$\begin{aligned} \frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \quad = \quad & \sum_{i \in S} -\frac{k_i^s v_i^s}{(T_i^s)^2} + \sum_{j \in R} -\frac{k_j^r v_j^r}{(T_j^r)^2} + \sum_{i \in S} \sum_{j \in L_i} v_i^s g^{ij} + \sum_{i \in S} \sum_{j \in G_i} v_j^r g^{ij} \\ & + \sum_{i \in S} \sum_{j \in E_i} \max(v_i^s, v_j^r) g^{ij} + \sum_{i \in S} \sum_{j \in R} v_j^r g_{ij}. \end{aligned} \qquad \text{(A.1)}$$

We first consider any positive and negative basic directions $(\boldsymbol{v}_+, \boldsymbol{v}_-)$ where given any subset $W \subset S \cup R$:

$$\boldsymbol{v}_+ \colon \begin{cases} v_i^s = 1, & \forall i \in W \\ v_j^r = 1, & \forall j \in W \\ v_i^s = v_j^r = 0, & otherwise. \end{cases} \quad \text{and } \boldsymbol{v}_- \colon \begin{cases} v_i^s = -1, & \forall i \in W \\ v_j^r = -1, & \forall j \in W \\ v_i^s = v_j^r = 0, & otherwise. \end{cases}$$

Hence

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}_+^+} = \sum_{i\in W} -\frac{k_i^s}{(T_i^s)^2} + \sum_{j\in W} -\frac{k_j^r}{(T_j^r)^2} + \sum_{i\in W}\sum_{j\in L_i} g^{ij} + \sum_{j\in W}\sum_{i:j\in G_i} g^{ij} + \sum_{\substack{i\in W \text{ or } j\in W,\\ j\in E_i}} g^{ij} + \sum_{j\in W} g_{ij} \tag{A.2}$$

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}_-^+} = \sum_{i\in W} \frac{k_i^s}{(T_i^s)^2} + \sum_{j\in W} \frac{k_j^r}{(T_j^r)^2} - \sum_{\substack{i\in W \text{ or } j\in W,\\ j\in E_i}} g^{ij} - \sum_{j\in W} g_{ij}. \tag{A.3}$$

If $W = P(U_l^*)$ for some l, then by first order necessary conditions $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} = 0$. Thus from (A.2) and (A.3) we obtain $(C1)$. If W is a subset of $P(U_l^*)$ for some l, then $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \geq 0$. From (A.2) and (A.3) we obtain $(C2)$. Therefore $(C1)$ and $(C2)$ are necessary conditions of optimality. To prove sufficiency, for any non-basic direction $\boldsymbol{v}$, we can decompose the direction $\boldsymbol{v} = b_1\boldsymbol{v_1} + (b_2 - b_1)\boldsymbol{v_2} + ... + (b_p - b_{p-1})\boldsymbol{v_p}$, where $\boldsymbol{v_x}$ are basic directions with their coefficients $v_x^i, v_x^j \in \{0, 1, -1\}$ and $b_1 \leq b_2 \leq \cdots \leq b_p$. Then we observe $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} = b_1 \frac{\partial C_{IR}}{\partial \boldsymbol{v_1}^+} + (b_2 - b_1)\frac{\partial C_{IR}}{\partial \boldsymbol{v_2}^+} + ... + (b_p - b_{p-1})\frac{\partial C_{IR}}{\partial \boldsymbol{v_p}^+}$. That means any direction can be decomposed into a summation of basic directions. Thus if $(C1)$ and $(C2)$ hold, $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \geq 0$, which guarantees optimality. □

A.2 Proof of Lemma 1

Proof. Proof. At the beginning of each round, we pick the largest order interval that can be obtained from the remaining suppliers and retailers. Then we delete that supplier, retailer, or the pair from the problem. By repeating the above procedure, $R \cup S$ is partitioned into several groups, which satisfies the condition $(C1)$ in Theorem 1. Notice that the order interval will increase after grouping, so we can prove optimality by contradiction.

Assume the solution $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ we obtained from MSIRR algorithm is not optimal, then there exist a base direction v such that $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} < 0$. That is, there exists at least one set A of suppliers and retailers that can form a new group with a larger order interval and smaller total cost. A cannot be a single supplier (or a single retailer), otherwise the supplier (retailer) will be selected and form a single cluster by the algorithm. Then we know the new order interval is:

$$T_A = \sqrt{\frac{\sum\limits_{i\in A} k_i^s + \sum\limits_{j\in A} k_j^r}{\sum\limits_{i\in A, j\in L_i} g^{ij} + \sum\limits_{j\in A, i:j\in G_i} g^{ij} + \sum\limits_{j\in A, i\in S} g_{ij} + \sum\limits_{i\in A, j\in A} g^{ij}}}.$$

Since $\max\limits_{l}(\frac{k_l}{h_l then,}) \geq \frac{\sum_i k_i}{\sum_i h_i}$, there exists at least one supplier $\hat{i}$ and one retailer $\hat{j}$

such that

$$T_{\hat{i}\hat{j}} = \sqrt{\frac{k^s_{\hat{i}} + k^r_{\hat{j}}}{\sum_{j\in L_{\hat{i}}} g^{\hat{i}j} + \sum_{i:\hat{j}\in G_i} g^{i\hat{j}} + \sum_{i\in S} g_{i\hat{j}} + g^{\hat{i}\hat{j}}}} > T_A.$$

However, $\hat{i}$ and $\hat{j}$ are grouped by the algorithm, and the grouping is made in the order of T calculated in step 2. Therefore, $T^{s*}_{\hat{i}} = T^{r*}_{\hat{j}} > T_{\hat{i}\hat{j}} > T_A > T^{s*}_{\hat{i}}$ by the assumption of T_A and the observation that T increases after grouping. □

A.3 Proof of Theorem 2

With this rounding rule, we preserve the order of intervals in the solution. That is;

$$T^{s*}_i \geq T^{r*}_j \Rightarrow T^{s*}_{i,P} \geq T^{r*}_{j,P}.$$

PoT rounding leads to a new partition:

$$\frac{U^*_l}{\sqrt{2}} \leq U^{P*}_l < \sqrt{2}U^*_l.$$

While some of the $P(U^*_l)$ may be combined after rounding, for the convenience of comparison, we continue to consider them separately. Hence, the previous partition for $R \cup S$ still applies, so that $P(U^*_l) = P(U^{P*}_l)$. Therefore

$$\begin{aligned}
& C_{POT}(\boldsymbol{T^{s*}_P}, \boldsymbol{T^{r*}_P}) \\
= & \sum_{i\in S} \frac{k^s_i}{T^{s*}_{i,P}} + \sum_{j\in R} \frac{k^r_j}{T^{r*}_{j,P}} + \sum_{i\in S, j\in R} \max(T^{s*}_{i,P}, T^{r*}_{j,P}) \cdot g^{ij} + \sum_{i\in S, j\in R} T^{r*}_{j,P} \cdot g_{ij} \\
= & \sum_{i\in S} \frac{k^s_i}{T^{s*}_{i,P}} + \sum_{j\in R} \frac{k^r_j}{T^{r*}_{j,P}} + \sum_{i\in S, j\in L_i\cup E_i} T^{s*}_{i,P} \cdot g^{ij} + \sum_{j\in R, i:j\in G_i} T^{r*}_{j,P} \cdot g^{ij} + \sum_{i\in S, j\in R} T^{r*}_{j,P} \cdot g_{ij} \\
= & \sum_{l\in\{1,\cdots,k\}} \Big(\frac{K(U^{P*}_l)}{U^{P*}_l} + H(U^{P*}_l) \cdot U^{P*}_l \Big).
\end{aligned}$$

Following the standard approach, for each $P(U^*_l)$ we can apply the PoT optimality bound:

$$\begin{aligned}
C^{U^*_l}_{IRS}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) &= \frac{K(U^{P*}_l)}{U^{P*}_l} + H(U^{P*}_l)) \cdot U^{P*}_l \\
&< C^{\sqrt{2}U^*_l}_{IRS}(\sqrt{2}\boldsymbol{T^{s*}}, \sqrt{2}\boldsymbol{T^{r*}}) \\
&= C^{\frac{U_{l*}}{\sqrt{2}}}_{IRS}(\frac{\boldsymbol{T^{s*}}}{\sqrt{2}}, \frac{\boldsymbol{T^{r*}}}{\sqrt{2}}).
\end{aligned}$$

Therefore

$$C_{IRS}^{U_l^{P*}}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) < \frac{1}{2}\Big(C_{IRS}^{\sqrt{2}U_l^*}(\sqrt{2}\boldsymbol{T^{s*}}, \sqrt{2}\boldsymbol{T^{r*}}) + C_{IRS}^{\frac{U_l^*}{\sqrt{2}}}(\frac{\boldsymbol{T^{s*}}}{\sqrt{2}}, \frac{\boldsymbol{T^{r*}}}{\sqrt{2}})\Big)$$

$$= \frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2})C_{IRS}^{U_l^*}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}).$$

Summing over the $C_{IRS}^{U_l^{P*}}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}})$, we get the standard Power-of-Two performance bound: $\frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2})$.

A.4 Proof of Theorem 3

Proof. Proof. We use a similar methodology to Roundy's approach (Roundy, 1985) to show $C_{IR}(\boldsymbol{T^{s*}, T^{r*}})$ is a cost lower bound. We start by creating exactly the same model except that cost g^{ij} and g^{ij} are replaced by η_{ij} and η^{ij}. We observe that the total cost of an arbitrary policy can be decomposed into several EOQ-like problems. Next we carefully select alternative parameters η_{ij} and η^{ij} to satisfy the following conditions:

(C3) The sum of optimal decomposed costs with alternative parameters is the same as $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$

(C4) Any feasible policy costs no more with alternative cost parameters η_{ij} and η^{ij} than with original parameters g^{ij} and g^{ij}.

We introduce additional notation for this proof:

- η_{ij} and η^{ij}: alternative cost parameters (the original parameters are g_{ij} and g^{ij})
- $\eta_i = \sum_{j \in R} \eta^{ij}$: total cost associated with supplier i
- $\eta_j = \sum_{i \in S} \eta_{ij}$: total cost associated with retailer j
- $n_i(t)$: the number of orders placed by warehouse to supplier i in $[0, t)$
- $n_j(t)$: the number of orders placed by retailer j to warehouse in $[0, t)$
- $I_{ij}(t)$: the inventory of product i at retailer j at time t
- $S_{ij}(t) \geq I_{ij}(t)$: the inventory of product i at retailer j or at warehouse but will be sent to retailer j at time t
- $I_i(t) = \frac{1}{\eta_i} \sum\limits_{j \in R} \eta^{ij} S_{ij}(t)$: weighted average inventory of product i at time t, weighted by the associated alternative cost over all retailers
- $I_j(t) = \frac{1}{\eta_j} \sum\limits_{i \in S} \eta_{ij} I_{ij}(t)$: weighted average inventory at retailer j, weighted by the associated cost over all products.

Therefore for an arbitrary policy, the total cost in $[0, t)$ can be decomposed:

$$\sum_{i\in S} n_i(t)k_i^s + \sum_{j\in R} n_j(t)k_j^r + \sum_{j\in R}\sum_{i\in S}\int_0^{t'} (\eta_{ij} I_{ij}(t) + \eta^{ij} S_{ij}(t))dt$$
$$= \sum_{i\in S}\left(n_i(t)k_i^s + \int_0^{t'} \eta_i I_i(t)dt\right) + \sum_{j\in R}\left(n_j(t)k_j^r + \int_0^{t'} \eta_j I_j(t)dt\right). \quad \text{(A.4)}$$

$I_i(t)$ and $I_j(t)$ are different from system inventory since they are weighted by cost. However, $I_i(t)$ is right continuous in t, decreases linearly with constant slope, and jumps only when the warehouse replenishes inventory of product i. Similarly $I_j(t)$ also decreases linearly with constant slope and jumps whenever retailer j places an order. Hence $I_j(t)$ and $I_i(t)$ functions as inventory does in EOQ models. Therefore the sum of optimal decomposed costs is a lower bound of cost of arbitrary policy.

To satisfy $(C3)$, we need the sum of optimal EOQ costs equal to $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. By (3.3) and (A.4), $(C3)$ is equivalent to:

$$\sum_{i\in S} 2\sqrt{k_i^s \eta_i} + \sum_{j\in R} 2\sqrt{k_j^r \eta_j} = \sum_{l\in\{1,\cdots,k\}} 2\sqrt{K(U_l^*)\cdot H(U_l^*)} = 2\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + 2\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}., \quad \text{(A.5)}$$

By the first order condition of the decomposed costs, it is sufficient to show $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ are the optimal solutions to the decomposed EOQ-like problems, that is

$$\begin{cases} \sum_{i\in S} \eta_{ij} = \eta_j = \frac{k_j^r}{(T_j^{r*})^2}, & \forall j \in R \\ \sum_{j\in R} \eta^{ij} = \eta_i = \frac{k_i^s}{(T_i^{s*})^2}, & \forall i \in S. \end{cases}. \quad \text{(A.6)}$$

To satisfy $(C5)$ we need

$$\sum_{j\in R}\sum_{i\in S}\int_0^{t'} (\eta_{ij} I_{ij}(t) + \eta^{ij} S_{ij}(t))dt \le \sum_{j\in R}\sum_{i\in S}\int_0^{t'} (g_{ij} I_{ij}(t) + g^{ij} S_{ij}(t))dt. \quad \text{(A.7)}$$

Since $I_{ij}(t) \le S_{ij}(t)$, we only need

$$\begin{cases} \eta_{ij} \ge g_{ij}, & \forall i \in S, j \in R \\ \eta^{ij} + \eta_{ij} = g_{ij} + g^{ij}, & \forall i \in S, j \in R \\ \eta^{ij}, \eta_{ij} \ge 0, & \forall i \in S, j \in R \end{cases} \quad \text{(A.8)}$$

Now we have shown (A.6) and (A.8) are sufficient conditions for $(C3)$ and $(C4)$. It remains to show that there exist alternative parameters η_{ij} and η^{ij} satisfying (A.6) and (A.8). In the following proof we use the subdifferential to show the existence of η_{ij} and η^{ij}.

Claim 1. *There exist alternative cost parameters satisfying (A.6) and (A.8).*

Proof. Proof. η_{ij} and η^{ij} defined in (A.9) satisfy (A.8) for all $0 \le A_{ij} \le g^{ij}$:

$$\eta_{ij} = \begin{cases} g_{ij} & \text{if } T_j^{r*} < T_i^{s*} \\ g_{ij} + g^{ij} & \text{if } T_j^{r*} > T_i^{s*} \\ g_{ij} + A_{ij} & \text{if } T_j^{r*} = T_i^{s*} \end{cases} \text{ and } \eta^{ij} = \begin{cases} g^{ij} & \text{if } T_j^{r*} < T_i^{s*} \\ 0 & \text{if } T_j^{r*} > T_i^{s*} \\ g^{ij} - A_{ij} & \text{if } T_j^{r*} = T_i^{s*} \end{cases} \tag{A.9}$$

We determine appropriate A_{ij} to satisfy (A.6) later.

By Danskin's theorem, the subdifferential of a finite pointwise maximum function is the convex hull of the subdifferential of corresponding active functions (Bertsekas, 1999). Hence

$$\partial(\max(T_i^s, T_j^r)) = \begin{cases} \partial T_j^r, & \text{if } T_i^s < T_j^r \\ \partial T_i^s, & \text{if } T_i^s > T_j^r \\ \alpha_{ij} \cdot \partial T_i^s + (1-\alpha_{ij}) \cdot \partial T_j^r, & \text{if } T_i^s = T_j^r \end{cases}$$

Therefore the subdifferential of $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is

$$\begin{aligned} \partial C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r}) = & \sum_{i \in S} \frac{-k_i^s}{(T_i^s)^2} \partial T_i^s + \sum_{j \in R} \frac{-k_j^r}{(T_j^r)^2} \partial T_j^r + \sum_{i \in S} \sum_{j \in L_i} g^{ij} \partial T_i^s + \sum_{i \in S} \sum_{j \in G_i} g^{ij} \partial T_j^r \\ & + \sum_{i \in S} \sum_{j \in E_i} g^{ij} \Big(\alpha_{ij} \cdot \partial T_i^s + (1-\alpha_{ij}) \cdot \partial T_j^r \Big) + \sum_{i \in S} \sum_{j \in R} g_{ij} \partial T_j^r. \end{aligned} \tag{A.10}$$

Since $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is the optimal solution to $(\boldsymbol{PIR})$, which is convex, we know $\mathbf{0} \in \partial C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$.

We substitute 0 for each entry in $\partial C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$, and by definition there exists $\alpha_{ij} \in [0,1]$ such that

$$\begin{cases} -\frac{k_i}{(T_i^*)^2} + \sum\limits_{j \in L_i} g^{ij} + \sum\limits_{j \in E_i} \alpha_{ij} g^{ij} = 0, & \forall i \in S \\ -\frac{c_j}{(T_j^*)^2} + \sum\limits_{i: j \in G_i} g^{ij} + \sum\limits_{i: j \in E_i} (1-\alpha_{ij}) g^{ij} + \sum\limits_{i \in S} g_{ij} = 0, & \forall j \in R \end{cases} \tag{A.11}$$

We let $A_{ij} = (1-\alpha_{ij}) g^{ij}$ and substitute into (A.11), then we get all the equations in (A.6). Therefore Claim 1 is proved. □

In summary, combining (A.4), (A.7) and (A.5), we know:

$$
\begin{aligned}
&\sum_{i\in S} k_i^s n_i(t') + \sum_{j\in R} k_j^r n_j(t') + \sum_{j\in R}\sum_{i\in S}\int_0^{t'} (g_{ij} I_{ij}(t) + g^{ij} S_{ij}(t))dt \\
\geq &\sum_{i\in S}(\frac{k_i^s}{T_i^s} + \eta_i T_i^s) + \sum_{j\in R}(\frac{k_j^r}{T_j^r} + \eta_j T_j^r) \\
\geq &C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}).
\end{aligned}
$$

□

A.5 Proof of Theorem 4

Proof. Proof. Compared to the optimal order interval in the centralized model,

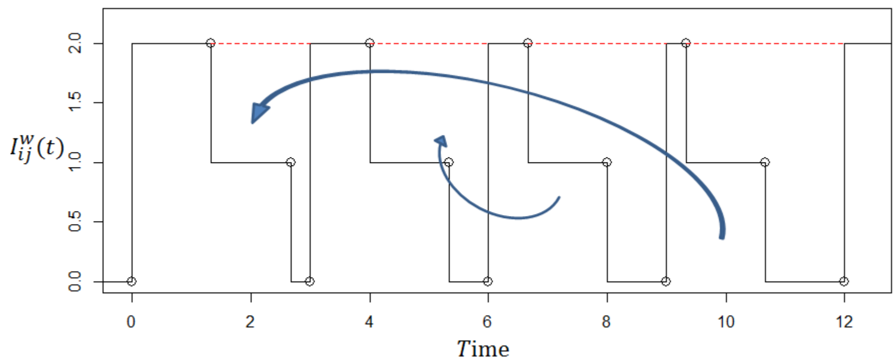
$$
\begin{aligned}
\Gamma_j^{r*} &= \sqrt{\frac{k_j^r}{\sum\limits_{i\in S}(g_{ij} + g^{ij})}} \\
&\leq \sqrt{\frac{k_j^r}{\sum\limits_{i\in S} g_{ij} + \sum\limits_{i:j\in L_i\cup E_i} g^{ij}}} \leq T_j^{r*}.
\end{aligned}
$$

□

A.6 Proof of Theorem 5

Proof. Proof. In this proof, we use $\{x\}$ to denote fractional part of x. We first consider the case when $a_{ij} = \frac{p_{ij}}{q_{ij}}$ is rational. Because the set $\{n\frac{\Gamma_i^s}{\Gamma_j^{r*}}\}, \{(n+1)\frac{\Gamma_i^s}{\Gamma_j^{r*}}\}, \cdots, \{(n + q_{ij} - 1)\frac{\Gamma_i^s}{\Gamma_j^{r*}}\}$ is equivalent to the set of $0, \frac{1}{q_{ij}}, \frac{2}{q_{ij}}, \cdots, \frac{q_{ij}-1}{q}$, $\forall n \in \mathbb{N}$, $I_{ij}^w(t)$ is cyclic with cycle length $q_{ij}\Gamma_i^s$. Due to symmetry, we can calculate the long run average inventory $\mathbb{A}(I_{ij}^w)$ by "moving and adding" inventory levels for every q_{ij} consecutive order cycles of supplier i to form "blocks", as illustrated in Figure A.1:

Figure A.1: $\Gamma_i^s > \Gamma_j^{r*}$, inventory from supplier i to retailer j



Next we need to determine the fraction of blocks that are of height $b_{ij}d_{ij}\Gamma_j^{r*}$. Notice that $\{n\frac{\Gamma_i^s}{\Gamma_j^{r*}}\} \cdot \Gamma_j^{r*}$ is the time between the n^{th} replenishment from supplier i and retailer j's last order until then, assuming the 0^{th} order is placed at time 0. Therefore, if $\frac{1}{q_{ij}} \le \{n\frac{\Gamma_j^{r*}}{\Gamma_i^s}\} \le \frac{p_{ij}-1}{q_{ij}}$, then in the n^{th} order cycle of supplier i, (between the n^{th} and $(n-1)^{th}$ order of supplier i), supplier i needs to raise inventory to $(b_{ij}+1)\Gamma_j^{r*}d_{ij}$. Otherwise it only raises the inventory level to $b_{ij}\Gamma_j^{r*}d_{ij}$. Notice that when $\{n\frac{\Gamma_j^{r*}}{\Gamma_i^s}\} = 0$, even though retailer j places $b_{ij}+1$ orders, supplier i only needs to raise inventory to $b_{ij}\Gamma_j^{r*}d_{ij}$. This is because one of retailer j's order is placed simultaneously with the replenishment of supplier i. Consequently,

$$\begin{aligned} Hij^l(\Gamma_i^s, \Gamma_j^{r*}) &= \frac{1}{2}\left((b_{ij}+1)\frac{p_{ij}-1}{q_{ij}} + b_{ij}\frac{q_{ij}+1-p_{ij}}{q_{ij}}\right)\Gamma_j h^i d_{ij} \\ &= \frac{1}{2}\left(b_{ij} + \frac{p_{ij}}{q_{ij}} - \frac{1}{q_{ij}}\right)\Gamma_j h^i d_{ij} \\ &= \frac{1}{2}(\frac{\Gamma_i^s}{\Gamma_j^r} - \frac{1}{q_{ij}})\Gamma_j h^i d_{ij} \\ &= (\Gamma_i^s - \frac{\Gamma_j^r}{q_{ij}})g^{ij} \end{aligned}$$

For the case of irrational b_{ij}, we let $q_{ij} \to \infty$ and get $Hij^l(\Gamma_i^s, \Gamma_j^{r*}) = \Gamma_i^s g^{ij}$ □

A.7 Proof of Lemma 2

Proof. We first claim the following technical result:

Claim 2. *Let $\{x\}$ return the fractional part of x. $\forall a > 0, \forall b \in (0,1)$ and $N \in \mathbb{N}$, we consider two cases. If a is irrational, then*

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}(\{na\} \in [b, b+\epsilon]) \xrightarrow{N\to\infty} \epsilon,\ \forall b\ s.t.\ b+\epsilon \leq 1.$$

Otherwise $a = \frac{p}{q}$, where p and q are coprime, then $\{na\}$ only takes value in $0, \frac{1}{q}, \cdots, \frac{q-1}{q}$, and

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}(\{na\} = \frac{k}{q}) \xrightarrow{N\to\infty} \frac{1}{q},\ k = 0, 1, \cdots, q-1.$$

Claim 2 is a corollary of a classical result, Weyl's Equidistribution Theorem (see, e.g., (Weyl, 1910)). Then we know

$$\Delta_{ij}(k) = \{\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\} \cdot \Gamma_i^s.$$

This is because the k^{th} order of retailer j occurs at $k\Gamma_j^{r*}$, and supplier i has placed $\lfloor\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\rfloor$ orders up to then. Thus, the last order from supplier i is placed at $\lfloor\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\rfloor \cdot \Gamma_i^s$. The time difference is therefore $k\Gamma_j^{r*} - \lfloor\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\rfloor \cdot \Gamma_i^s = \{\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\} \cdot \Gamma_i^s$.

Then we substitute a by $\frac{\Gamma_j^{r*}}{\Gamma_i^s}$ in Claim 2. If $\frac{\Gamma_j^{r*}}{\Gamma_i^s}$ is rational, we have

$$\frac{1}{N}\sum_{k=1}^{N}\Delta_{ij}(k) = \frac{\Gamma_i^s}{N}\sum_{k=1}^{N}\{\frac{k\Gamma_j^{r*}}{\Gamma_i^s}\} \xrightarrow{N\to\infty} \frac{\Gamma_i^s}{\tilde{q}_{ij}}\sum_{k=0}^{\tilde{q}_{ij}-1}\frac{k}{\tilde{q}_{ij}} = \frac{\Gamma_i^s(\tilde{q}_{ij}-1)}{2\tilde{q}_{ij}}.$$

On the other hand, $\Delta_{ij}(k) + \overline{\Delta}_{ij}(k) = \Gamma_i^s$ because the sum is the time between last and next order of supplier i. Therefore

$$\frac{1}{N}\sum_{k=1}^{N}\overline{\Delta}_{ij}(k) = \frac{1}{N}\sum_{k=1}^{N}(\Gamma_i^s - \Delta_{ij}(k)) \xrightarrow{N\to\infty} \Gamma_i^s - \frac{\Gamma_i^s(\tilde{q}_{ij}-1)}{2\tilde{q}_{ij}} = \frac{\Gamma_i^s(\tilde{q}_{ij}+1)}{2\tilde{q}_{ij}}.$$

The proof is similar when $\frac{\Gamma_j^{r*}}{\Gamma_i^s}$ is irrational. □

A.8 Proof of Theorem 6

Proof. We first consider the case when b_{ij} is irrational. In each order cycle of retailer j, inventory is held from the last replenishment of supplier i in this cycle, to the end of the cycle when retailer j orders the inventory. The long run average of this holding time is $\frac{\Gamma_i^s}{2}$ (from Lemma 2). The inventory level is $d_{ij}\Gamma_j^r$, the size of each order from retailer j. Hence the long run average total cost in a cycle of retailer j is $\frac{\Gamma_i^s}{2}d_{ij}\cdot\Gamma_j^r\cdot h^i$. Cost per unit time of $I_{ij}^w(t)$ is $\frac{\Gamma_i^s}{2}d_{ij}\cdot\Gamma_j^r\cdot h^i/\Gamma_j^{r*}=g^{ij}\Gamma_i^s$. The case when b_{ij} is rational is similar. In summary,

$$\underline{H}_{ij}^g(\Gamma_i^s,\Gamma_j^{r*})=\begin{cases}\frac{1}{\Gamma_j^{r*}}(\Gamma_j^{r*}h^i d_{ij}\cdot\frac{\tilde{q}_{ij}-1}{2\tilde{q}_{ij}}\Gamma_i^s)=\frac{\tilde{q}_{ij}-1}{\tilde{q}_{ij}}g^{ij}\Gamma_i^s, & if\ \frac{\Gamma_j^{r*}}{\Gamma_i^s}\in\mathbb{Q}\\ \frac{1}{\Gamma_j^{r*}}(\Gamma_j^{r*}h^i d_{ij}\cdot\frac{1}{2}\Gamma_i^s)=g^{ij}\Gamma_i^s, & otherwise\end{cases}.$$

□

A.9 Proof of Theorem 7

To solve ($\mathbf{PDS_i^{zio}}$), we separate the feasible region into two parts, depending on whether or not the supplier i has the same order interval as any retailer. We find the local optimal order interval in each case, and then search for the global optimal solution.

1. $\forall j\in R,\Gamma_i^s\neq\Gamma_j^{r*}$:
   In this case $\sum\limits_{j\in L_i\cup G_i}g^{ij}=\sum\limits_{j\in R}g^{ij}$ is a constant, so

   $$C_i^{zio}(\Gamma_i^s)=\frac{k_i}{\Gamma_i^s}+\sum_{j\in L_i\cup G_i}g^{ij}\Gamma_i^s=\frac{k_i}{\Gamma_i^s}+\sum_{j\in R}g^{ij}\Gamma_i^s. \tag{A.12}$$

   By the first order condition,

   $$\tilde{\Gamma}_i^{s*}=\sqrt{\frac{k_i}{\sum\limits_{j\in R}g^{ij}}}. \tag{A.13}$$

2. $\Gamma_i^s=\Gamma_j^{r*}$ for some $j\in R$:

   In this case $\sum\limits_{l\in L_i\cup G_i}g^{il}=\sum\limits_{l\in R}g^{il}-g^{ij}$, and we know

   $$C_i^{zio}(\Gamma_j^{r*})=\frac{k_i}{\Gamma_j^{r*}}+\sum_{l\in R}g^{il}\Gamma_j^{r*}-g^{ij}\Gamma_j^{r*}.$$

   Compared to (A.12), $C_i^{zio}(\Gamma_i^s)$ decreases by $g^{ij}\Gamma_j^{r*}$ at all Γ_j^{r*}, so Γ_j^{r*} is a local optimum of $C_i^{zio}(\Gamma_i^s)$.

A.10 Proof of Theorem 8

Proof. The proof is also a corollary of the classical Weyl's Equidistribution Theorem (see, e.g., (Weyl, 1910)). We first consider the case where $\tilde{a}_{ij}$ is irrational. In each order cycle of retailer j, inventory is held from the first replenishment of supplier i in this cycle until the end cycle when retailer j orders the inventory. Applying Lemma 2, the average time between retailer j's order to next supplier i's order is $\frac{\Gamma_i^s}{2}$, and this is the average time in retailer j's order cycle that supplier does not have any inventory. Hence the remaining time $\Gamma_j^{r*} - \frac{\Gamma_i^s}{2}$ is the long run average time that supplier i holds inventory in each cycle. The order-up-to level is $d_{ij}\Gamma_j^r$, which is the size of each order placed by retailer j. Hence the long run average total cost in a cycle of retailer j is $(\Gamma_j^{r*} - \frac{\Gamma_i^s}{2})d_{ij} \cdot \Gamma_j^r \cdot h^i$, so that the holding cost per unit time of $I_{ij}^w(t)$ is

$$(\Gamma_j^{r*} - \frac{\Gamma_i^s}{2}) \cdot d_{ij}\Gamma_j^r \cdot h^i/\Gamma_j^r = (\Gamma_j^{r*} - \frac{\Gamma_i^s}{2})d_{ij} \cdot h^i.$$

The case when b_{ij} is rational is similar. In summary,

$$\underline{H}_{ij}^g(\Gamma_i^s, \Gamma_j^{r*}) = \begin{cases} \frac{1}{\Gamma_j^{r*}}(\Gamma_j^{r*} h^i d_{ij} \cdot (\Gamma_j^{r*} - \frac{\tilde{q}_{ij}+1}{2\tilde{q}_{ij}}\Gamma_i^s)) = (2\Gamma_j^{r*} - \frac{\tilde{q}_{ij}+1}{\tilde{q}_{ij}}\Gamma_i^s)g^{ij}, & if\ \frac{\Gamma_j^{r*}}{\Gamma_i^s} \in \mathbb{Q} \\ \frac{1}{\Gamma_j^{r*}}(\Gamma_j^{r*} h^i d_{ij} \cdot (\Gamma_j^{r*} - \frac{1}{2}\Gamma_i^s)) = g^{ij}(2\Gamma_j^{r*} - \Gamma_i^s), & otherwise. \end{cases}.$$

□

A.11 Proof of Lemma 3

Proof. Throughout the proof, we define $\mathbb{A}(x)$ as long run average of x. That is,

$$\mathbb{A}(x(t)) \triangleq \begin{cases} \lim\limits_{T\to\infty} \frac{1}{T}\int_{t=0}^{T} x(t)dt, & \text{if } x(t) \text{ is continuous} \\ \lim\limits_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} x(k), & \text{if } x(k) \text{ is disctete} \end{cases}$$

We do not worry about existence of $\mathbb{A}(\cdot)$ because all inventory levels we consider here are bounded and piecewise continuous.

To evaluate $\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*})$, we use echelon inventory:

$$\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) = h_i \cdot \left(\lim_{T\to\infty}\frac{1}{T}\int_{t=0}^{T} I_{ij}^w(t)dt\right) = h^i \cdot \mathbb{A}(I_{ij}^w(t)) = h^i \cdot \mathbb{A}\Big(EI_{ij}^w(t) - I_{ij}(t)\Big). \tag{A.14}$$

Since the inventory level at retailer j is cyclic, we obtain

$$\mathbb{A}(I_{ij}(t)) = \lim_{T\to\infty}\frac{1}{T}\int_{t=0}^{T} I^{ij}(t)dt = \frac{1}{2}d_{ij}\Gamma_j^{r*}. \tag{A.15}$$

For $\mathbb{A}\Big(EI_{ij}^w(t)\Big)$, we consider two cases:

1. If $\frac{\Gamma_i^s}{\Gamma_j^{r*}}$ is rational:
We first calculate the long run average of starting and ending inventory levels of supplier i's replenishment cycles:

$$\mathbb{A}(EI_{ij}^w(t_0-)) = \lim_{k\to\infty} \frac{1}{K}\sum_{k=1}^{k} EI_{ij}^w(k\Gamma_i^s-),$$
$$\mathbb{A}(EI_{ij}^w(t_0+)) = \lim_{k\to\infty} \frac{1}{K}\sum_{k=1}^{k} EI_{ij}^w(k\Gamma_i^s+),$$

where $t_0 \in \{\Gamma_i^s, 2\Gamma_i^s, \cdots, k\Gamma_i^s, \cdots\}$ is a replenishment time for supplier i. Here t_0+ and t_0- denote the time immediately before and after replenishment. Thus $EI_{ij}^w(t_0+)$ and $EI_{ij}^w(t_0-)$ are starting and ending inventory levels of supplier i's replenishing cycles. Because EI_{ij}^w decreases at a constant rate except at replenishment time, it follows that

$$\mathbb{A}\Big(EI_{ij}^w(t)\Big) = \frac{1}{2}\Big(\mathbb{A}(EI_{ij}^w(t_0-)) + \mathbb{A}(EI_{ij}^w(t_0+))\Big).$$

For the average starting inventory, we claim that immediately after k^{th} replenishment of supplier i, product i's inventory level at retailer j is:

$$I_{ij}(k\Gamma_i^s+) = \Big(\Gamma_j^{r*} - \Gamma_j^{r*}\{\frac{k\Gamma_i^s}{\Gamma_j^{r*}}\}\Big)d_{ij}.$$

This is because $\Gamma_j^{r*}\{\frac{k\Gamma_i^s}{\Gamma_j^{r*}}\}d_{ij}$ is the quantity of product i that has been consumed at retailer j in current cycle of retailer j. Hence $\Big(\Gamma_j^{r*} - \Gamma_j^{r*}\{\frac{k\Gamma_i^s}{\Gamma_j^{r*}}\}\Big)d_{ij}$ is the remaining inventory.

From Claim 2, we derive the long run average of $I_{ij}(t_0+)$ as follows:

$$\mathbb{A}(I_{ij}(t_0+)) = \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K} I_{ij}(k\Gamma_i^s+) = \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\Big(1 - \{\frac{k\Gamma_i^s}{\Gamma_j^{r*}}\}\Big)\Gamma_j^{r*}d_{ij} = \frac{q_{ij}+1}{2q_{ij}}\cdot\Gamma_j^{r*}d_{ij}.$$

Notice that $\mathbb{A}(I_{ij}(t_0+)) \neq \mathbb{A}(I_{ij}(t))$ because t_0 is restricted in supplier i's replenishing time. From the setting of the *order-up-to* policy, we also know

$$I_{ij}^w(t_0+) = (b_{ij} + \mathbb{1}_{a_{ij}\neq 0})\Gamma_j^{r*}d_{ij} = \mathbb{A}(I_{ij}^w(t_0+)).$$

Therefore

$$\mathbb{A}(EI_{ij}^w(t_0+)) = \mathbb{A}(I_{ij}(t_0+)) + \mathbb{A}(I_{ij}^w(t_0+)) = \frac{q_{ij}+1}{2q_{ij}} \cdot \Gamma_j^{r*} d_{ij} + (b_{ij} + \mathbb{1}_{a_{ij}\neq 0})\Gamma_j^{r*} d_{ij}. \tag{A.16}$$

Since product i is consumed at retailer j at a constant rate, in each order cycle of supplier i, echelon inventory level I_{ij}^w drops by the same amount $\Gamma_i^s d_{ij}$. Hence

$$\mathbb{A}(EI_{ij}^w(t_0-)) = \mathbb{A}(EI_{ij}^w(t_0+)) - \Gamma_i^s d_{ij}. \tag{A.17}$$

Combining (A.16) and (A.17), we derive the long run average of $EI_{ij}^w(t)$ as:

$$\begin{aligned}
\mathbb{A}(EI_{ij}^w(t)) &= \frac{1}{2}\Big(\mathbb{A}(EI_{ij}^w(t_0+)) + \mathbb{A}(EI_{ij}^w(t_0-))\Big) \\
&= \frac{q_{ij}+1}{2q_{ij}} \cdot \Gamma_j^{r*} d_{ij} + (b_{ij} + \mathbb{1}_{a_{ij}\neq 0})\Gamma_j^{r*} d_{ij} - \frac{1}{2}\Gamma_i^s d_{ij} \\
&\leq \frac{q_{ij}+1}{2q_{ij}} \cdot \Gamma_j^{r*} d_{ij} + (\frac{\Gamma_i^s}{\Gamma_j^{r*}} - \frac{1}{q_{ij}} + 1)\Gamma_j^{r*} d_{ij} - \frac{1}{2}\Gamma_i^s d_{ij} \qquad \text{(A.18)} \\
&\leq \frac{1}{2}\Gamma_i^s d_{ij} + \frac{3}{2}\Gamma_j^{r*} d_{ij} \qquad \text{(A.19)}
\end{aligned}$$

In the above derivation, (A.18) holds because b_{ij} is the integer part of $\frac{\Gamma_i^s}{\Gamma_j^{r*}}$. Thus if $a_{ij} = 0$, then $b_{ij} = \frac{\Gamma_i^s}{\Gamma_j^{r*}} \leq \frac{\Gamma_i^s}{\Gamma_j^{r*}} - \frac{1}{q_{ij}} + 1$; else $\frac{\Gamma_i^s}{\Gamma_j^{r*}} \geq b_{ij} + \frac{1}{q_{ij}}$.

2. If $\frac{\Gamma_i^s}{\Gamma_j^{r*}}$ is irrational:
   From Lemma 2, we derive $I_{ij}^w(t_0+) = (b_{ij}+1)\Gamma_j^{r*} d_{ij}$ and $\mathbb{A}(I_{ij}(t_0+)) = \frac{1}{2}\Gamma_j^{r*} d_{ij}$. Similar to the previous case,

$$\begin{aligned}
\mathbb{A}(EI_{ij}^w(t)) &= \frac{1}{2}\Big(\mathbb{A}(EI_{ij}^w(t_0+)) + \mathbb{A}(EI_{ij}^w(t_0-))\Big) \\
&= \frac{1}{2} \cdot \Gamma_j^{r*} d_{ij} + (b_{ij}+1)\Gamma_j^{r*} d_{ij} - \frac{1}{2}\Gamma_i^s d_{ij} \\
&\leq \frac{1}{2}\Gamma_i^s d_{ij} + \frac{3}{2}\Gamma_j^{r*} d_{ij} \qquad \text{(A.20)}
\end{aligned}$$

From (A.19) and (A.20),

$$\mathbb{A}(EI_{ij}^w(t)) \leq \frac{1}{2}\Gamma_i^s d_{ij} + \frac{3}{2}\Gamma_j^{r*} d_{ij}. \tag{A.21}$$

Finally, substituting (A.15) and (A.21) into (A.14),

$$\underline{H}_{ij}^l(\Gamma_i^s, \Gamma_j^{r*}) \leq \frac{1}{2}\Gamma_i^s d_{ij} + \frac{3}{2}\Gamma_j^{r*} d_{ij} - \frac{1}{2}\Gamma_j^{r*} d_{ij} = (\Gamma_i^s + 2\Gamma_j^{r*})g^{ij}.$$

□

A.12 Proof of Theorem 9

Proof. We first show that there exist at most one k satisfying $\hat{\Gamma}^{s*}_{i,k} \in (\Gamma^{s*}_k, \Gamma^{s*}_{k+1})$. We assume by contradiction that $\exists k_1 > k_2$ that both satisfy the condition. Then we come to

$$\hat{\Gamma}^{s*}_{i,k_1} > \hat{\Gamma}^{s*}_{i,k_2} = \sqrt{\frac{k^s_i}{\sum_{j=1}^{k_2-1} g^{ij} - \sum_{j=k_2+1}^{m} g^{ij}}} > \sqrt{\frac{k^s_i}{\sum_{j=1}^{k_1-1} g^{ij} - \sum_{j=k_1+1}^{m} g^{ij}}} = \hat{\Gamma}^{s*}_{i,k_1},$$

which is a contradiction.

Next, we claim that if $\hat{\Gamma}^{s*}_{i,k}$ does not exist, $C^{out}_i(\Gamma^s_i)$ is lower bounded by one of the two end points. This is because $C^{out}_i(\Gamma^s_i)$ is piecewise convex, and it has discontinuities at Γ^{r*}_j. We detail our proof below.

If $\sum_{j=1}^{k} g^{ij} - \sum_{j=k+1}^{m} g^{ij} < 0$, i.e. $\hat{\Gamma}^{s*}_{i,k}$ does not exist, then $C^{out}_i(\Gamma^s_i)$ is decreasing on $(\Gamma^{s*}_k, \Gamma^{s*}_{k+1}]$.

If $\hat{\Gamma}^{s*}_{i,k} > \Gamma^{s*}_{k+1}$, from piecewise convexity we know $\forall \Gamma^s_i \in (\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$,

$$\begin{aligned} C^{out}_i(\Gamma^s_i) &> \frac{k^s_i}{\Gamma^{r*}_{k+1}} + \Big(\sum_{j\in L_i} g^{ij} - \sum_{j\in G_i} g^{ij}\Big)\Gamma^{r*}_{k+1} + \sum_{j\in L_i\cup G_i} 2g^{ij}\Gamma^{r*}_j \\ &= C^{out}_i(\Gamma^{r*}_{k+1}) \end{aligned}$$

Similarly If $\hat{\Gamma}^{s*}_{i,k} < \Gamma^{s*}_k$, we conclude $C^{out}_i(\Gamma^s_i) > C^{out}_i(\Gamma^{r*}_k)$ holds for $\forall \Gamma^s_i \in (\Gamma^{r*}_k, \Gamma^{r*}_{k+1})$.

We have partitioned the real line into many open intervals and many isolated points, and analyzed the local optimal solution in each interval. Therefore the globally optimal solution to $C^{out}_i(\Gamma^s_i)$ must be selected from $\boldsymbol{\Gamma^{r*}}$ and $\Gamma^{s*}_{i,k'}$. □

A.13 Proof of Theorem 11

Proof. From Theorem 3, we know that $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on the cost of the centralized model.

$$C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) = \sum_{i\in S} \frac{k^s_i}{T^{s*}_i} + \sum_{j\in R} \frac{k^r_j}{T^{r*}_j} + \sum_{i\in S, j\in R} \max(T^{s*}_i, T^{r*}_j) g^{ij} + \sum_{i\in S, j\in R} T^{r*}_j g_{ij}.$$

In the following we show

$$\frac{3}{2} C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) \geq \sum_{i\in S} C^{zio}_i(\underline{\Gamma}^{s*}_i) + \sum_{j\in R} C^r_j(\Gamma^{r*}_j),$$

where the right hand side is the cost of the optimal ZIO policy we obtained in Section 3.2.

$$\begin{aligned}
\frac{3}{2}C_{IR}(\boldsymbol{T^{s*}},\boldsymbol{T^{r*}}) &\geq \Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}T_j^{r*}g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\} \\
&\quad+\frac{1}{2}\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}\max{(T_i^{s*},T_j^{r*})}g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\} \qquad \text{(A.22)} \\
&=\Big\{\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}T_j^{r*}g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\}+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} \\
&\quad+\Big\{\sum_{i\in S,j\in R}\max{(T_i^{s*},T_j^{r*})}g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\} \qquad \text{(A.23)} \\
&\geq \sum_{j\in R}C_j^r(\Gamma_j^{r*})+\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{i\in S,j\in R}T_i^{s*}g^{ij}\Big\} \qquad \text{(A.24)} \\
&\geq \sum_{i\in S}C_i^{zio}(\underline{\Gamma}_i^{s*})+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.25)}
\end{aligned}$$

In the above derivation, (A.22) is true because of the property of maximum function; (A.23) is true because of convexity in the optimal centralized solution; (A.24) is true since Γ_j^{r*} is the optimal decentralized policy of retailers in Section 4.4.1, and we eliminate positive terms $\sum T_j^{r*}g_{ij}$; (A.25) is true because $\tilde{\Gamma}_i^{s*}$ minimize $C_i^{zio}(\Gamma_i^s)$.

From the analysis above, we know that an upper bound on the decentralized cost is no more than $\frac{3}{2}$ of the lower bound of the centralized cost. Hence, the ratio of decentralized to optimal centralized cost is bounded by $\frac{3}{2}$. □

A.14 Proof of Theorem 12

Proof.

$$
\begin{aligned}
&\frac{5}{2}\cdot C_{IR}(\boldsymbol{T^{s*}},\boldsymbol{T^{r*}})\\
&=\frac{3}{2}\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}\max\left(T_i^{s*},T_j^{r*}\right)g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\}\\
&\quad+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\Big\{\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}T_j^{r*}(g_{ij}+g^{ij})\Big\}+\sum_{i\in S,j\in R}\left(\max\left(T_i^{s*},T_j^{r*}\right)-T_j^{r*}\right)g^{ij} \qquad \text{(A.26)}\\
&\geq\frac{3}{2}\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}\frac{k_j^r}{T_j^{r*}}+\sum_{i\in S,j\in R}\max\left(T_i^{s*},T_j^{r*}\right)g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}g_{ij}\Big\}+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.27)}\\
&=\frac{3}{2}\Big\{\sum_{i\in S,j\in R}2\max\left(T_i^{s*},T_j^{r*}\right)g^{ij}+\sum_{i\in S,j\in R}2T_j^{r*}g_{ij}\Big\}+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.28)}\\
&\geq\Big\{\sum_{i\in S,j\in L_i}(2T_j^{r*}+T_i^{s*})g^{ij}+\sum_{i\in S,j\in G_i}3(T_i^{s*}+(T_j^{r*}-T_i^{s*}))g^{ij}\Big\}+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.29)}\\
&\geq\Big\{\sum_{i\in S,j\in L_i}(2\Gamma_j^{r*}+T_i^{s*})g^{ij}+\sum_{i\in S,j\in G_i}T_i^{s*}g^{ij}+\sum_{i\in S,j\in G_i}2(T_j^{r*}-T_i^{s*})g^{ij}\Big\}+\sum_{i\in S}\frac{k_i^s}{T_i^{s*}}+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.30)}\\
&\geq\sum_{i\in S}C_i^{out}(\Gamma_i^{s*})+\sum_{j\in R}C_j^r(\Gamma_j^{r*}) \qquad \text{(A.31)}
\end{aligned}
$$

In the above derivation, (A.26) is true due to rearrangement of the maximum function in cost evaluation $C_{IR}(\boldsymbol{T^{s*}},\boldsymbol{T^{r*}})$; (A.27) is true because we eliminate $\sum_{i\in S,j\in R}(\max(T_i^{s*},T_j^{r*})-T_j^{r*})g^{ij}$, and because Γ_j^{r*} is the optimal decentralized policy of retailers in Section 4.4.1; (A.28) is implied by $(C1)$ and first order condition of $C_{IR}(\boldsymbol{T^{s}},\boldsymbol{T^{r}})$; (A.29) is true by the definition of maximum function and G_i, L_i; (A.30) comes from deleting and rearranging terms; (A.31) is true because Γ_i^{s*} is the minimizer of $C_i^{out}(\Gamma_i^{s*})$. □

A.15 Proof of Theorem 13

Proof. Since $\hat{\tau}_i^s$ is the optimal solution to ($\boldsymbol{PDPSR_i}$), we have

$$C_{i,p}^s(\hat{\tau}_i^{s*}) = \frac{k_i^s}{\hat{\tau}_i^{s*}} + \sum_{j:\tau_i^{s*}>\tau_j^{r*}} g^{ij}(\hat{\tau}_i^{s*} - \tau_j^{r*}) < \frac{k_i^s}{\Gamma_i^{s*}} + \sum_{j:\Gamma_i^{s*}>\tau_j^{r*}} g^{ij}\Gamma_i^{s*} \leq C_i^{'}(\tilde{\Gamma}_i^{s*}).$$

We let $\tilde{\Gamma}_{i,p}^{s*} = min\{2^m\Gamma_i^{s*} : 2^m\Gamma_i^{s*} \geq \frac{\Gamma_i^{s*}}{\sqrt{2}}\}$ be PoT rounding of Γ_i^{s*}, then

$$C_{i,p}(\tau_i^{s*}) \leq C_{i,p}^s(\tilde{\Gamma}_{i,p}^{s*}) < C_i^{'}(\tilde{\Gamma}_{i,p}^{s*}) \leq \frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2})C_i^{'}(\tilde{\Gamma}_i^{s*}).$$

□

A.16 Proof of Theorem 14

Proof. We denote $\theta = \frac{1}{2}(\sqrt{2} + \frac{\sqrt{2}}{2})$ for convenience.

$$\begin{aligned}
\frac{3}{2}\theta C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r}) &= \frac{\theta}{2}\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R}\max(T_i^{s*}, T_j^{r*})g^{ij} + \sum_{i\in S, j\in R}T_j^{r*}g_{ij}\Big\} \\
&\quad + \theta\Big(\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R}(\max(T_i^{s*}, T_j^{r*}) - T_j^{r*})g^{ij} + \sum_{i\in S, j\in R}T_j^{r*}(g_{ij} + g^{ij})\Big) && \text{(A.32)} \\
&\geq \frac{\theta}{2}\Big\{\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R}\max(T_i^{s*}, T_j^{r*})g^{ij} + \sum_{i\in S, j\in R}T_j^{r*}g_{ij}\Big\} \\
&\quad + \theta\Big(\sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{i\in S, j\in R}(\max(T_i^{s*}, T_j^{r*}) - T_j^{r*})g^{ij}\Big) + \sum_{j\in R}C_{j,p}^r(\tau_j^{r*}) && \text{(A.33)} \\
&\geq \frac{\theta}{2}\Big\{\sum_{i\in S, j\in R}2\max(T_i^{s*}, T_j^{r*})g^{ij} + \sum_{i\in S, j\in R}2T_j^{r*}g_{ij}\Big\} + \sum_{i\in S}\frac{\theta k_i^s}{T_i^{s*}} + \sum_{j\in R}C_{j,p}^r(\tau_j^{r*}) && \text{(A.34)} \\
&\geq \theta\Big(\sum_{i\in S, j\in R}\max(T_i^{s*}, T_j^{r*})g^{ij} + \sum_{i\in S}\frac{k_i^s}{T_i^{s*}}\Big) + \sum_{j\in R}C_{j,p}^r(\tau_j^{r*}) && \text{(A.35)} \\
&\geq \sum_{i\in S}C_{i,p}^s(\tau_i^{s*}) + \sum_{j\in R}C_{j,p}^r(\tau_j^{r*}) && \text{(A.36)}
\end{aligned}$$

In the above derivation, (A.32) is true by rearrangement of $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$; (A.33) is true by Theorem 10; (A.34) is true because we eliminate $\sum_{i\in S, j\in R}(\max(T_i^{s*}, T_j^{r*}) - T_j^{r*})$ which is nonnegative, and because of first order condition of $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$; (A.35) is true because $\max(T_i^{s*}, T_j^{r*}) \geq T_i^{s*}$ and by Theorem 13.

From the analysis above, we know that the ratio of decentralized to centralized cost is $< \frac{3}{2}\theta$. □

A.17 Proof of Lemma 4

Proof. We define the one-full-truck reorder cycle as $T_j^{r0} = \frac{q}{\sum_{i\in S} d_{ij}}$.

By linear rounding, we know transportation cost is minimized when we deliver full truckload:

$$\frac{c_j}{T_j^r}\lceil\frac{\sum_{i\in S} d_{ij}T_j^r}{q}\rceil \geq \frac{c_j}{T_j^r}\frac{\sum_{i\in S} d_{ij}T_j^r}{q} = \frac{c_j\sum_{i\in S} d_{ij}}{q}$$

Holding cost is non-decreasing in T_j^r:

$$\frac{1}{2}\sum_{i\in S, j\in R}\max(T_i^s, T_j^r)d_{ij}h^i + \frac{1}{2}\sum_{i\in S, j\in R} T_j^r d_{ij}h'_{ij}$$

Therefore if $T_j^r > \frac{q}{\sum_{i\in S} d_{ij}}$, that is, retailer j orders more than one truckload, both holding cost and truck transportation cost will be higher than if $T_j^r = T_j^{r0}$. □

A.18 Proof of Lemma 5

Proof. By the property of ($\mathbf{PC_l}$), Slater's condition holds thus strong duality holds. By KKT conditions we know,

$$\begin{cases} 0 \in \partial C_{IRS}^{U_l^*} + \lambda_l - \mu_l \\ \lambda_l \cdot (U_l^* - T_j^{r0}) = 0 \\ U_l^* - T_j^{r0} \leq 0 \\ \mu_l \cdot U_l^* = 0 \\ \lambda_l, \mu_l \geq 0 \end{cases} \tag{A.37}$$

It is natural to see $U_l^* > 0$, hence (A.37) is equivalent to

$$\begin{cases} 0 \in \frac{-K(U_l^*)}{(U_l^*)^2} + H(U_l^*) + \lambda_l \\ \lambda_l \cdot (U_l^* - T_j^{r0}) = 0 \\ U_l^* - T_j^{r0} \leq 0 \\ \lambda_l \geq 0 \end{cases}$$

Therefore we obtain the optimal solution to ($\mathbf{PC_l}$) as:

$$U_l^* = \min_{j \in G(U_l)} \left(\sqrt{\frac{K(U_l)}{H(U_l)}}, T_j^{r0} \right).$$

□

A.19 Proof of Theorem 15

Proof. Notice $C_C(\boldsymbol{T^s}, \boldsymbol{T^r})$ is the sum of several maximum functions which are strictly convex, thus $C_C(\boldsymbol{T^s}, \boldsymbol{T^r})$ is strictly convex. The feasible region of $C_C(\boldsymbol{T^s}, \boldsymbol{T^r})$ is also convex, therefore the local minimum is global minimum.

Let $\boldsymbol{T} = (\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ and $\forall$ direction $\boldsymbol{v} = (\boldsymbol{v^s}, \boldsymbol{v^r}) = (v_1^s, \cdots, v_n^s, v_1^r, \cdots, v_m^r)$. $\boldsymbol{T}$ is a local minimum if and only if all the feasible directional derivatives is nonnegative, so that:

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} = \lim_{t \to 0^+} \frac{C_{IR}(\boldsymbol{T} + t\boldsymbol{v}) - C_{IR}(\boldsymbol{T})}{t} \geq 0.$$

We do not need to consider left derivative because:

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}^-} = \frac{\partial C_{IR}}{\partial (-\boldsymbol{v})^+}.$$

So

$$\begin{aligned} \frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \quad = \quad & \sum_{i \in S} -\frac{k_i^s v_i^s}{(T_i^s)^2} + \sum_{j \in R} -\frac{c_j^r v_j^r}{(T_j^r)^2} + \sum_{i \in S} \sum_{j \in L_i} v_i^s g^{ij} + \sum_{i \in S} \sum_{j \in G_i} v_j^r g^{ij} \\ & + \sum_{i \in S} \sum_{j \in E_i} \max(v_i^s, v_j^r) g^{ij} + \sum_{i \in S} \sum_{j \in R} v_j^r g_{ij} \qquad \text{(A.38)} \end{aligned}$$

We first consider any positive and negative basic direction $\boldsymbol{v}_+$ $\boldsymbol{v}_-$ where given any subset $W \subset S \cup R$:

$$\boldsymbol{v}_{+}\colon \begin{cases} v_i^s = 1, & \forall i \in W \\ v_j^r = 1, & \forall j \in W \\ v_i^s = v_j^r = 0, & otherwise. \end{cases} \quad \text{and } \boldsymbol{v}_{-}\colon \begin{cases} v_i^s = -1, & \forall i \in W \\ v_j^r = -1, & \forall j \in W \\ v_i^s = v_j^r = 0, & otherwise. \end{cases}$$

Hence

$$\begin{aligned} \frac{\partial C_{IR}}{\partial \boldsymbol{v}_+^+} &= \sum_{i\in W} -\frac{k_i^s}{(T_i^s)^2} + \sum_{j\in W} -\frac{c_j^r}{(T_j^r)^2} + \sum_{i\in W}\sum_{j\in L_i} g^{ij} + \sum_{j\in W}\sum_{i:j\in G_i} g^{ij} \\ &+ \sum_{\substack{i\in W \text{ or } j\in W, \\ j\in E_i}} g^{ij} + \sum_{j\in W} g_{ij} \end{aligned} \tag{A.39}$$

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}_-^+} = \sum_{i\in W} \frac{k_i^s}{(T_i^s)^2} + \sum_{j\in W} \frac{c_j^r}{(T_j^r)^2} - \sum_{\substack{i\in W \text{ or } j\in W, \\ j\in E_i}} g^{ij} - \sum_{j\in W} g_{ij} \tag{A.40}$$

A direction $\boldsymbol{v}$ is feasible at $\boldsymbol{T}$ if and only if $\boldsymbol{T} + t\boldsymbol{v}$ feasible for some $t > 0$. That means in basic directions, $v_j^r = 1$ only when $T_j^r < T_j^{r0}$.

If $W = P(U_l^*)$ for some l, then form KKT condition we obtain $(C1)$. Therefore $(C1)$ and $(C2)$ are necessary conditions of optimality. If W is a subset of $P(U_l^*)$ for some l, then $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \geq 0$ for all feasible direction. From requirement of feasibility, $U_l^* \leq T_j^{r0}$ for any $j \in P(U_l^*)$. From (A.38) and (A.40) we obtain upper bound and lower bounds on U_l^*. Combining these bounds we obtain $(C2)$. To prove sufficiency, for any non-basic direction, we can decompose the direction

$$\boldsymbol{v} = b_1\boldsymbol{v_1} + (b_2 - b_1)\boldsymbol{v_2} + ... + (b_p - b_{p-1})\boldsymbol{v_p},$$

where $\boldsymbol{v_x}$ are basic directions with their coefficients $v_x^i, v_x^j \in \{0, 1, -1\}$ and $b_1 \leq b_2 \leq \cdots \leq b_p$. Then we observe

$$\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} = b_1\frac{\partial C_{IR}}{\partial \boldsymbol{v_1}^+} + (b_2 - b_1)\frac{\partial C_{IR}}{\partial \boldsymbol{v_2}^+} + ... + (b_p - b_{p-1})\frac{\partial C_{IR}}{\partial \boldsymbol{v_p}^+}.$$

That means any direction can be decomposed into summation of basic directions. Thus if $(C1)$ and $(C2)$ hold, $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} \geq 0$, which guarantees optimality.

□

A.20 Proof of Lemma 6

Proof. At the beginning of each round, we pick the largest order interval that can be obtained from the remaining suppliers and retailers. Then we delete that supplier, retailer, or the pair from the problem. By repeating the above procedure, $R \cup S$

is partitioned into several groups, which satisfies the condition $(C1)$ in Theorem 15. Notice that order interval will increase after grouping, so we can prove optimality by contradiction.

Assume the solution $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ we obtained from MSIRR algorithm is not optimal, then there exist a base direction v such that $\frac{\partial C_{IR}}{\partial \boldsymbol{v}^+} < 0$. That is, there exists at least one set A of suppliers and retailers that can form a new group with a larger order interval and smaller total cost. A cannot be a single supplier (or a single retailer), otherwise the supplier (retailer) will be selected and form a single cluster by the algorithm. Then we know the new order interval is:

$$T_A = \min_{j \in A} \left\{ T_j^{r0}, \sqrt{\frac{\sum\limits_{i \in A} k_i^s + \sum\limits_{j \in A} c_j^r}{\sum\limits_{i \in A, j \in L_i} g^{ij} + \sum\limits_{j \in A, i: j \in G_i} g^{ij} + \sum\limits_{j \in A, i \in S} g_{ij} + \sum\limits_{i \in A, j \in A} g^{ij}}} \right\}.$$

Since $\max\limits_i(\frac{k_i}{h_i}) \geq \frac{\sum_i k_i}{\sum_i h_i}$, there exist at least one supplier $\hat{i}$ and one retailer $\hat{j}$ such that

$$T_{\hat{i}\hat{j}} = \min \left\{ T_{\hat{j}}^{r0}, \sqrt{\frac{k_{\hat{i}}^s + k_{\hat{j}}^r}{\sum\limits_{j \in L_{\hat{i}}} g^{\hat{i}j} + \sum\limits_{i: \hat{j} \in G_i} g^{i\hat{j}} + \sum\limits_{i \in S} g_{i\hat{j}} + g^{\hat{i}\hat{j}}}} \right\} > T_A.$$

However, $\hat{i}$ and $\hat{j}$ are grouped by the algorithm, and the grouping is made in the order of T calculated in step 2. Therefore, $T_{\hat{i}}^{s*} = T_{\hat{j}}^{r*} > T_{\hat{i}\hat{j}} > T_A > T_{\hat{i}}^{s*}$ by the assumption of T_A and the observation that T increases after grouping.

□

A.21 Proof of Theorem 16

Proof. We use a similar methodology to Roundy's approach Roundy (1985) to show $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a cost lower bound. We start by creating exactly the same model except that cost g^{ij} and g^{ij} are replaced by η_{ij} and η^{ij}. We observe that total cost of arbitrary policy can be decomposed into several capacitated EOQ-like problems. Next we carefully select alternative parameters η_{ij} and η^{ij} to satisfy following conditions:

(C3) The sum optimal decomposed costs with alternative parameters is the same as $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$

(C4) Any feasible policy costs no more with alternative cost parameters η_{ij} and η^{ij} than with original parameters g^{ij} and g^{ij}.

We introduce more notation for this proof:

- η_{ij} and η^{ij}: alternative cost parameters (the original parameters are g_{ij} and g^{ij})
- $\eta_i = \sum_{j\in R} \eta^{ij}$: total cost associated with supplier i
- $\eta_j = \sum_{i\in S} \eta_{ij}$: total cost associated with retailer j
- $n_i(t)$: the number of orders placed by warehouse to supplier i in $[0, t)$
- $n_j(t)$: the number of orders placed by retailer j to warehouse in $[0, t)$
- $I_{ij}(t)$: the inventory of product i at retailer j at time t
- $S_{ij}(t) \geq I_{ij}(t)$: the inventory of product i at retailer j or at warehouse but will be sent to retailer j at time t
- $I_i(t) = \frac{1}{\eta_i} \sum\limits_{j\in R} \eta^{ij} S_{ij}(t)$: weighted average inventory of product i at time t, weighted by the associated alternative cost over all retailers
- $I_j(t) = \frac{1}{\eta_j} \sum\limits_{i\in S} \eta_{ij} I_{ij}(t)$: weighted average inventory at retailer j, weighted by the associated cost over all products.

Therefore for arbitrary policy, the total cost in $[0, t)$ can be decomposed:

$$\begin{aligned} &\sum_{i\in S} n_i(t)k_i^s + \sum_{j\in R} n_j(t)c_j^r + \sum_{j\in R}\sum_{i\in S}\int_0^{t'} (\eta_{ij} I_{ij}(t) + \eta^{ij} S_{ij}(t))dt \\ =&\sum_{i\in S}\left(n_i(t)k_i^s + \int_0^{t'} \eta_i I_i(t)dt\right) + \sum_{j\in R}\left(n_j(t)c_j^r + \int_0^{t'} \eta_j I_j(t)dt\right) \end{aligned} \tag{A.41}$$

$I_i(t)$ and $I_j(t)$ are different from system inventory since they are weighted by cost. However, $I_i(t)$ is right continuous in t, decreases linearly with constant slope, and jumps only when warehouse replenishes inventory of product i. Similarly $I_j(t)$ also decreases linearly with constant slope and jumps whenever retailer j places an order. Hence $I_j(t)$ and $I_i(t)$ works exactly the same as inventory of capacitated EOQ models. The optimal decomposed cost is therefore

$$\frac{k_i}{\hat{T}_i} + \eta_i \hat{T}_i \text{ and } \frac{k_j}{\hat{T}_j} + \eta_j \hat{T}_j,$$

where $\hat{T}_i = \sqrt{\frac{k_i}{\eta_i}}$ and $\hat{T}_j = min\{T_j^0, \sqrt{\frac{k_j}{\eta_j}}\}$.

To satisfy $(C3)$, we want the sum of optimal capacitated EOQ costs equal to $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$. By (4.2) and (A.41), $(C3)$ is equivalent to

$$\sum_{i\in S} \frac{k_i}{\hat{T}_i} + \eta_i \hat{T}_i + \sum_{j\in R} \frac{k_j}{\hat{T}_j} + \eta_j \hat{T}_j = \sum_{l\in\{1,\cdots,k\}} \frac{K(U_l^*)}{U_l^*} + U_l^* \cdot H(U_l^*) \tag{A.42}$$

To prove (A.42), it is sufficient to show $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ are the optimal solutions in decomposed EOQ-like problems and the total cost remains the same, that is

$$\begin{cases} \hat{T}_i = T_i^{s*}, & \forall i \in S \\ \hat{T}_j = T_j^{r*}, & \forall j \in R \\ \sum\limits_{i \in P(U_l^*)} \eta^i + \sum\limits_{j \in P(U_l^*)} \eta^j = H(U_l^*), & \forall P(U_l^*) \end{cases} \tag{A.43}$$

By the definitions of $\hat{T}_i$ and $\hat{T}_j$, (A.43) is equivalent to

$$\begin{cases} \sum\limits_{j \in R} \eta^{ij} = \eta_i = \frac{k_i^s}{(T_i^{s*})^2}, & \forall i \in S \\ \sum\limits_{i \in S} \eta_{ij} = \eta_j = \frac{c_j^r}{(T_j^{r*})^2}, & \forall j \in R, \text{ where } T_j^{r*} < T_j^{r0} \\ \sum\limits_{i \in S} \eta_{ij} = \eta_j = \frac{c_j^r}{(T_j^{r*})^2} - \lambda_j \leq \frac{c_j^r}{(T_j^{r*})^2}, & \forall j \in R, \text{ where } T_j^{r*} = T_j^{r0} \\ \sum\limits_{i \in P(U_l^*)} \eta^i + \sum\limits_{j \in P(U_l^*)} \eta^j = H(U_l^*), & \forall P(U_l^*) \end{cases} \tag{A.44}$$

To satisfy $(C5)$ we need

$$\sum_{j \in R} \sum_{i \in S} \int_0^{t'} (\eta_{ij} I_{ij}(t) + \eta^{ij} S_{ij}(t)) dt \leq \sum_{j \in R} \sum_{i \in S} \int_0^{t'} (g_{ij} I_{ij}(t) + g^{ij} S_{ij}(t)) dt \tag{A.45}$$

Since $I_{ij}(t) \leq S_{ij}(t)$, we only need

$$\begin{cases} \eta_{ij} \geq g_{ij}, & \forall i \in S, j \in R \\ \eta^{ij} + \eta_{ij} = g_{ij} + g^{ij}, & \forall i \in S, j \in R \\ \eta^{ij}, \eta_{ij} \geq 0, & \forall i \in S, j \in R \end{cases} \tag{A.46}$$

Now we have shown (A.44) and (A.46) are sufficient conditions for $(C3)$ and $(C4)$. It remains to show there exist such alternative parameters η_{ij}, η^{ij} and λ_j satisfying (A.44) and (A.46). In the following lemma we use subdifferential to prove the existence of these parameters.

Lemma 7. *There exists alternative cost parameters satisfying (A.44) and (A.46).*

Proof. η_{ij} and η^{ij} defined in (A.47) satisfy (A.46) for all $0 \leq A_{ij} \leq g^{ij}$:

$$\eta_{ij} = \begin{cases} g_{ij} & \text{if } T_j^{r*} < T_i^{s*} \\ g_{ij} + g^{ij} & \text{if } T_j^{r*} > T_i^{s*} \\ g_{ij} + A_{ij} & \text{if } T_j^{r*} = T_i^{s*} \end{cases} \text{ and } \eta^{ij} = \begin{cases} g^{ij} & \text{if } T_j^{r*} < T_i^{s*} \\ 0 & \text{if } T_j^{r*} > T_i^{s*} \\ g^{ij} - A_{ij} & \text{if } T_j^{r*} = T_i^{s*} \end{cases} \tag{A.47}$$

By Danskin's theorem, the subdifferential of a finite pointwise maximum function is the convex hull of the subdifferential of corresponding active functions Bertsekas (1999). Hence

$$\partial(\max(T_i^s, T_j^r)) = \begin{cases} \partial T_j^r, & \text{if } T_i^s < T_j^r \\ \partial T_i^s, & \text{if } T_i^s > T_j^r \\ \alpha_{ij} \cdot \partial T_i^s + (1-\alpha_{ij}) \cdot \partial T_j^r, & \text{if } T_i^s = T_j^r \end{cases}$$

Therefore the subdifferential of $C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r})$ is

$$\begin{aligned} \partial C_{IR}(\boldsymbol{T^s}, \boldsymbol{T^r}) = & \sum_{i \in S} \frac{-k_i^s}{(T_i^s)^2} \partial T_i^s + \sum_{j \in R} \frac{-c_j^r}{(T_j^r)^2} \partial T_j^r + \sum_{i \in S} \sum_{j \in L_i} g^{ij} \partial T_i^s + \sum_{i \in S} \sum_{j \in G_i} g^{ij} \partial T_j^r \\ & + \sum_{i \in S} \sum_{j \in E_i} g^{ij} \Big(\alpha_{ij} \cdot \partial T_i^s + (1-\alpha_{ij}) \cdot \partial T_j^r \Big) + \sum_{i \in S} \sum_{j \in R} g_{ij} \partial T_j^r \end{aligned} \quad \text{(A.48)}$$

Since $(\boldsymbol{T^{s*}, T^{r*}})$ is the optimal solution to $(\boldsymbol{PC})$, which is convex. Slater condition is satisfied in $\boldsymbol{(PC)}$, thus from KKT condition we know

$$\begin{cases} \mathbf{0} \in \partial C_{IR}(\boldsymbol{T^{s*}, T^{r*}}) + \sum_{j \in R} \lambda_j^* \partial T_j^{r*} \\ \lambda_j^* (T_j^{r*} - T_j^{r0}) = 0 \\ \lambda_j^* \geq 0 \end{cases} \quad \text{(A.49)}$$

where λ_j^* is the optimal solution of the dual problem. By definition of subdifferential, there exist $\alpha_{ij} \in [0,1]$ such that

$$\begin{cases} -\frac{k_i}{(T_i^*)^2} + \sum\limits_{j \in L_i} g^{ij} + \sum\limits_{j \in E_i} \alpha_{ij} g^{ij} = 0, & \forall i \in S \\ -\frac{c_j}{(T_j^*)^2} + \sum\limits_{i: j \in G_i} g^{ij} + \sum\limits_{i: j \in E_i} (1-\alpha_{ij}) g^{ij} + \sum\limits_{i \in S} g_{ij} = 0, & \forall j \in R, \text{ where } T_j^{r*} < T_j^{r0} \\ -\frac{c_j}{(T_j^*)^2} + \sum\limits_{i: j \in G_i} g^{ij} + \sum\limits_{i: j \in E_i} (1-\alpha_{ij}) g^{ij} + \sum\limits_{i \in S} g_{ij} + \lambda_j^* = 0, & \forall j \in R, \text{ where } T_j^{r*} = T_j^{r0} \end{cases} \quad \text{(A.50)}$$

We let $A_{ij} = (1-\alpha_{ij}) g^{ij}$ and $\lambda_j = \lambda_j^*$, then all the equations in (A.44) are satisfied from (A.50). □

In summary, combining (A.41), (A.45) and (A.42) we know:

$$\begin{aligned} & \sum_{i \in S} k_i^s n_i(t') + \sum_{j \in R} c_j^r n_j(t') + \sum_{j \in R} \sum_{i \in S} \int_0^{t'} (g_{ij} I_{ij}(t) + g^{ij} S_{ij}(t)) dt \\ \geq & \sum_{i \in S} \big(\frac{k_i^s}{T_i^s} + \eta_i T_i^s \big) + \sum_{j \in R} \big(\frac{c_j^r}{T_j^r} + \eta_j T_j^r \big) \\ \geq & C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) \end{aligned}$$

□

A.22 Proof of Theorem 17

We first decompose the cost according to the same partitions used previously:

$$\begin{aligned}
& C_{POT}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) \\
= & \sum_{i\in S}\frac{k_i}{T^{s*}_{i,P}} + \sum_{j\in R}\frac{c_j}{T^{r*}_{j,P}} + \sum_{i\in S, j\in R}\max\,(T^{s*}_{i,P}, T^{r*}_{j,P})\cdot g^{ij} + \sum_{i\in S, j\in R} T^{r*}_{j,P}\cdot g_{ij} \\
= & \sum_{i\in S}\frac{k_i}{T^{s*}_{i,P}} + \sum_{j\in R}\frac{c_j}{T^{r*}_{j,P}} + \sum_{i\in S, j\in L_i\cup E_i} T^{s*}_{i,P}\cdot g^{ij} + \sum_{j\in R, i:j\in G_i} T^{r*}_{j,P}\cdot g^{ij} + \sum_{i\in S, j\in R} T^{r*}_{j,P}\cdot g_{ij} \\
= & \sum_{l\in\{1,\cdots,k\}}\Big(\frac{K(U_l^{P*})}{U_l^{P*}} + H(U_l^{P*})\cdot U_l^{P*}\Big).
\end{aligned}$$

Thus, for each group $P(U_l^*)$ in the partition, we consider its capacity constrained problem ($\mathbf{PCR_l}$), and obtain the bound of 2 from PoT rounding:

$$C_{IRS}^{U_l^{P*}}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) \;=\; \frac{K(U_l^{P*})}{U_l^{P*}} + H(U_l^{P*})\cdot U_l^{P*} < C_{IRS}^{\frac{U_l^*}{2}}(\frac{1}{2}\boldsymbol{T^{s*}}, \frac{1}{2}\boldsymbol{T^{r*}}) < 2C_{IRS}^{U_l^*}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$$

Summing over cost of each of the decomposed problem, we get

$$C_{POT}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) = \sum_{l\in\{1,\cdots,k\}} C_{IRS}^{U_l^{P*}}(\boldsymbol{T_P^{s*}}, \boldsymbol{T_P^{r*}}) < \sum_{l\in\{1,\cdots,k\}} 2C_{IRS}^{U_l^*}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) = 2C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$$

From Theorem 3, we know $C_{IR}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on an arbitrary policy. Thus the worst case ratio for this centralized PoT policy is at most 2.

A.23 Proof of Theorem 18

We discuss rounding $(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ to nearest PoT solution in two cases:

1. $U_l^* \leq U_l^{\hat{P}*} \leq \sqrt{2}U_l^*$:
   When we round up order intervals to the nearest PoT solution, at most two trucks

will be used in each shipment. Thus we can show

$$
\begin{aligned}
C_{IRS}^{U_l^{\hat{P}*}}(\boldsymbol{T}_{\hat{\boldsymbol{P}}}^{\boldsymbol{s*}}, \boldsymbol{T}_{\hat{\boldsymbol{P}}}^{\boldsymbol{r*}}) &\leq \sum_{i \in P(U_l^{\hat{P}*})} \frac{k_i}{U_l^{\hat{P}*}} + \sum_{j \in P(U_l^{\hat{P}*})} \frac{2c_j}{U_l^{\hat{P}*}} + H(U_l^{\hat{P}*}) \cdot U_l^{\hat{P}*} \\
&\leq \frac{2K(U_l^{\hat{P}*})}{U_l^{\hat{P}*}} + H(U_l^{\hat{P}*}) \cdot U_l^{\hat{P}*} \\
&\leq \frac{2K(U_l^*)}{U_l^*} + H(U_l^{\hat{P}*}) \cdot \sqrt{2} U_l^* \\
&\leq 2 \cdot C_{IRS}^{U_l^*}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})
\end{aligned}
$$

2. $\frac{1}{\sqrt{2}} U_l^* \leq U_l^{\hat{P}*} \leq U_l^*$:
When we round down order intervals, we can show a tighter bound:

$$
C_{IRS}^{U_l^{\hat{P}*}}(\boldsymbol{T}_{\hat{\boldsymbol{P}}}^{\boldsymbol{s*}}, \boldsymbol{T}_{\hat{\boldsymbol{P}}}^{\boldsymbol{r*}}) \leq \frac{K(U_l^{\hat{P}*})}{\frac{1}{\sqrt{2}} U_l^{\hat{P}*}} + H(U_l^{\hat{P}*}) \cdot U_l^{\hat{P}*} \leq \sqrt{2} \cdot C_{IRS}^{U_l^*}(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})
$$

Therefore, the worst case performance for this PoT rounding is also 2.

A.24 Proof of Theorem 19

Proof. From Theorem 16 we know $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$ is a lower bound on the cost of the centralized model. From KKT conditions we also know

$$
\sum_{i \in S} \frac{k_i}{T_i^{s*}} + \sum_{j \in R} \frac{c_j}{T_j^{r*}} \geq \sum_{i \in S, j \in R} \max(T_i^{s*}, T_j^{r*}) \cdot g^{ij} + \sum_{i \in S, j \in R} T_j^{r*} \cdot g_{ij} \tag{A.51}
$$

In the following we show

$$
\frac{3}{2} C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}}) \geq \sum_{i \in S} \frac{k_i}{T_i^{s*}} + \sum_{j \in R} \frac{c_j}{T_j^{r*}} + \sum_{i \in S, j \in L_i \cup G_i} T_i^{s*} \cdot g^{ij} + \sum_{i \in S, j \in R} T_j^{r*} \cdot (g_{ij} + g^{ij}),
$$

where the right hand side is the cost of ZIO policy we obtaind in (4.5).

$$
\begin{aligned}
\frac{3}{2}C_C(\boldsymbol{T^{s*}},\boldsymbol{T^{r*}}) &= \frac{3}{2}\Big(\sum_{i\in S}\frac{k_i}{T_i^{s*}}+\sum_{j\in R}\frac{c_j}{T_j^{r*}}+\sum_{i\in S,j\in R}\max\left(T_i^{s*},T_j^{r*}\right)\cdot g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}\cdot g_{ij}\Big)\\
&\geq \sum_{i\in S}\frac{k_i}{T_i^{s*}}+\sum_{j\in R}\frac{c_j}{T_j^{r*}}+2\Big(\sum_{i\in S,j\in R}\max\left(T_i^{s*},T_j^{r*}\right)\cdot g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}\cdot g_{ij}\Big) \quad \text{(A.52)}\\
&\geq \sum_{i\in S}\frac{k_i}{T_i^{s*}}+\sum_{j\in R}\frac{c_j}{T_j^{r*}}+\Big(\sum_{i\in S,j\in R}(T_i^{s*}+T_j^{r*})\cdot g^{ij}+\sum_{i\in S,j\in R}T_j^{r*}\cdot g_{ij}\Big) \quad \text{(A.53)}\\
&= C_{zio}(\boldsymbol{T^{s*}},\boldsymbol{T^{r*}})
\end{aligned}
$$

In the above derivation, (A.52) is because of (A.51), and (A.53) is by the definition of maximum function.

□

A.25 Proof of Theorem 20

We start from a simple case with one retailer and one supplier. From results in previous paper, we know they must share the same order cycle and we can combine cost and omit subscripts in notation for simplicity. The naive model in our previous paper degenerated to the classical EOQ model, with cost function $\frac{k}{T}+hT$ and optimal solution $T^*_{sim}=\sqrt{\frac{k}{h}}$. If $T^*_{sim}\leq T^0$ then the optimal solution in simple model only uses one truckload, which is trivial. Otherwise we denote $T^*_{sim}=pT^0$ where $p>1$, and

$$C_{sim}(T^*_{sim})=\frac{k}{T^*_{sim}}\lceil p\rceil+hT^*_{sim}=\frac{k}{pT^0}\lceil p\rceil+hpT^0, \quad \text{(A.54)}$$

while the optimal solution to new model is T^0, and the corresponding cost is

$$C_C(T^0)=\frac{k}{T^0}+hT^0 \quad \text{(A.55)}$$

Comparing (A.54) and (A.55), we come to

$$\frac{C_{sim}(T^*_{sim})}{C_C(T^0)}=\frac{\frac{k}{pT^0}\lceil p\rceil+hpT^0}{\frac{k}{T^0}+hT^0}=\frac{hpT^0\lceil p\rceil+hpT^0}{hpT^0p+\frac{hpT^0}{p}}=\frac{p\lceil p\rceil+p}{p^2+1} \quad \text{(A.56)}$$

To evaluate the worst case bound, we only need to maximize (A.81) when $p \geq 1$. We discuss in different unit integer intervals. When $p \in (n, n+1), n \in \mathbb{N}$

$$(A.81) = \frac{p(n+1)+p}{p^2+1} = (n+2)\frac{p}{p^2+1}, \tag{A.57}$$

which is decreasing in $(n, n+1)$. Thus

$$(A.82) < (n+2)\frac{n}{n^2+1} = \frac{n(n+2)}{n^2+1}, \tag{A.58}$$

and the right hand side is tight when $p \to n^+$. The derivative of (A.58) is

$$\frac{(2n+2)(n^2+1) - 2n(n^2+2)}{(n^2+1)^2} = 2 \cdot \frac{-n^2+n+1}{(n^2+1)^2} \tag{A.59}$$

(A.85)> 0 only when $n = 1$, so the maximizer of (A.58) is obtained when $n = 1$ or $n = 2$. Therefore the tight upper bound of (A.81) is $\frac{8}{5}$ when $p \to 2^+$.

Next, we consider more general case of multi-supplier multi-retailer. If optimal partition in two models are the same, we can decompose and show that the worst case performance of uncapacitated model is $\frac{8}{5}$, which is attainable. But in general partition can be different, and the upper bound for worst case ratio is not clear yet. Therefore we conclude the worst case cost increase of applying simple model is at least 60%.

A.26 Proof of Theorem 21

Proof. Compared to the optimal order interval in the centralized model,

$$\Gamma_j^{r*} = min\left\{T_j^{r0}, \sqrt{\frac{k_j^r}{\sum\limits_{i\in S}(g_{ij}+g^{ij})}}\right\} \leq min\left\{T_j^{r0}, \sqrt{\frac{k_j^r}{\sum\limits_{i\in S} g_{ij} + \sum\limits_{i:j\in L_i\cup E_i} g^{ij}}}\right\} \leq T_j^{r*}.$$

□

A.27 Proof of Theorem 24

Proof. From Theorem 16, we know $C_C(\boldsymbol{T^s}, \boldsymbol{T^r})$ is a lower bound on cost of centralized model.

$$C_C(\boldsymbol{T^s}, \boldsymbol{T^r}) = \sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{c_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R}\max{(T_i^{s*}, T_j^{r*})}g^{ij} + \sum_{i\in S, j\in R} T_j^{r*}g_{ij}.$$

In the following we show

$$\frac{3}{2}C_{IR}(\boldsymbol{T^s},\boldsymbol{T^r}) \geq \sum_{i\in S} C_i^{zio}(\Gamma_i^{s*}) + \sum_{j\in R} C_j^r(\Gamma_j^{r*}),$$

where the right hand side is cost of the optimal ZIO policy we obtained in Section 4.2.3.

$$\begin{aligned}
\frac{3}{2}C_C(\boldsymbol{T^s},\boldsymbol{T^r}) \geq & \left\{ \sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{c_j^r}{T_j^{r*}} + \sum_{i\in S,j\in R} T_j^{r*}g^{ij} + \sum_{i\in S,j\in R} T_j^{r*}g_{ij} \right\} \\
& + \frac{1}{2}\left\{ \sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{j\in R}\frac{c_j^r}{T_j^{r*}} + \sum_{i\in S,j\in R} \max{(T_i^{s*},T_j^{r*})}g^{ij} + \sum_{i\in S,j\in R} T_j^{r*}g_{ij} \right\} \qquad \text{(A.60)} \\
\geq & \left\{ \sum_{j\in R}\frac{c_j^r}{T_j^{r*}} + \sum_{i\in S,j\in R} T_j^{r*}g^{ij} + \sum_{i\in S,j\in R} T_j^{r*}g_{ij} \right\} + \sum_{i\in S}\frac{k_i^s}{T_i^{s*}} \\
& + \left\{ \sum_{i\in S,j\in R} \max{(T_i^{s*},T_j^{r*})}g^{ij} + \sum_{i\in S,j\in R} T_j^{r*}g_{ij} \right\} \qquad \text{(A.61)} \\
\geq & \sum_{j\in R} C_j^r(\Gamma_j^{r*}) + \left\{ \sum_{i\in S}\frac{k_i^s}{T_i^{s*}} + \sum_{i\in S,j\in R} T_i^{s*}g^{ij} \right\} \qquad \text{(A.62)} \\
\geq & \sum_{i\in S} C_i^{zio}(\Gamma_i^{s*}) + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \qquad \text{(A.63)}
\end{aligned}$$

In the above derivation, (A.60) is true because the property of maximum function; (A.61) is true because of (A.51); (A.62) is true since Γ_j^{r*} is the optimal decentralized policy of retailers in Section 4.2.3, and we eliminate positive terms $\sum T_j^{r*}g_{ij}$; (A.63) is true because $\tilde{\Gamma}_i^{s*}$ minimize $C_i^{zio}(\Gamma_i^s)$.

From the analysis above, we know that an upper bound of decentralized cost is no more than $\frac{3}{2}$ of the lower bound of centralized cost. Hence, the price of anarchy is bounded by $\frac{3}{2}$. □

A.28 Proof of Theorem 25

$$\frac{5}{2} \cdot C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$$

$$= \frac{3}{2}\Big\{ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} \frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R} \max(T_i^{s*}, T_j^{r*}) g^{ij} + \sum_{i\in S, j\in R} T_j^{r*} g_{ij} \Big\}$$
$$+ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \Big\{ \sum_{j\in R} \frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R} T_j^{r*}(g_{ij} + g^{ij}) \Big\} + \sum_{i\in S, j\in R} (\max(T_i^{s*}, T_j^{r*}) - T_j^{r*}) g^{ij} \tag{A.64}$$

$$\geq \frac{3}{2}\Big\{ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} \frac{k_j^r}{T_j^{r*}} + \sum_{i\in S, j\in R} \max(T_i^{s*}, T_j^{r*}) g^{ij} + \sum_{i\in S, j\in R} T_j^{r*} g_{ij} \Big\}$$
$$+ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \tag{A.65}$$

$$\geq \frac{3}{2}\Big\{ \sum_{i\in S, j\in R} 2\max(T_i^{s*}, T_j^{r*}) g^{ij} + \sum_{i\in S, j\in R} 2T_j^{r*} g_{ij} \Big\} + \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \tag{A.66}$$

$$\geq \Big\{ \sum_{i\in S, j\in L_i} (2T_j^{r*} + T_i^{s*}) g^{ij} + \sum_{i\in S, j\in G_i} 3(T_i^{s*} + (T_j^{r*} - T_i^{s*})) g^{ij} \Big\}$$
$$+ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \tag{A.67}$$

$$\geq \Big\{ \sum_{i\in S, j\in L_i} (2\Gamma_j^{r*} + T_i^{s*}) g^{ij} + \sum_{i\in S, j\in G_i} T_i^{s*} g^{ij} + \sum_{i\in S, j\in G_i} 2(T_j^{r*} - T_i^{s*}) g^{ij} \Big\}$$
$$+ \sum_{i\in S} \frac{k_i^s}{T_i^{s*}} + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \tag{A.68}$$

$$\geq \sum_{i\in S} C_i^{out}(\Gamma_i^{s*}) + \sum_{j\in R} C_j^r(\Gamma_j^{r*}) \tag{A.69}$$

In the above derivation, (A.64) is true due to rearrangement of maximum function in cost evaluation $C_C(\boldsymbol{T^{s*}}, \boldsymbol{T^{r*}})$; (A.65) is true because we eliminate $\sum_{i\in S, j\in R} (\max(T_i^{s*}, T_j^{r*}) - T_j^{r*}) g^{ij}$, and because Γ_j^{r*} is the optimal decentralized policy of retailers in Section 4.2.3; (A.66) is implied by $(C1)$ and (A.51); (A.67) is true by the definition of maximum function and G_i, L_i; (A.68) comes from deleting and rearranging terms; (A.69) is true because Γ_i^{s*} is the minimizer of $C_i^{out}(\Gamma_i^{s*})$.

A.29 Proof of Theorem 26

Proof. We first show $C_j(Q_j, S_{1j}, \cdots, S_{nj})$ is convex in S_{ij}.

$$\begin{aligned}
\Delta C_{ij}(S_{ij}) &\triangleq C_j(Q_j, S_{1j}, \cdots, S_{ij}+1, \cdots, S_{nj}) - C_j(Q_j, S_{1j}, \cdots, S_{ij}, \cdots, S_{nj}) \\
&= h_{ij} + (p_{ij}+h_{ij})\Big(\sum_{v \geq S_{ij}+1} (v - S_{ij} - 1) m_i(v) - \sum_{v \geq S_{ij}} (v - S_{ij}) m_i(v)\Big) \\
&= h_{ij} - (p_{ij}+h_{ij}) \sum_{v \geq S_{ij}+1} m_i(v) - (S_{ij} - S_{ij}) m_i(S_{ij}) \\
&= h_{ij} - (p_{ij}+h_{ij})\Big(1 - \sum_{v \leq S_{ij}} m_i(v)\Big) \\
&= -p_{ij} + (p_{ij}+h_{ij}) M_i(S_{ij}),
\end{aligned}$$

It is easy to see $\Delta C_{ij}(S_{ij})$ is monotone increasing in S_{ij}. Thus

$$\begin{aligned}
S^*_{ij}(Q_j) &= \min\Big\{S_{ij} : -p_{ij} + (p_{ij}+h_{ij}) M_i(S_{ij}) \geq 0\Big\} \\
&= M_i^{-1}\Big(\frac{p_{ij}}{p_{ij}+h_{ij}}\Big)
\end{aligned}$$

□

A.30 Proof of Theorem 27

Proof. To show the monotonicity, it is equivalent to show that $M_i(v)$ is decreasing in Q.

$$M_i(v) = P(V_i(t+L_j) \leq v) = \sum_{x=0}^{v} \Big\{P(D^{ij}_{[t,t+L_j)} = x) \sum_{m=0}^{v-x} u_i(m)\Big\} \tag{A.70}$$

Therefore by (5.1) it is sufficient to show that the following $U(x)$ is decreasing in Q_j:

$$U(x) \triangleq \sum_{m=0}^{x} u_i(m) = \sum_{m=0}^{x} \frac{1 - B(m; Q_j, \theta_i)}{\theta_i Q_j} \tag{A.71}$$

That is to show:

$$\sum_{m=0}^{x} \frac{1 - B(m; Q+1, \theta_i)}{\theta_i (Q+1)} \leq \sum_{m=0}^{x} \frac{1 - B(m; Q, \theta_i)}{\theta_i Q} \tag{A.72}$$

(A.72) is equivalent to:

$$\sum_{m=0}^{x} \frac{P(\geq \text{m+1 successes at Q+1})}{Q+1} \leq \sum_{m=0}^{x} \frac{P(\geq \text{m+1 successes at Q})}{Q} \tag{A.73}$$

By combinatorial identity, left hand side of (A.73) is equivalent to:

$$\sum_{m=0}^{x} \frac{P(\geq \text{m+1 successes at Q}) + P(\text{m successes at Q, and success at Q+1})}{Q+1} \tag{A.74}$$

Therefore (A.73) is tranformed to:

$$\begin{aligned}\sum_{m=0}^{x} \frac{P(\geq \text{m+1 successes at Q+1})}{Q(Q+1)} &\geq \sum_{m=0}^{x} \frac{P(\text{m successes at Q, and success at Q+1})}{Q+1} \\ &= \theta_i \sum_{m=0}^{x} \frac{P(\text{m successes at Q})}{Q+1} \\ &= \theta_i \frac{P(\leq \text{x successes at Q})}{Q+1} \end{aligned} \tag{A.75}$$

Rearrange terms, to prove (A.75) is equivalent to showing:

$$\sum_{m=0}^{x} P(\geq \text{m+1 successes at Q+1}) \geq \theta_i Q(1 - P(\geq \text{x+1 successes at Q})) \tag{A.76}$$

We also notice:

$$\theta_i P(\geq \text{x+1 successes at Q}) \geq P(\geq \text{x+2 successes at Q+1 }). \tag{A.77}$$

Iteratively using (A.77), we get the bound

$$\begin{aligned}\theta_i P(\geq \text{x+1 successes at Q}) &\geq P(\geq \text{x+1+k successes at Q+k}) \cdot (\theta_i)^{-k+1} \\ &\geq P(\geq \text{x+1+k successes at } Q+1) \end{aligned} \tag{A.78}$$

Hence

$$\begin{aligned}&\sum_{m=0}^{x} P(\geq \text{m+1 successes at Q+1}) + \theta_i Q P(\geq \text{x+1 successes at Q}) \\ &\geq \sum_{m=0}^{x} P(\geq \text{m+1 successes at Q+1}) + \sum_{m=x+1}^{Q} P(\geq \text{m+1 successes at Q+1}) \quad \text{(A.79)} \\ &= \sum_{m=0}^{Q} P(\geq \text{m+1 successes at Q+1}) \\ &= \mathbb{E}(\text{Total successes at } Q+1) \quad \text{(A.80)} \\ &= \theta_i (Q+1) \end{aligned}$$

In the above derivation, (A.79) applies (A.78) and $x \geq 0$; (A.80) count the total successes by probabilities. After rearanging the terms, (A.76) is proved, and $S^*_{ij}(Q_j)$ is increasing in Q_j. □

A.31 Proof of Theorem 28

Proof. We first notice In (5.3), the fixed ordering cost $\frac{k_j}{Q_j}\lambda_j$ and inventory holding cost $\sum_{i\in S} h_{ij}\Big(S_{ij} - \frac{\lambda_{ij}(Q_j-1)}{2} - \lambda_{ij}L_j\Big)$ are convex in Q_j, thus we only need to show backorder cost is also convex in Q_j.

Lemma 8. *When aggregate demand $X_0(t) = n$ is fixed, the average backorder level at time $t + L_j$ is*

$$f_{ij}(n) \triangleq \mathbb{E}\Big\{max(0, Binomial(n, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij})\Big\},$$

then $f_{ij}(n)$ is convex in n.

Proof. When$X_0(t) = n$ and by independence of demand of different products, the number of product i follows a binomial distribution. $D^{ij}_{[t,t+L_j]}$ is the demand occurs during lead time, and it follows a Poisson distribution with $\lambda_{ij}L_j$. Thus their sum is the total demand of product i at time $t + L_j$.

$$f_{ij}(n-1) + f_{ij}(n+1) - 2f_{ij}(n) = \mathbb{E}\Big\{max(0, Binomial(n+1, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij})$$
$$+ max(0, Binomial(n-1, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij}) - 2max(0, Binomial(n, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij})\Big\}$$
$$= \mathbb{E}\Big\{max(0, Binomial(n-1, \lambda_{ij}) + \mathbb{1}_p + \mathbb{1}_p + D^{ij}_{[t,t+L_j]} - S_{ij})$$
$$+ max(0, Binomial(n-1, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij}) - 2max(0, Binomial(n-1, \lambda_{ij}) + \mathbb{1}_p + D^{ij}_{[t,t+L_j]} - S_{ij}) \tag{A.81}$$
$$= \mathbb{E}\Big\{max(0, Z + \mathbb{1}_p + \mathbb{1}_p) + max(0, Z) - 2max(0, Z + \mathbb{1}_p)\Big\}, \tag{A.82}$$

where $\mathbb{1}_p$ is a bernoulli random variable with successful rate of λ_{ij}, and $Z = Binomial(n-1, \lambda_{ij}) + D^{ij}_{[t,t+L_j]} - S_{ij}$ is a random variable takeing value in integer. In the above derivation, (A.81) is true because binomial random variable is the sum of independent bernoulli random variables.

Thus by tower law property,

$$
\begin{aligned}
(A.82) &= \mathbb{E}\Big\{\mathbb{E}\Big\{max(0, Z + \mathbb{1}_p + \mathbb{1}_p) + max(0, Z) - 2max(0, Z + \mathbb{1}_p)\Big| Z\Big\}\Big\} \\
&= \mathbb{E}\Big\{max(0, Z + \mathbb{1}_p + \mathbb{1}_p) + max(0, Z) - 2max(0, Z + \mathbb{1}_p)\Big| Z \neq -1\Big\} P(Z \neq -1) \\
&\quad + \mathbb{E}\Big\{max(0, Z + \mathbb{1}_p + \mathbb{1}_p) + max(0, Z) - 2max(0, Z + \mathbb{1}_p)\Big| Z = -1\Big\} P(Z = -1) \qquad \text{(A.83)} \\
&= \mathbb{E}\Big\{max(0, Z) + \mathbb{1}_p + \mathbb{1}_p + max(0, Z) - 2max(0, Z) - 2\mathbb{1}_p\Big| Z \neq -1\Big\} P(Z \neq -1) \\
&\quad + \mathbb{E}\Big\{max(0, -1 + \mathbb{1}_p + \mathbb{1}_p) + 0 - 0\Big| Z = -1\Big\} P(Z = -1) \qquad \text{(A.84)} \\
&\geq 0
\end{aligned}
$$

In (A.83), we calculate the conditional mean in two cases; in (A.84), $\mathbb{1}$ can be pulled out of the maximum function when $Z \neq 1$, and maximum function is easy to calculate whne $Z = 1$. Therefore we have shown $f_{ij}(n)$ is convex in n. □

Next we introduce a technical lemma to help prove Theorem 28:

Lemma 9. *If $f(i) : \mathbb{N} \to \mathbb{R}$ is convex in i, where $\mathbb{N}$ is the set of integers, then its running average $g(Q) = \frac{1}{Q}\sum_{i=0}^{Q-1} f(i)$ is convex in Q.*

Proof. We denote $\Delta_i = f(i) - f(i-1)$ and let $\Delta_0 = 0$. By convexity of $f(i)$ we know Δ_i is non-decreasing. That is,

$$0 = \Delta_0 \leq \Delta_1 \leq \cdots \leq \Delta_n \leq \cdots. \qquad \text{(A.85)}$$

Hence we rewrite $g(Q)$:

$$
\begin{aligned}
g(Q) &= \frac{1}{Q}\sum_{i=0}^{Q-1}\{f(0) + \sum_{j=0}^{i}\Delta_j\} \\
&= \frac{1}{Q}\sum_{i=0}^{Q-1}\{f(0) + (Q-i)\Delta_i\} \qquad \text{(A.86)} \\
&= f(0) + \frac{1}{Q}\sum_{i=0}^{Q-1}(Q-i)\Delta_i,
\end{aligned}
$$

where in (A.86) we exchange the order of summation. Thus the step difference is

$$\begin{aligned}\Delta g(Q) &\triangleq g(Q+1) - g(Q) \\ &= f(0) + \frac{1}{Q+1}\sum_{i=0}^{Q}(Q+1-i)\Delta_i - f(0) - \frac{1}{Q}\sum_{i=0}^{Q-1}(Q-i)\Delta_i \\ &= \frac{1}{Q+1}\Delta_Q + \sum_{i=0}^{Q-1}\Big(\frac{Q+1-i}{Q+1} - \frac{Q-i}{Q}\Big)\Delta_i \qquad \text{(A.87)} \\ &= \frac{1}{Q+1}\Delta_Q + \frac{1}{Q(Q+1)}\sum_{i=0}^{Q-1} i\cdot\Delta_i \qquad \text{(A.88)}\end{aligned}$$

Next we show the difference of $g(\cdot)$ is non-decreasing:

$$\begin{aligned}\Delta^2 g(Q) &\triangleq \Delta g(Q+1) - \Delta g(Q) \\ &= \frac{1}{Q+2}\Delta_{Q+1} + \frac{1}{(Q+1)(Q+2)}\sum_{i=0}^{Q} i\cdot\Delta_i - \frac{1}{Q+1}\Delta_Q - \frac{1}{Q(Q+1)}\sum_{i=0}^{Q-1} i\cdot\Delta_i \\ &\geq \Big(\frac{1}{Q+2} + \frac{Q}{(Q+1)(Q+2)} - \frac{1}{Q+1}\Big)\Delta_Q + \Big(\frac{1}{(Q+1)(Q+2)} - \frac{1}{Q(Q+1)}\Big)\sum_{i=0}^{Q-1} i\cdot\Delta_i \qquad \text{(A.89)} \\ &= \frac{Q-1}{(Q+1)(Q+2)}\Delta_Q - \frac{2}{Q(Q+1)(Q+2)}\sum_{i=0}^{Q-1} i\cdot\Delta_i \\ &= \frac{2}{Q(Q+1)(Q+2)}\sum_{i=0}^{Q-1} i\cdot(\Delta_Q - \Delta_i) \qquad \text{(A.90)} \\ &\geq 0 \qquad \text{(A.91)}\end{aligned}$$

In the above derivation, (A.89) and (A.91) are true because of convexity (A.85); (A.90) is true by Gaussian summation formula: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Therefore we show running average of $f(i)$ is also covex. □

Now we come back to proof of Theorem 28. The average backorder level

$$\sum_{v\geq S_{ij}}(v - S_{ij})m_i(v) = \frac{1}{Q}\sum_{q=0}^{Q-1} f_{ij}(q) \qquad \text{(A.92)}$$

By Lemma 8 we know $f_{ij}(q)$ is convex in q, and therefore we show the running average (A.92) is convex in upper limit Q, from the result of technical Lemma 9. In summary,

$C_j(Q_j, S_{1j}, \cdots, S_{nj})$, the total cost of retailer i, is the sum of convex functions, thus remains convex in Q. Therefore by the convexity of C_j in Q, the optimal aggregate level $Q*$ at given order-up-to levels is

$$argmin_{Q \in \mathbb{Z}} \{Q : C_j(Q+1, S_{1j}, \cdots, S_{nj}) \geq C_j(Q, S_{1j}, \cdots, S_{nj})\}.$$

□