STOCHASTIC JOINT REPLENISHMENT PROBLEM: A NEW POLICY AND ANALYSIS FOR SINGLE LOCATION AND TWO ECHELON INVENTORY SYSTEMS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCE

OF BİLKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Banu Yüksel Özkaya

December 2005

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Prof. Ülkü Gürler (Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Emre Berk (Co-advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Prof. Nesim Erkip

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Müge Avşar

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Doğan Serel

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Asst. Prof. Alper Şen

Approved for the Institute of Engineering and Science:

Prof. Mehmet Baray,
Director of Institute of Engineering and Science

Abstract

STOCHASTIC JOINT REPLENISHMENT PROBLEM: A NEW POLICY AND ANALYSIS FOR SINGLE LOCATION AND TWO ECHELON INVENTORY SYSTEMS

Banu Yüksel Özkaya
Ph.D. in Industrial Engineering
Advisor: Prof. Ülkü Gürler
December 2005

In this study, we examine replenishment coordination strategies for multiple item or multiple location inventory systems. In particular, we propose a new, parsimonious control policy for the stochastic joint replenishment problem. We first study the single location setting with multiple items under this policy. An extensive numerical study indicates that the proposed policy achieves significant cost improvements in comparison with the existing policies. The single location model also represents a two-echelon supply chain for a single item with multiple locations, where the upper echelon employs cross docking. We then extend our model to incorporate multi-location settings where the upper echelon also holds inventory. Our modeling methodology based on the development of the ordering process by the lower echelon provides an analytical tool to investigate various joint replenishment policies. An extensive numerical study is conducted to determine the performance of the system and identify regions of dominance across policies.

Keywords: Stochastic joint replenishment problem, multi-item inventory system, two-echelon inventory system

Özet

RASSAL TOPLU SİPARİŞ PROBLEMİ: YENİ BİR POLİTİKA VE TEK VE İKİ DÜZEYLİ ENVANTER SİSTEMLERİNİN ANALİZİ

Banu Yüksel Özkaya
Endüstri Mühendisliği Doktora
Tez Yöneticisi: Prof.Dr. Ülkü Gürler
Aralık 2005

Bu çalışmada çok ürünlü ve çok yerleşimli envanter sistemleri için koordineli sipariş verme stratejileri incelenmiştir. Rassal toplu sipariş verme problemi için kolay uygulanabilen yeni bir kontrol politikası önerilmiştir. İlk olarak bu politika altında tek yerleşimli ve çok ürünlü bir envanter sistemi incelenmiştir. Kapsamlı olarak yapılan sayısal bir çalışma ile önerilen politikanın mevcut politikalara göre önemli maliyet azalmaları sağladığı saptanmıştır. İncelenen tek yerleşimli model aynı zamanda tek ürünlü ve üst düzeyin geçiş noktası olarak kullanıldığı iki düzeyli bir tedarik zincirini de temsil etmektedir. Bu model üst düzeyin envanter tuttuğu iki düzeyli envanter sistemlerini de incelemek üzere genişletilmiştir. Alt düzeyin sipariş verme sürecinin geliştirilmesine dayalı olan metodoloji farklı toplu sipariş verme politikalarını analitik olarak incelememizi sağlamıştır. İncelenen sistemde farklı toplu sipariş verme politikaların üstünlük sağladığı bölgeleri tanımlamak için kapsamlı bir sayısal çalışma gerçekleştirilmiştir.

Anahtar sözcükler: Rassal toplu sipariş problemi, çok ürünlü envanter sistemleri, iki düzeyli envanter sistemleri

my family...

Acknowledgement

First and foremost, I would like to express my sincere gratitude to Prof. Ülkü Gürler for giving me the opportunity to study my Ph.D. with her as well as my M.S. I feel myself fortunate to have worked under the supervision of her. She has been ready to provide help, support and trust even during times of slow or no progress. Next, I would like to thank Asst. Prof. Emre Berk for his valuable contribution to my Ph.D. study as well as my academic perspectives. His quick thinking and extreme knowledge with generous discussions made it very pleasant to work with him. I would like to thank them both for everything they have done for this research and also for my future career. I hope I will be a kind of researcher as they want me to be.

I am also indebted to Prof. Nesim Erkip, Asst. Prof. Alper Şen, Asst. Prof. Müge Avşar and Asst. Prof. Doğan Serel for devoting their valuable time to review this thesis and their invaluable suggestions and feedback.

I also want to thank Prof. Mustafa Pınar who was a committee member in the earlier stages of this thesis.

I would like to express my deepest thanks to Asst. Prof. Ayten Türkcan for all encouragement and academic support even she was very far away. We went through a similar experience and she helped make me feel not alone. Without her continuous morale support during my desperate times, I would not be able to bear all.

I am grateful to Sinan Gürel for his understanding, support at all times. I admire his calm and rational attitude toward life. My special thanks are for Ayşegül Altın who is one of the kindest person I have ever met. I would like to extend my warmest thanks to Berrin Keyik Çelik and Burcu Uslu Özdemir for their intimate friendship.

My sincere thanks are for Çağdaş Büyükkaramıklıfor generously sharing many ideas about life. I would especially like to thank Eray Yücel for his encouragement and support at hopeless times. I would also like to thank my office mate Önder Bulut for his understanding during my last semester.

I express thanks to all, whose names I can not mention, for their friendship, help and morale support during my study at Bilkent.

I am indebted to my father Prof. Mithat Yüksel, my mother Prof. Ümran Yüksel in creating an environment in which pursuing a Ph.D. study seemed so natural. They have given me everlasting support, encouragement and love. I owe gratitude to my grandmother Zehra Çıray, my sister Dr. Mehtap Yüksel Eğrilmez and her husband Dr. Murat Eğrilmez for their support, trust and love. I want to thank them all for all the beauties of my life.

I also want to take this opportunity to thank Ahmet Özkaya, Dr. Nurhayat Özkaya and Algı Özkaya for their constant help and support in all parts of my life.

Last but not the least, I wish to express my gratitude to Ongun Can Özkaya for his constant understanding. I can never bring back the days and nights that I left him alone during my study. Without his support, this work would probably never be possible. I owe so much to him in my life for the happiness he brought to me with his love and kindness.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The management of multi-echelon or multi-item inventory systems has been one of the most challenging issues both in practice and theory for years. This has become even more critical in the recent years with the concept of supply chain management. The coordination and integration are the key terms to reduce costs and increase the efficiency in an inventory system, which has become possible with the recent advances in information technology. Therefore, effective utilization of available information about the inventory status among the different locations/items in the inventory system is crucial. In this study, we study coordinated replenishment policies in both single-location/multi-item and single-item/multi-location inventory settings. Specifically, we study the stochastic joint replenishment problem (SJRP) in different settings.

SJRP was originally described in a single-location and multi-item inventory system. It is the determination of replenishment and stocking decisions for N different items to minimize the expected total ordering, holding and shortage costs per unit time in the presence of random demands and ordering cost structures with *first-order-interaction*. The *first-order-interaction* structure for ordering costs is defined as the setting where (see Balintfy [15] and *e.g.* Federgruen and Zheng [32]) there are *(i)* a common fixed cost associated with a replenishment order regardless of its composition, and *(ii)* an item-specific fixed cost for each item that is included in the replenishment order.

The ordering cost structure presents an opportunity to exploit the economies of scale in replenishment by ordering items jointly. Such joint replenishment opportunities occur when it is possible to include several different items in the same delivery order or when the items are purchased from the same supplier or they share the same transportation vehicle. Hence, effective joint replenishment policies are needed.

Under a stochastic joint replenishment policy, when an item is taken in isolation, it experiences exogenously generated opportunities of replenishment with reduced fixed costs, *ie.*, it can order with item-specific cost rather than common fixed cost. When a reordering decision for an item is triggered by its own inventory position, this may generate opportunities of replenishment at reduced cost for the other items. Clearly, these aspects are inter-related and may influence the performances of the policies. However, we believe that the generation of the replenishment opportunity arrivals is crucial in understanding the SJRP. In a multi-item setting, the employed policy is the generator of the opportunity arrival process. Hence, by choosing a particular policy to employ, we also choose a particular mechanism to generate the replenishment opportunities to the system. In the presence of such replenishment opportunities with reduced costs, it is intuitive that it may be beneficial to reorder an item at some (or all) of these opportunity arrivals which are no longer the demand instances for the items. Obviously, the overall costs incurred by the inventory system depends greatly on how these opportunities arrive at the system, which, to our understanding, also differentiates the performances of the policies. It is also important to have parsimonious joint replenishment policies, operating with fewer control policy parameters and easier to model and optimize.

The determination of these opportunity generation mechanisms and hence joint replenishment policies in multi-item inventory systems is a real problem faced by retailers and is an integral part of supply chain management in general. Moreover, it is becoming an increasingly important problem due to the recent trend among manufacturers and retailers to reduce their supplier bases (Harland [43]). It is estimated that major Original Equipment Manufacturers

(OEM's) have reduced the number of their suppliers by 25% since the mid-1990s. A best practice study reports that world-class companies operate with 97% fewer suppliers for A-category items, when compared with the average (The Hackett Group, www.thehackettgroup.com). Another survey reveals that, 80% of the firms directly considered the potential cost savings due to the reduction of transaction costs among multiple suppliers (Cousin [26]). In their recent works, Erhun and Tayur [29] and Cachon [17] also report particular instances of considerable cost savings achieved by exploiting the economies of scale due to joint replenishment opportunities.

As will be explained in the next Chapter, the SJRP has been usually addressed in single-location and multi-item inventory systems. Despite the successful implementation of efficient coordinated replenishment policies in many retail companies (www.smartops.com) and considerable cost savings achieved, as reported in Erhun and Tayur [29] and Cachon [17], little theoretical work has been done to evaluate the benefits of these coordinated policies in multi-echelon inventory theory.

In this study, we consider the stochastic joint replenishment problem both in single-location/multi-item and single-item/two-echelon inventory settings.

We begin with a review on the relevant literature of this study in Chapter 2. In Chapter 3, we propose a new class of control policy for the stochastic joint replenishment problem in a single-location/multi-item inventory system. The proposed $(Q, \mathbf{S}, T)$ policy makes use of the advantages of both continuous and periodic review policies in a parsimonious manner. We derive the expressions for the key operating characteristics of the inventory system for both unit and compound Poisson demands.

Chapter 4 presents the results of an extensive numerical study which has been conducted to study the sensitivity of the policy to various system parameters and to assess the performance of the proposed policy over the existing policies in the literature. We have found that the proposed policy provides significant savings over the existing policies for items similar in their cost structures and individual demand rates. The proposed policy achieves its performance levels with parsimony which reduces the computational requirements for the optimization and provides an easy implementation in practice.

The single location model provided also represents a two-echelon supply chain for a single item, where the upper echelon employs cross docking. In Chapter 5, we extend our model to incorporate a single-item, multi-location setting where the upper echelon also holds inventory. We study a policy class under the stochastic joint replenishment problem in a two-echelon divergent inventory system. We propose a general methodology to analyze the considered policy class. The framework we provide is only based on the development of the ordering process by the lower echelon.

Our modeling methodology provides us an analytical tool to investigate various joint replenishment policies under the considered policy class. Chapter 6 presents the detailed analysis for four different joint replenishment policies within the considered policy class and present expressions and approximations for the key operating characteristics of the model under each policy. We also give insights on the behaviour of the operating characteristics of these policies.

Chapter 7, we present the results of the detailed numerical study which assesses the performance of the policies within the considered policy class in a two-echelon divergent inventory system. We provide discussions on the allocation of the costs within the echelons and the comparison of echelon costs across the policies. We also present the advantage of allowing the warehouse to hold stock instead of employing cross-dock at the warehouse.

In the last chapter, some concluding remarks about the study and future research directions are provided. We also provide a table for the notation we use throughout the study in Appendix.

Chapter 2

Literature Review

In this chapter, we provide a review on the relevant literature about this study. In Section 2.1, the literature on the stochastic joint replenishment problem will be provided. Section 2.2 discusses the analytical models on common policies studied in two-echelon divergent inventory systems.

2.1 Literature on SJRP

Although the stochastic joint replenishment problem is practically important, the solution for this problem is notoriously difficult. To our knowledge, Ignall [45] is the only study that attempts to find the structure of the optimal joint replenishment policy with stochastic demand. It has been shown that the optimal policy may have a very complex structure even for two items with zero lead time, due to the dependence between the order quantity of an item and the inventory level of the other at an ordering instance. Based on this finding, one may conjecture that the optimal policy for N items would involve control surfaces defined by the inventory levels of other items considered in the replenishment. Even if the exact structure is found, it would be too complex to compute and implement it in practice. Hence, most of the existing approaches to the problem have been confined to the evaluation of some intuitive policy classes that are relatively easy to compute and implement.

The stochastic joint replenishment problem differs from its deterministic counterpart (JRP) greatly in terms of modeling methodologies and the employed policy structures arising from the deterministic nature of demand. Therefore, the vast body of research on JRP falls outside the scope of this study. We refer the reader to Aksoy and Erenguc [2] and Goyal and Satir [39] for extensive reviews of the works in deterministic demand environments. The literature on the stochastic joint replenishment problem can be classified into two major streams based on the type of policy class under consideration. In our review, we follow this classification.

2.1.1 Can-order Policies

This stream of research has begun with the earliest work on joint replenishment with stochastic demand by Balintfy [15] who introduced the continuous review $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ joint ordering policy - also called the can-order policy. The policy operates as follows. When the inventory position of an item i crosses s_i, a replenishment order is triggered to raise its inventory position to S_i. At the same time, any other item j with an inventory position at or below its can-order point, c_j $(s_j < c_j < S_j)$ is also included in the replenishment, raising its inventory position to S_j. Despite its benign structure, the analytical treatment of the system under this policy is extremely difficult even in the presence of unit Poisson demands. Balintfy [15] only provides an initial insight into the problem with a queuing-based approach. A special case with $\mathbf{c} = \mathbf{S} - \mathbf{1}$ and $\mathbf{s} = \mathbf{0}$ in a 2-item inventory system facing identical unit Poisson demands with zero lead-time has been analyzed by Silver [67]. Under the assumption that shortages are not allowed and with the objective of minimizing ordering and holding costs per unit time, Silver [67] proves that the can-order policy is always better than independent control if the cost of placing an order for two items is equal to that for a single item; and, otherwise, there exists a critical value of the joint ordering cost only below which it is preferable to use joint replenishment. An exact analysis has been possible for this special case because the inventory levels of both items provide regeneration points at the order instances and, hence, the renewal reward theorem is applicable. However, the same approach cannot be used for the general case. Therefore, different approximate models and solution methods have been proposed in the literature.

A common approximation technique proposed by Silver [69] is to decompose the N-item problem with unit Poisson demands into N single-item problems facing unit Poisson demands and Poisson special replenishment opportunities. The resulting single-item problem has been analyzed by Silver [68] and solved optimally by Zheng [80]. The same decomposition technique has later been extended to compound Poisson demand by Thompson and Silver [75] and Silver [70]. Using a similar decomposition approach, Federgruen *et al.* [31] propose a semi-Markov decision model and use a policy-iteration algorithm to solve for the optimal values of the control policy parameters. We denote this policy by $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$. Van Eijs [77] and Schultz and Johansen [65] have illustrated that the decomposition method assuming a Poisson arrival process for the special replenishment opportunities can lead to poor performance of the can-order policies. Instead, they propose using Erlang distributions in the decomposition. The optimal values of the policy parameters are obtained through policy iteration and simulation-based updating of the stochastic process governing the opportunities. Melchiors [53] has proposed to use a new compensation approach and been able to improve the previous approximations of the continuous can-order policies for unit Poisson demands. We denote this policy by $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$. However, the approach and the approximations used require extensive iterative computations and may result in significant deviations from simulated costs in some cases. Recently, Johansen and Melchiors [46] proposed a periodic review version of the can-order policy which performs well when there is high demand variation across the items.

As the above summary indicates, almost all of the works on the can-order policy have focused on alleviating the inherent modeling complexities arising from the nature of the policy class. Another major difficulty with the can-order policy is the size of the optimization problem. For an N-item setting, the continuous review $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ policy employs $3N$ control policy parameters, whereas the periodic review counterpart has $3N + 1$ policy parameters.

For completeness, we also cite Liu and Yuan [52] who study the can-order policy in a two-item inventory system with correlated demand processes, and Van der Duyn Schouten [76] who considers quantity discounts within the framework of can-order policies.

For realistic operating environments, this implies extensive numerical optimization effort. Coupled with the iterative nature of the decomposition techniques developed in the literature, the can-order policy appears to be a prohibitively tedious control policy class. Therefore, a number of researchers have proposed control policies that are more parsimonious (*i.e.* with fewer control policy parameters) and/or easier to model and optimize. We discuss such policies next.

2.1.2 Other Policies

The continuous review $(Q, \mathbf{S})$ policy was first proposed by Renberg and Planche [60], and subsequently studied by Pantumsinchai [58] with Poisson demand. Under the $(Q, \mathbf{S})$ policy, when the aggregate consumption since the previous order reaches Q, all items are raised up to the vector of order-up-to levels, $\mathbf{S}$. The policy employs $N + 1$ policy parameters in an N-item setting. This policy has been motivated by, and is suitable for, environments where the items have to be procured at a pre-determined quantity, such as a truckload size due to transportation limitations. An exact analysis is presented in Pantumsinchai [58] and the numerical findings indicate that the performance of $(Q, \mathbf{S})$ policy *vis a vis* the can-order policy is remarkable for high ordering cost, small number of items and low shortage costs, whereas, the latter performs better only with small ordering costs.

Cheung and Leung [24] study the $(Q, \mathbf{S})$ policy for a two-item inventory system in a replenishment/quality control context and illustrated that the sampling plan in coordinated replenishments is more complex than that of independent replenishments and therefore decreases the cost savings owing due to joint replenishment.

Atkins and Iyogun [4] propose two base-stock periodic review policies for unit Poisson demands, developed on the basis of a lower bound on the cost rate established previously by the authors (Atkins and Iyogun [3]). The first policy P, imposes the same review period length T for all items, and the inventory levels of all items are raised to their order-up-to levels defined by $\mathbf{S}$. The policy employs $N+1$ policy parameters. The second policy MP is a modified periodic policy that utilizes item-specific review period lengths based on the afore-mentioned lower bound; it uses $2N$ policy parameters. Their numerical study indicates that the proposed policies dominate the $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ policy except when the fixed ordering costs are small.

As reported in Pantumsinchai [58], the performance of the MP policy is comparable to that of the $(Q, \mathbf{S})$ policy. An extension of the P policy of Atkins and Iyogun [4] to compound Poisson demand is provided by Fung *et al.* [37] under a service level constraint. They observe that this extension results in significant cost reductions over can-order policy especially when the lead-time is large.

Viswanathan [79] recommends a new policy class. Under the proposed policy, $P(\mathbf{s}, \mathbf{S})$, one uses an independent, periodic review (s, S) policy for each item with a common review interval, T. This policy employs $2N+1$ policy parameters for an N item setting. An approximate solution is provided under the assumption that an order is placed at each review epoch. An extensive comparison of the $P(\mathbf{s}, \mathbf{S})$ policy is made with the MP, $(Q, \mathbf{S})$, $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ policies. It is found that $P(\mathbf{s}, \mathbf{S})$ dominates the other policies especially when the holding costs are high compared to the backorder costs.

Cachon [17] proposes another periodic review policy - called the $(Q, \mathbf{S}|T)$ or minimum quantity periodic review policy. Under $(Q, \mathbf{S}|T)$ policy, the system is reviewed every T time units, and any item j is ordered up to its maximum level S_j if a total of at least Q demands have accumulated for the items. In an N-item inventory setting, the $(Q, \mathbf{S}|T)$ policy employs $N+2$ parameters. Cachon [17] also considered shelf-space and truck capacities for the SJRP.

In a very recent study, Nielsen and Larsen [56] proposed the $Q(\mathbf{s}, \mathbf{S})$ policy in which inventories are reviewed only when Q total demands accumulate since the last review instance. At the review instance, any item j, the inventory position of which is less than or equal to its reorder level s_j, is ordered up to S_j. This policy employs $2N + 1$ policy parameters for an N -item setting. In operating environments with identical demand and cost structures for the items, the policy reduces to the $(Q, \mathbf{S})$ policy. Over a small test bed, the policy was found to be superior to the previously proposed policies.

As the above discussion of the existing policies illustrates, the stochastic joint replenishment problem is an open research area for the development of more efficient computational methods and control policies.

2.2 Literature on Two-Echelon Divergent Inventory Systems

The theory of stochastic multi-echelon inventory models has been essentially developed during the last two decades. For a general overview of this development, we refer to Axsäter [6] and Federgruen [30]. Since there are a vast number of studies in this area, we will restrict ourselves only to the literature on two-echelon divergent inventory systems. Note that in two-echelon divergent systems, each retailer at the lower echelon is supplied from only one stocking point at the upper echelon.

Most of the ordering policies studied in the literature are built around two major policy classes. In our review, we will follow these classes and also mention a few studies that utilize the centralized information in a two-echelon inventory system.

2.2.1 Installation Stock Policies

One of the most common policies used in multi-echelon inventory systems is the installation stock policy. Here, the inventory control is completely decentralized in the sense that the ordering decisions at a certain installation are solely based on installation stock, i.e., the inventory position at this installation and do not require any information about the inventory situation at the other installations.

There are three main approaches for the evaluation of these policies in divergent systems:

1. The first approach is to approximate the effective leadtime time of a retailer order, which consists of a deterministic leadtime and a random waiting time resulting from stock-outs at the warehouse. This approximation is the basis of the approach of Sherbrooke [66] for the METRIC model where each facility employs a one-for-one ordering policy.

2. The second approach is to aggregate all retailers as a single retailer and determine the outstanding orders of this retailer. The outstanding order at this retailer is disaggregated among the retailers which provides the computation of the inventory and backorder levels of the retailers. Using this approach, Simon [71] provided the exact expressions for the METRIC model. Graves [40] used this exact approach to optimize the inventory levels in the system. Graves [40] also provides a two-moment fit for the number of outstanding orders at a retailer.

   Moinzadeh and Lee [55] and Lee and Moinzadeh [51], [50] presented several approximations for the number of outstanding orders and provided optimization procedures for both one-for-one and batch ordering policies.

3. The last approach matches every supply unit with a demand unit. By keeping track of an arbitrary supply unit from the moment it enters the system until it exits by fulfilling a demand, it is possible to calculate the holding and backorder costs associated with this unit.

   This idea first appeared in Svoronos and Zipkin [73] to calculate the average backorders at the retailers and the average inventory level at the retailers and the warehouse under (Q, R) policy at each installation.

Later, Axsäter [5] calculated the holding and backorder cost of an arbitrary unit for the case where one-for-one replenishment policy is employed at each installation. Axsäter [5] also derives lower and upper bounds on the optimal base stock levels. The cost function derived in Axsäter [5] was later used by Axsäter [7] to calculate the cost function of (Q, R) policy with unit Poisson demand and identical retailers. Forsberg [36] extended the analysis to non-identical retailers. Forsberg [35] presented an exact model based on the model developed in Forsberg [36] to analyze the case of Erlang inter-demand times. In Forsberg [35], approximations based on the analysis of Erlang inter-demand times were also presented to analyze the case of more general inter-demand time distributions. This approach was also used by Axsäter [9] to calculate the exact probability distribution of the inventory level of the retailers under (Q, R) policy with compound Poisson demand and identical retailers. With non-identical retailers and compound Poisson demand, Forsberg [34] and Axsäter [8] have used the cost function of Axsäter [5] to provide an exact cost rate function of order-up-to policy and an approximate solution for (Q, R) policy, respectively. More recently, Cachon [16] used this approach to calculate the average inventory, backorders and fill rates for periodic review (R, nQ) policies with discrete batch demand.

2.2.2 Echelon Stock Policies

The cost effectiveness of an installation stock policy is obviously limited due to the lack of information about the entire system. A simple way to eliminate this disadvantage is to incorporate the information about the inventory levels at the lower echelons. The echelon inventory position at an installation is obtained by adding the inventory positions at the installation and all of its downstream installations.

The echelon stock concept was first introduced by Clark and Scarf [25]. They proved that order-up-to policies based on echelon stock are optimal for serial inventory systems under periodic review and ordering costs incurred only at the highest echelon. Rosling [61] proved that the assembly systems can be interpreted as serial systems and hence echelon stock order-up-to policies are also optimal for assembly systems when the ordering costs are zero. Similarly, Axsäter and Rosling [11] have shown that echelon stock (Q, R) policies dominate installation stock (Q, R) policies.

Axsäter and Junnti [12] compared the installation and echelon stock policies through simulation for random demands in a two-echelon divergent inventory system and illustrated that neither policies dominate the other in all settings. On the other hand, Axsäter and Junnti [12],[13] calculated the worst case performance of the installation stock policy compared with echelon stock policy for constant demand case.

Chen and Zheng [22] considered a two-echelon inventory system where each facility operates under an echelon stock (R, nQ) policy. For unit Poisson demand at the retailers, they provide an exact method to compute the average holding and backorder costs in the system. The exact method is based on disaggregating the backorders at the warehouse among the retailers. For compound Poisson demand, they also provide an approximate solution.

2.2.3 Joint Replenishment Policies

To the best of our knowledge, there are a few studies that consider joint ordering decisions in a two-echelon divergent inventory system.

Axsäter and Zhang [14] have proposed a model where the warehouse uses a regular installation stock policy but the retailers employ a new type of policy, (Q_r, R_r). Under the proposed policy, when the sum of the inventory positions decline to a joint reorder point, R_r (the number of demands accumulated in the system reaches Q_r units), the retailer with the lowest inventory position orders a batch quantity, Q_r. The proposed policy, in comparison with installation and echelon stock policies, gives slightly higher costs.

In a more recent study, Cheung and Lee [23] have studied the $(Q, \mathbf{S})$ policy for the retailers. Similar to multi-item setting the policy operates as follows: When the cumulative demands over all the retailers reach Q units, an order is placed at the supplier to replenish the retailer to their maximum levels S_i. Under a continuous review (Q, R) policy employed at the warehouse, they present an exact analysis of the model and give lower and upper bounds for the case where the stock rebalancing is carried out at the retailers.

Observe that, under both of these policies, although the warehouse employs a (Q, R) policy, the material flow in the inventory system is identical to a system where the warehouse operates under an echelon stock policy. The mentioned two policies only differ in the way the ordered units are distributed among the retailers.

Recently, Gurbuz *et al.* [41] proposed a hybrid policy for a two-echelon inventory system with the upper echelon employing cross-dock. The proposed policy is a hybrid combination of the special can-order policy with $\mathbf{c} = \mathbf{s} - \mathbf{1}$ and $(Q, \mathbf{S})$ policy, *ie.* the inventory position of all retailers are raised up to $\mathbf{S}$ whenever any retailer's inventory position drops to s or the number of total demands accumulated at the retailers reaches Q units. The proposed policy is compared with $(Q, \mathbf{S})$, the special can-order policy and a periodic review order-up-to policy.

Lastly, we also cite recent studies by Cetinkaya and Lee [20], Axsäter [10], Cetinkaya and Bookbinder [18], Kiesmüller and de Kok [48], Cetinkaya *et al.* [19], [21] which study different aspects of consolidation policies under VMI programs. These consolidation studies differ from the joint replenishment studies because the consolidation policies let the replenishment orders coming from the retailers wait for a certain time or until a certain quantity is consolidated at the warehouse. The mentioned studies except Kiesmüller and de Kok [48] usually consider the problem from the perspective of the vendor, *ie.* the warehouse and the effect of the consolidation policies on the performance of retailer is ignored.

The above review on the existing policies in divergent inventory system illustrates that the stochastic joint replenishment problem in multi-echelon inventory theory is also an open research area for the development and implementation of new models and policies and analysis of them.

Chapter 3

A New Policy for the SJRP

As explained in the literature review in the previous chapter, the solution of the stochastic joint replenishment problem is extremely difficult. Hence, most of the existing approaches to the problem have been restricted to the evaluation of some intuitive policy classes that are relatively easy to compute and implement. In this chapter, we propose a new class of control policy for the stochastic joint replenishment problem. The $(Q, \mathbf{S}, T)$ policy, proposed herein, makes use of the advantages of both continuous and periodic review policies in a parsimonious manner.

The main assumptions of the model and the proposed policy will be explained in Section 3.1. Section 3.2 presents a preliminary analysis which will be followed by the development of the expressions for the key operating characteristics in Section 3.3. In Section 3.4, we will generalize the proposed policy to the case with compound Poisson demand.

3.1 The Proposed Policy

We consider a continuous review, multi-item inventory system with $N \geq 2$ items facing unit external demands generated by independent and stationary unit Poisson processes with rate λ_i $(i = 1, 2, \ldots, N)$. All unmet demands are assumed to be backordered. Items are supplied from an ample supplier and delivery lead times are constants given by L_i for item i. Although we consider a single-location model, we assume that the lead times may differ across the items since, as indicated in Tersine [74] and Ouyang *et al.* [59], lead time usually consists of the transportation time, which is common for the items and setup and load-unload times, which may be different.

The system is continuously reviewed; and, hence, the records for the last replenishment epoch, as well as the time elapsed since then and the total demand arrived to the system after the last order are all available in the system.

The fixed ordering costs in the system have two components: a common ordering cost, K, which is charged every time a replenishment order is placed and a fixed item specific ordering cost k_i, for item i that is added if item i is included in the order. The common ordering cost, K is associated with the fixed transportation/ordering cost and is independent of the number of items involved in the order. The item specific ordering cost is the cost of adding one more item in the replenishment order and possibly results from reviewing the individual items as well as load-unload processes. This ordering cost structure, so-called *first-order interaction* was first introduced by Balintfy [15] and presents an opportunity to exploit the economies of scale in replenishment by ordering items jointly and, hence, requires an effective coordination mechanism among the items.

Holding cost is charged at h_i per unit of item i held in stock per unit time. Two types of shortage costs are incurred: a time weighted shortage cost at ρ_i per unit backordered of item i per unit time and a fixed penalty cost of π_i for every unit of item i that is not immediately satisfied. We assume that the cost of monitoring the inventory system is negligible and we ignore the unit purchasing costs since all demand is eventually satisfied.

Under the assumed cost structure, the objective is to minimize the expected total cost per unit time. We propose below a joint replenishment policy that unifies the time and the inventory position considerations for the placement of orders. Note that the *inventory position* at any point in time is defined as the on-hand inventory plus on order minus backorders. The proposed policy is formally stated as below:

Policy: *Monitor all inventory positions continuously, and raise the inventory positions of the items up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$
i) whenever a total of Q *demands accumulate for the items or*
ii) at time kT *if at least one demand occurs in* $((k-1)T, kT]$ *with no demand arrivals in* $(0, (k-1)T]$,
whichever occurs first.

We shall refer to this proposed policy as the $(Q, \mathbf{S}, T)$ policy, where $\mathbf{S}$ is the vector denoting the maximum inventory positions of the items, and T and Q correspond, respectively, to the time and inventory triggers. In the sequel, we use the term *decision epoch* to refer to an instance at which either a replenishment order is placed or merely an inventory review is made without any order placement. To clarify the distinction, consider the following cases. Suppose that a total of Q demands have arrived before T time units have elapsed since the last decision epoch; then, an order is placed at the instance of the Q'th demand arrival, which constitutes a decision epoch. Alternatively, suppose that T time units have elapsed before a total of Q demands have arrived. At this instance, the inventory review may or may not result in an order placement. If at least one demand has arrived in T units of time, reordering will occur and the placement of an order constitutes the decision epoch. However, if no demand has arrived within the T units of time, then the decision is not to order anything, and the decision epoch coincides with an inventory position review instance. Thus, we use a decision epoch to refer to an instance at which either a replenishment order is placed or only an inventory review action is taken. Due to the Poisson demand process, we immediately see that decision epochs constitute regenerative instances for the system. We will also elaborate on the implementation of the policy in Section 3.2.

The $(Q, \mathbf{S}, T)$ policy is a hybrid of the continuous review $(Q, \mathbf{S})$ policy, first proposed by Renberg and Planche [60], and the periodic review $(\mathbf{S}, T)$ or P policy of Atkins and Iyogun [4]. Thus, it attempts to exploit the benefits of two separate policies. As expected, it reduces to these two policies in the limit: as $T \to \infty$, we obtain the $(Q, \mathbf{S})$ policy; and, as $Q \to \infty$, we obtain the $(\mathbf{S}, T)$ policy.

The replenishment quantity under the $(Q, \mathbf{S}, T)$ policy is a random variable; it may be as small as one unit and cannot exceed Q units. This is in contrast with the $(Q, \mathbf{S})$ policy, which imposes a constant reorder size. Hence, the $(Q, \mathbf{S}, T)$ policy may not fully exploit the economies of scale in joint ordering in every order instance in comparison with the $(Q, \mathbf{S})$ policy. We have observed this disadvantage in many cases in our numerical results presented in Chapter 4. However, the cause of this diseconomy, namely, the introduction of the time trigger, T, helps in another way and compensates for this inefficiency. Under the $(Q, \mathbf{S})$ policy, the inter-order times are random. To be specific, they have $Erlang\_Q$ distribution, which may have quite long tails. The introduction of T cuts such long tails, as it imposes an upper bound on the time between two consecutive decision epochs (and, thereby, reorder times). Therefore, $(Q, \mathbf{S}, T)$ policy also aims to decrease the variance of the inter-order time. The $(Q, \mathbf{S}, T)$ policy also makes use of the advantages of continuous and periodic review policies by providing opportunities either at demand arrivals or review instances.

Previously, we have indicated that the generation of replenishment opportunity arrivals is crucial in understanding the idea behind SJRP. Under $(Q, \mathbf{S})$ policy, the internally generated joint replenishment opportunities arrive in a non-Markovian fashion (*e.g.* time between two consecutive opportunities is $Erlang_Q$ distributed). The presence of a time-based reorder trigger provides the opportunity of pro-active reordering in the presence of non-Markovian total demand process/replenishment opportunity arrivals. We know from Katircioglu [47] that a time-based reorder trigger is optimal for single-location models with non-Markovian demands (see also Moinzadeh [54] and Tekin *et al.* [28]).

Time trigger also provides a check against the excessive imbalances of demands across the items. To see this, consider a hypothetical case when we have, say, $Q-3$ total demand arrivals since the last decision epoch. It may be the case that all of those demands have come for only one item, say j. The inventory level of item j may then be dangerously low - we may even be experiencing shortages.

If we were using the $(Q, \mathbf{S})$ policy, item j would have to wait for three more demands to arrive to the system to give its order. However, if we are using the $(Q, \mathbf{S}, T)$ policy, there is the possibility that T time units since the last decision epoch will have elapsed much before the arrival of those next three demands to the system, and item j will give its order at the time trigger. This will protect item j against shortages better than the $(Q, \mathbf{S})$ policy. If after T time units since the last review instance or the replenishment order, an order has not been placed yet, *i.e.*, Q demands have not accumulated, the policy places an order for the items in anticipation of the placement of a possible near future order. By doing so, the items can be replenished in a more reliable way to handle for the leadtime uncertainty and to protect against shortages. Hence, we would expect the introduction of T to improve the $(Q, \mathbf{S})$ policy.

Next, we present some preliminary results needed to derive the operating characteristics of the system.

3.2 Preliminary Analysis

In this section, we obtain two entities: the joint distribution of the order size and the inter-order time; and the steady-state distribution of the individual inventory positions of the items.

First, we introduce some notation. Let r_i be the probability that the demand is for item i, given that a demand arrival has occurred. Since the demand process is Poisson, $r_i = \lambda_i/\lambda_0$, where $\lambda_0 = \sum_{j=1}^{N} \lambda_j$ is the system demand rate. Let $X_n, n = 1, 2, \ldots$, denote the random variable representing the arrival time of the nth system demand after the last decision epoch which could be either a demand instance or a time trigger. Since inter-arrival times of the demands are exponential, the time until next demand (forward recurrence time for the demand process since the last decision epoch) is also exponential and therefore X_n has an $Erlang\_n$ distribution with scale parameter λ_0. Let $f(x, k, \lambda)$ and $F(x, k, \lambda)$ be the probability density and the cumulative distribution functions of an Erlang random variable with shape and scale parameters k and λ, respectively. For any cumulative distribution function F, we use $\overline{F} = 1 - F$.

Under the $(Q, \mathbf{S}, T)$ policy, we define a *cycle* as the time between two consecutive order placement decisions. A cycle starts every time a positive replenishment order is given (raising the inventory positions to $\mathbf{S}$). Under the proposed policy, there may be multiple decision epochs, separated by intervals of length T within a cycle. We denote the total number of such decision epochs by M, which is a geometric random variable. We present two realizations of the evolution of a cycle in Figure 3.1.

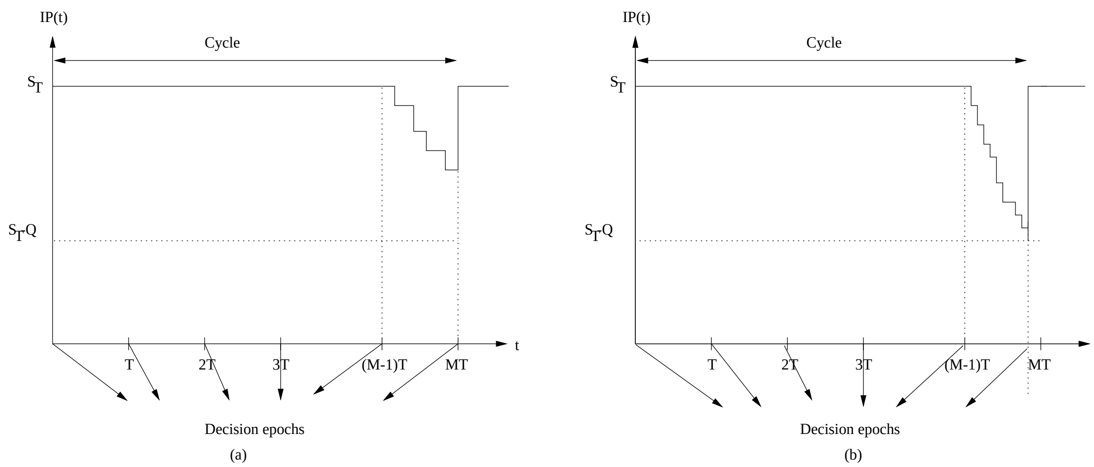

Figure 3.1: Realizations for a cycle

Figure 3.1(a) refers to a realization where, in the first $(M-1) \geq 0$ intervals of length T since the last order placement decision, no demand has arrived and in the next interval of length T, less than Q but more than one demands have arrived to the system, triggering a reorder decision based on the time threshold. Hence, the length of the cycle is MT. Figure 3.1(b) refers to a realization where, in the first $(M-1)$ intervals of length T since the last order placement decision, no demand has arrived as in Figure 3.1(a), but before T more time units elapse, Q demands arrive, triggering a replenishment. Hence, the length of the cycle is random with a value between $(M-1)T$ and MT. As mentioned above, M is a random variable which is geometrically distributed, with parameter $\phi_0 = p_0(0, \lambda_0 T)$, where $p_0(x, \lambda)$ denotes the probability mass function of a Poisson random variable at x, with rate λ.

For clarity and later use, we make the following definitions. Let $IP_i(t)$ denote the inventory position of item i and $IP(t)$ denote the total inventory position of the system at time t. Then, $IP(t) = \sum_{i=1}^{N} IP_i(t) \leq \sum_{i=1}^{N} S_i = S_T$. Also let $NI_i(t)$ denote the net inventory level of item i at time t. In order to illustrate the behavior of the inventory system under the proposed policy, we depict a particular realization in Figure 3.2. Figures 3.2(a) and 3.2(b) show the inventory positions and net inventory of item 1 and item 2, respectively. Figure 3.2(c) displays the corresponding total inventory position. In the following, we briefly narrate the time sequence of the events and the decisions taken. In this illustration, we have $S_1 = 5, S_2 = 3$, $Q = 3$ and some $T > 0$ as the policy parameters; initially both items are at their maximum stocking levels. For generality, we assume that lead times for individual items are different. That is, an order consisting of units for both items will be received at different times by the two items. We assume $L_1 > L_2 > 0$. At time $t = t_1$, a demand arrives for item 1, at $t = t_2$, a demand arrives for item 2 and at time $t = t_3 (< T)$, another demand arrives for item 1. At this instance, the number of demands accumulated in the system reaches $Q = 3$. This triggers an order placement at $t = t_3$ which brings the inventory position of item 1 to S_1 and of item 2 to S_2. This order consists of three units, two of which are for item 1 and the remaining one unit is for item 2. At this point, there is one outstanding order in the system and both items are awaiting some delivery. At time $t_4 = t_3 + L_2$, the unit for item 2 in the order placed at t_3 arrives, raising the net inventory of item 2 to three. At time t_5, a demand arrives for item 1 and drops its inventory position to four and its net inventory to two (since item 1 is still awaiting its delivery). At time $t_6 = t_3 + T$, a total of T time units have elapsed since the last order was placed; therefore, an order is placed as triggered by the policy. The order size is one and only item 1 is included in this order since no demand has arrived for item 2 between $t = t_3$ and $t = t_6$. At time t_7, another demand arrives for item 1 decreasing its inventory position to four and its net inventory to one. Note that, between t_6 and $t_8 = t_3 + L_1$, there are two outstanding orders for item 1 whereas there is no outstanding order for item 2. At time $t = t_8$, the units in the order given at time t_3 are received by item 1 and its net inventory is raised to three. A demand for item 2 arrives at time $t = t_9$ dropping both the inventory position and net inventory to two. At time $t_{10} = t_6 + T$, another order is placed; its order size is two, one unit for each item. At t_{10}, there are two outstanding orders for item 1 and one outstanding order for item 2. The process goes on further.

Let Y and Q_0 denote random variables corresponding to the cycle length (*i.e.* the inter-order time) and the order size, respectively. For convenience, we shall use the term *joint density* for joint density/probability mass function of random vectors when some components are discrete and others are continuous random variables. Let $f_{Y,Q_0}(y,q)$ denote the joint probability density function of Y and Q_0. We have the following result as proved in the Appendix.

Lemma 3.2.1

$$f_{Y,Q_0}(y,q) = \begin{cases} \phi_0^{m-1} p_0(q, \lambda_0 T) & \text{if } y = mT, m \geq 1, 0 < q < Q \\ \\ \phi_0^{m-1} f(y-(m-1)T, Q, \lambda_0) & \text{if } (m-1)T < y < mT, m \geq 1, q = Q \end{cases}$$

Proof: *See Appendix.*

Using the above lemma, we can find the marginals, which will be of use in the sequel.

Corollary 3.2.1

(a) The probability mass function $P_{Q_0}(q) = P(Q_0 = q)$ *of* Q_0 *is given by:*

$$P_{Q_0}(q) = \begin{cases} p_0(q, \lambda_0 T)/(1-\phi_0) & \text{if} \quad 0 < q < Q \\ \\ \overline{P}_0(Q-1, \lambda_0 T)/(1-\phi_0) & \text{if} \quad q = Q \end{cases}$$

(b) The p.d.f., $f_Y(y)$, *of* Y *is given by:*

$$f_Y(y) = \begin{cases} \phi_0^{m-1}[P_0(Q-1, \lambda_0 T) - \phi_0] & \text{if} \quad m \geq 1, y = mT \\ \\ \phi_0^{m-1} f(y-(m-1)T, Q, \lambda_0) & \text{if} \quad m \geq 1, (m-1)T < y < mT \end{cases}$$

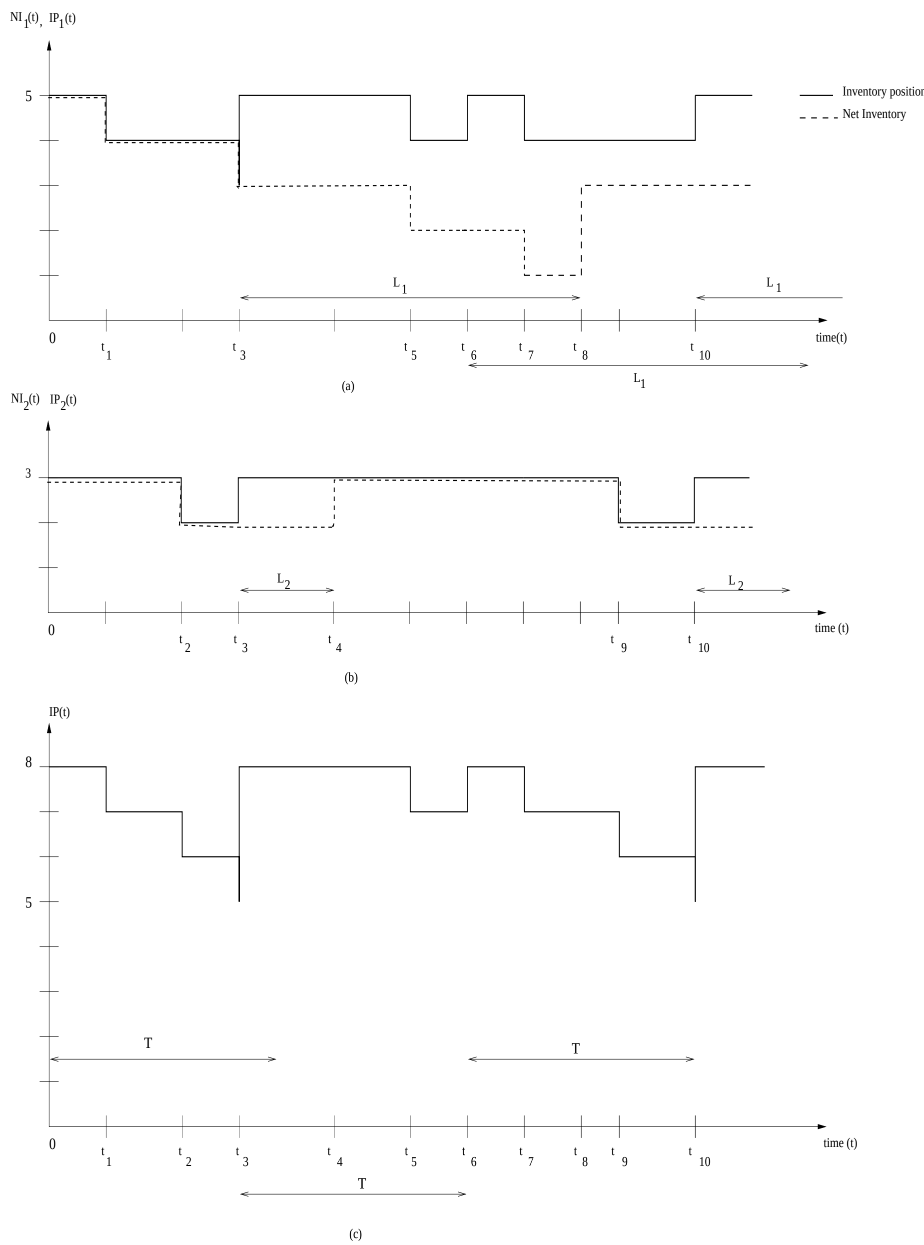

Figure 3.2: Evolution of Ordering Process

where $P_0(x, \lambda)$ denotes the Poisson cumulative distribution function with rate λ.
Proof: *See Appendix.*

The next step is to obtain the steady-state distribution of inventory positions of the items.

As already mentioned in Section 3.1, each decision epoch is a regeneration point for the system, since the inventory positions of all the items are at their base-stock levels at these instances under the $(Q, \mathbf{S}, T)$ policy. Referring to Stidham [72], we know that the steady-state distributions of the inventory positions of items exist.

For $t > 0$ and $1 \leq i \leq N$, define the three-dimensional stochastic process, $\xi_i(t) = \{N_i(t), N_0(t), Z(t)\}$, where $Z(t)$ denotes the time elapsed at time t since the last decision epoch, and $N_i(t)$ and $N_0(t)$ denote, respectively, the number of demands for item i and for all other items that have arrived over $Z(t)$ time units. A particular state that $\xi_i(t)$ visits at time t will be denoted by $\{n_i, n_0, z\}$. Then, $g_i(t, n_i, n_0, z)$ denotes the probability density function of $\xi_i(t)$. Assuming that a steady state density exists, we have the following result:

Proposition 3.2.1 *The steady state p.d.f., denoted by $g_i(n_i, n_0, z)$ is given by the following expression:*

$$g_i(n_i, n_0, z) \quad = \quad C_0 p_0(n_i, \lambda_i z) p_0(n_0, (\lambda_0 - \lambda_i) z) \tag{3.1}$$

for $0 < z \leq T$ and $0 \leq n_0 + n_i \leq Q - 1, n_0 \geq 0, n_i \geq 0$, where C_0 is the normalizing constant given by

$$C_0 = \left[\int_{t=0}^{T} P_0(Q-1, \lambda_0 t) dt \right]^{-1}$$

Proof: *See Appendix.*

Due to the nature of the control policy which ensures constant inventory positions at decision epochs, there is a one-to-one correspondence between the observed demands and the inventory positions of items. If n_i demands have arrived for item i after the last decision epoch, the inventory position of item i

is $S_i - n_i$. Hence, from Proposition 3.2.1, we can immediately obtain the steady-state distribution of the inventory position of item i.

Proposition 3.2.2 *Let $\varphi_i(x)$ denote the steady-state probability that the inventory position of item i is x. Then,*

$$\varphi_i(S_i - n_i) = \frac{C_0}{\lambda_0} \sum_{k=0}^{Q-1-n_i} \binom{k+n_i}{n_i} r_i^{n_i}(1-r_i)^k F(T, k+n_i+1, \lambda_0)$$

for $0 \le n_i \le Q-1$.

Proof: *Using Proposition 3.2.1, we have,*

$$\begin{aligned}
\varphi_i(S_i - n_i) &= \sum_{k=0}^{Q-1-n_i} \int_{z=0}^{T} g_i(n_i, k, z)dz \\
&= C_0 \sum_{k=0}^{Q-1-n_i} \int_{z=0}^{T} p_0(n_i, \lambda_i z) p_0(k, (\lambda_0 - \lambda_i)z)dz \\
&= C_0 \sum_{n_0=0}^{Q-1-n_i} \int_{z=0}^{T} \frac{e^{-\lambda_i z}(\lambda_i z)^{n_i}}{n_i!} \frac{e^{-(\lambda_0-\lambda_i)z}((\lambda_0 - \lambda_i)z)^k}{k!} dz \\
&= C_0 \sum_{k=0}^{Q-1-n_i} \frac{\lambda_i^{n_i}(\lambda_0 - \lambda_i)^k}{\lambda_0^{n_i+k+1}} \frac{(k+n_i)!}{k!n_i!} \int_{z=0}^{T} \lambda_0 \frac{e^{-\lambda_0 z}(\lambda_0 z)^{k+n_i}}{(k+n_i)!} dz \\
&= \frac{C_0}{\lambda_0} \sum_{k=0}^{Q-1-n_i} \binom{k+n_i}{n_i} r_i^{n_i}(1-r_i)^k F(T, k+n_i+1, \lambda_0)
\end{aligned}$$

Now, we are ready to formulate the operating characteristics of the inventory system.

3.3 Operating Characteristics

In this section, we derive the expressions for the expected cycle length, the order placement rate, the probability that a particular item is included in a replenishment order, and, the expected values of the steady state on-hand inventory and backorder levels. These expressions are then used to construct the expected cost rate function.

We begin with expected cycle length, $E[Y]$. As detailed in the Appendix, we have:

$$E[Y] = \frac{TP_0(Q-1,\lambda_0 T)}{1-\phi_0} + \frac{Q\overline{P}_0(Q,\lambda_0 T)}{\lambda_0(1-\phi_0)} \tag{3.2}$$

In each cycle, the common ordering cost is incurred once. Hence, the common ordering cost rate is simply $K/E[Y]$. In each replenishment, item specific ordering costs are also incurred. To obtain the item specific ordering cost rate, one needs to find the items that are included in any given order. The probability that item i is included in an order of size q $(1 < q < Q\ )$ is $1-(1-r_i)^q$, where $r_i = \lambda_i/\lambda_0$ as defined before. Letting θ_i denote the probability that item i is included in a replenishment order, we have

$$\theta_i = \sum_{q=1}^{Q} P_{Q_0}(q)[1-(1-r_i)^q] \tag{3.3}$$

where $P_{Q_0}(q)$ is given in Corollary 3.2.1.

To compute the expected on-hand inventory level and the expected number of backorders at any time, we employ the standard argument of Hadley and Whitin [42] as follows: Consider the system at time instances t and $t+L_i$, where L_i is the constant replenishment leadtime of item i. Note that all outstanding orders at time t will have arrived in the system by time $t+L_i$ and nothing on order at time t will have arrived by time $t+L_i$. Then, the on-hand inventory of item i, $OH_i(t+L_i)$, and the backorder level of item i, $BO_i(t+L_i)$ at time $t+L_i$ can be written as:

$$OH_i(t+L_i) = \max(IP_i(t) - D_i(t,t+L_i],0) \tag{3.4}$$

$$BO_i(t+L_i) = \max(D_i(t,t+L_i] - IP_i(t),0) \tag{3.5}$$

Here, $D_i(t,t+L_i]$ is the number of demands arriving for item i during $(t,t+L_i]$ and has a Poisson distribution with rate $\lambda_i L_i$. Notice that since the demand is Poisson, $D_i(t,t+L_i]$ is independent of $IP_i(t)$.

In view of Equations (3.4)-(3.5), we can find the steady state inventory levels at time $t+L_i$ by conditioning on the steady state distribution of the inventory position at time t.

At steady state, we have the probability mass function of on-hand inventory level OH_i and backorder level, BO_i as follows:

$$P(OH_i = y_i) = \sum_{n_i=S_i-Q+1}^{\min(S_i,y_i)} \varphi_i(n_i)p_0(n_i - y_i, \lambda_i L_i) \quad 0 \leq y_i \leq S_i \tag{3.6}$$

$$P(BO_i = y_i) = \sum_{n_i=S_i-Q+1}^{S_i} \varphi_i(n_i)p_0(n_i + y_i, \lambda_i L_i) \quad y_i \geq 0 \tag{3.7}$$

Hence, at steady state, we have $E[OH_i]$ and $E[BO_i]$ as follows:

$$E[OH_i] = \sum_{y_i=1}^{S_i} y_i P(OH_i = y_i) \tag{3.8}$$

$$E[BO_i] = \sum_{y_i=1}^{\infty} y_i P(BO_i = y_i) \tag{3.9}$$

The steady state probability that there is no stock on hand of item i, ψ_i is given as follows:

$$\psi_i = 1 - \sum_{y_i=1}^{S_i} P(OH_i = y_i) \tag{3.10}$$

We can now construct the expected cost rate $AC(Q, \mathbf{S}, T)$ for the whole system using Equations (3.2) - (3.10).

$$AC(Q, \mathbf{S}, T) = \frac{K + \sum_{i=1}^{N} k_i\theta_i}{E[Y]} + \sum_{i=1}^{N} h_i E[OH_i] + \sum_{i=1}^{N} \rho_i E[BO_i] + \sum_{i=1}^{N} \pi_i \lambda_i \psi_i \tag{3.11}$$

Then, the optimization problem is defined as follows:

$$\begin{gathered} \min_{Q,\mathbf{S},T} AC(Q, \mathbf{S}, T) \\ \text{s.t.} \\ Q \in \mathcal{Z}^+, \mathbf{S} \in \mathcal{Z}^N, T > 0 \end{gathered} \tag{3.12}$$

Although an explicit expression is provided in Proposition 3.2.1 for the steady state distribution of inventory positions, the complicated nature of the expressions for the operating characteristics does not allow for an analytical investigation of the unimodality or the convexity of the objective function. We comment on the numerical observations about this issue in Chapter 4.

3.4 Extension to Compound Poisson Demand

Although unit Poisson demand assumption is commonly made in inventory models, the Poisson distribution may exhibit a poor fit to demand data in certain environments since it may not capture the variability of demands sufficiently. But the data may have a coefficient of variation significantly greater than that of a Poisson random variable with the appropriate mean. In this section, we, therefore, extend our results to a more general setting where items face batch demands that arrive according to a Poisson process but with a random batch size which is independent of the arrivals. Specifically, we assume that customers who demand item i arrive according to a Poisson process with rate λ_i and demand x units of item i with probability $v_i(x)$, for $i = 1, 2, \cdots, N$ and $x = 1, 2, \cdots$. Let $v_i^{(k)}(x)$, $k = 1, 2, \cdots$ denote the probability that x units of item i have been demanded by k customers who arrived for item i. Incidentally, $v_i^{(k)}(x)$ is the kth convolution of the demand size distribution $v_i(x)$. Also let $V_i(x)$ be the distribution function of demand size for item i. We retain all of the other assumptions and the corresponding notation introduced in the previous sections. Additionally, we assume that if the on-hand inventory is not sufficient to satisfy fully an arriving customer's demand, the demand is partially filled with the available stock and the rest is backordered. We propose the following *generalized* $(Q, \mathbf{S}, T)$ policy:

Policy: *Monitor all inventory positions continuously, and raise the inventory positions of the items up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$ *whenever*
i) the total inventory position of the items crosses $\sum_{i=1}^{N} S_i - Q$ *or*
ii) at time kT *if at least one demand occurs in* $((k-1)T, kT]$ *with no demands in* $(0, (k-1)T]$,
whichever occurs first.

We call this policy as *generalized* because there are two fundamental differences between the unit and compound Poisson demand cases: *(i)* the order size with compound Poisson demand may now exceed Q units since the total inventory position is allowed to cross $\sum_{i=1}^{N} S_i - Q$ whereas it is limited by Q for unit Poisson demand, and *(ii)* the number of units demanded in a replenishment cycle may not be equal to the number of customer arrivals since each customer may demand more than one unit of an item.

The derivation of the expressions for the operating characteristics for the compound Poisson case is based on the methodology used for the unit Poisson demands but is modified slightly to account for the mentioned differences as explained below.

Let $\mathcal{N}$ denote the set of all the items comprising the inventory system, and Θ denote a subset of $\mathcal{N}$. Also let $w_\Theta(q,k)$ be the probability that k customers demand a total of q units for the items in the set Θ. Then, for $q \geq k \geq 1$, $\Theta = \{i\}$, $i = 1, 2, \cdots, N$, $w_\Theta(q,k) = v_i^{(k)}(q)$. For $q \geq k \geq 1$, $\Theta = \mathcal{N}$, we have

$$w_\Theta(q,k) \;=\; \sum_{\left\{\begin{array}{c}\sum_{i=1}^N x_i = k\\ \sum_{i=1}^N q_i = q\end{array}\right\}} \frac{k!}{x_1!x_2!\ldots x_N!} r_1^{x_1} r_2^{x_2} \ldots r_n^{x_N} v_1^{(x_1)}(q_1) v_2^{(x_2)}(q_2) \ldots v_N^{(x_N)}(q_N)$$

and for $q \geq k \geq 1$, $\Theta = \mathcal{N}\backslash\{i\}$, $i = 1, 2, \cdots, N$, we can write

$$w_\Theta(q,k) \;=\; \sum_{\left\{\begin{array}{c}\sum_{j\neq i} x_j = k\\ \sum_{j\neq i} q_j = q\end{array}\right\}} \frac{k!}{x_1!..x_{i-1}!x_{i+1}!..x_N!} \prod_{j\neq i} \acute{r}_j^{x_j} v_j^{(x_j)}(q_j) \tag{3.13}$$

where $\acute{r}_j = \lambda_j/(\lambda_0 - \lambda_i)$ for $j \neq i$. Observe that, for unit Poisson demand, we have $w_\Theta(q,k) = 1$ only for $q = k$.

Now, let $\tilde{p}_0(q, \lambda_\Theta z, \Theta)$ be the probability that a total of q units are demanded of items in set Θ in z time units by the customers arriving according to a compound Poisson process with rate λ_Θ $(= \sum_{i\in\Theta} \lambda_i)$ and batch size with p.m.f. given by $w_\Theta(q,k)$. Then,

$$\tilde{p}_0(q, \lambda_\Theta z, \Theta) = \sum_{k=0}^{q} p_0(k, \lambda_\Theta z) w_\Theta(q,k).$$

The joint probability density function of Y and Q_0 for the compound Poisson demand case can now be expressed as follows.

Lemma 3.4.1

$$f_{Y,Q_0}(y,q) = \begin{cases} \phi_0^{m-1}\tilde{p}_0(q,\lambda_{\mathcal{N}}T,\mathcal{N}) & \text{if} \quad y = mT, m \geq 1, 0 < q < Q \\ \phi_0^{m-1}\sum_{k=0}^{Q-1}\sum_{j=k}^{Q-1} f(y-(m-1)T, k+1, \lambda_0) w_{\mathcal{N}}(j,k)\left[\sum_{i=1}^{N} r_i v_i(q-j)\right] & \\ & \text{if} \quad (m-1)T < y < mT, m \geq 1, q \geq Q \end{cases}$$

Proof: *See Appendix.*

We next present the marginal distributions of Y and Q_0, without giving the proof, which are directly obtained from Lemma 3.4.1.

Corollary 3.4.1 *Under compound Poisson demand,*
(a) The probability mass function $P_{Q_0}(q) = P(Q_0 = q)$ of Q_0 is given by:

$$P_{Q_0}(q) = \begin{cases} \tilde{p}_0(q,\lambda_{\mathcal{N}}T,\mathcal{N})/(1-\phi_0) & \text{if} \quad 0 < q < Q \\ \sum_{k=0}^{Q-1}\sum_{j=k}^{Q-1}\overline{P}_0(k,\lambda_0 T) w_{\mathcal{N}}(j,k)\left[\sum_{i=1}^{N} r_i v_i(q-j)\right]/(1-\phi_0) & \text{if} \quad q \geq Q \end{cases}$$

(b) The p.d.f., $f_Y(y)$, of Y is given by:

$$f_Y(y) = \begin{cases} \phi_0^{m-1}\sum_{q=1}^{Q-1}\tilde{p}_0(q,\lambda_{\mathcal{N}}T,\mathcal{N}) & \text{if} \quad m \geq 1, y = mT \\ \phi_0^{m-1}\sum_{k=0}^{Q-1}\sum_{j=k}^{Q-1} f(y-(m-1)T, k+1, \lambda_0) w_{\mathcal{N}}(j,k)\left[\sum_{i=1}^{N} r_i \overline{V}_i(Q-1-j)\right] & \\ & \text{if} \quad m \geq 1, (m-1)T < y < mT \end{cases}$$

Using Corollary 3.4.1(b), $E[Y]$ can be written as:

$$\begin{aligned} E[Y] &= \sum_{q=1}^{Q-1}\tilde{p}_0(q,\lambda_{\mathcal{N}}T,\mathcal{N})\frac{T}{(1-\phi_0)^2} \\ &+ \sum_{k=0}^{Q-1}\sum_{j=k}^{Q-1}\left[\begin{array}{l} \left[\frac{(k+1)}{\lambda_0(1-\phi_0)}F(T,k+2,\lambda_0) + \frac{T\phi_0}{(1-\phi_0)^2}F(T,k+1,\lambda_0)\right] \\ w_{\mathcal{N}}(j,k)\left[\sum_{i=1}^{N} r_i\overline{V}_i(Q-1-j)\right] \end{array}\right] \end{aligned} \quad (3.14)$$

We next present the steady state p.d.f. of $\xi_i(t)$. Recall that $\xi_i(t)$ is a three dimensional stochastic process, $\{N_i(t), N_0(t), Z(t)\}$, where $Z(t)$ denotes the time elapsed at time t since the last decision epoch, and $N_i(t)$ and $N_0(t)$ denote, respectively, the number of demands for item i and for all other items that have arrived over $Z(t)$ time units.

Analogous to the unit Poisson demand case, we have the following result:

Proposition 3.4.1 *The steady state p.d.f. of the stochastic process $\xi_i(t)$ is given as follows:*

$$g_i(n_i, n_0, z) = C_1 \tilde{p}_0(n_i, \lambda_{\{i\}} z, \{i\}) \tilde{p}_0(n_0, \lambda_{\{\mathcal{N}\setminus\{i\}\}} z, \{\mathcal{N}\setminus\{i\}\})$$

for $0 < z < T$ and $0 \leq n_0 + n_i \leq Q-1, n_0 \geq 0, n_i \geq 0$, $i = 1, 2, \cdots, N$ where C_1 is the normalizing constant given by

$$C_1 = \left[\sum_{n_0=0}^{Q-1} \sum_{n_i=0}^{Q-1-n_0} \int_{z=0}^{T} \tilde{p}_0(n_i, \lambda_{\{i\}} z, \{i\}) \tilde{p}_0(n_0, \lambda_{\{\mathcal{N}\setminus\{i\}\}} z, \{\mathcal{N}\setminus\{i\}\}) dz \right]^{-1}$$

Proof: *See Appendix.*

Using the fact that the inventory positions of the items are at their order-up-to levels at the ordering instances, we obtain, as before, the steady state distribution of the inventory position of item i from proposition 3.4.1:

Proposition 3.4.2

$$\varphi_i(S_i - n_i) = C_1 \sum_{n_0=1}^{Q-1-n_i} \int_{z=0}^{T} \tilde{p}_0(n_i, \lambda_{\{i\}} z, \{i\}) \;\; \tilde{p}_0(n_0, \lambda_{\{\mathcal{N}\setminus\{i\}\}} z, \{\mathcal{N}\setminus\{i\}\}) dz$$

for $0 \leq n_i \leq Q-1, i = 1, 2, \cdots, N$.

Finally, for $i = 1, 2, \cdots, N$, we can write

$$P(OH_i = y_i) = \sum_{n_i=S_i-Q-1}^{\min(S_i, y_i)} \varphi_i(n_i) \tilde{p}_0(n_i - y_i, \lambda_{\{i\}} L_i, \{i\}) \quad 0 \leq y_i \leq S_i \quad (3.15)$$

$$P(BO_i = y_i) = \sum_{n_i=S_i-Q-1}^{S_i} \varphi_i(n_i) \tilde{p}_0(n_i + y_i, \lambda_{\{i\}} L_i, \{i\}) \quad y_i \geq 0 \quad (3.16)$$

Note that the results of Lemma 3.4.1, Propositions 3.4.1 and 3.4.2 and the expressions in Equations (3.15) and (3.16) for the compound Poisson demand case are similar to those given in Lemma 3.2.1, Propositions 3.2.1 and 3.2.2 and the expressions in Equations (3.6) and (3.7) for the unit Poisson demand case except for the modified probabilities. We note that the equations provided in this section reduce to the ones of unit Poisson demand given in Sections 3.2 and 3.3 for $v_i(1) = 1$.

Since the $(\mathbf{S}, T)$ and $(Q, \mathbf{S})$ policies are special cases of the $(Q, \mathbf{S}, T)$ policy, the above generalization provides the compound Poisson demand counterparts of these policies, as well.

Chapter 4

Numerical Results for $(Q, \mathbf{S}, T)$ Policy

In the previous chapter, we have proposed a new parsinomious policy for the stochastic joint replenishment problem in a single-location multi-item setting. In this chapter, our aim is to discuss the computational results regarding the proposed $(Q, \mathbf{S}, T)$ policy. We first present our results for unit Poisson demand and then provide some results regarding the extension to compound Poisson.

In Section 4.1, we point out some issues regarding the behaviour of $AC(Q, \mathbf{S}, T)$ with respect to decision variables and the search algorithm employed. In Section 4.2 we discuss the sensitivity of the optimal policy parameters with respect to various cost and system parameters. Section 4.3 presents the performance of the proposed $(Q, \mathbf{S}, T)$ policy with unit Poisson demand over a wide range of experimental setting. The experimental test beds used include the standard one which was previously used in the literature for comparison of any proposed stochastic joint replenishment policy as well as new test beds to illustrate the impact of different system parameters on the performance of the proposed policy. In Section 4.4, we include a discussion on the performance of the proposed policy under compound Poisson demand.

4.1 Computational Issues

Before we proceed with the results of our numerical study, we note that, we use two types of search algorithms in this chapter, the details of which will be explained below. In this section, we also present some remarks on the behavior of the $AC(Q, \mathbf{S}, T)$.

In view of the optimization problem (3.12) presented in Section 3.3, under $(Q, \mathbf{S}, T)$ policy, it is easy to observe that for a given (Q, T) pair, the optimization problem to find $\mathbf{S}^*$ can be decomposed into N independent sub-problems in each of which we solve for S_i^* separately. This separability property greatly reduces the complexity of the optimization problem.

In a preliminary study, we investigated and observed the unimodality of $AC(Q, \mathbf{S}, T)$ through an iterative search algorithm over a broad solution space with randomized initial points. A total of 100 initial points $\hat{Q}$ and $\hat{\mathbf{S}}$ were randomly selected over the following ranges:

$\hat{Q} \in [1, \max(10Q_m, 1000)]$, $\hat{S}_i \in [1, Q_i + 10\lceil \lambda_i L_i \rceil]$ where

$$\begin{aligned} Q_m &= \sqrt{2\lambda_0(K + \sum_{i=1}^{N} k_i) / \sum_{i=1}^{N} r_i h_i} \\ Q_i &= \sqrt{2\lambda_i(Kr_i + k_i)/h_i} \text{ for } i = 1, 2, \ldots, N \end{aligned}$$

and $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Q_m and Q_i values correspond to the optimal order quantities of all and the individual items under EOQ model with corresponding ordering, holding costs and demand rates and provide a basis to determine the search space for the optimal policy parameters (see Pantumsinchai [58] and Golany and Lev-er [38]).

One iteration of our iterative search algorithm consisted of three consecutive optimization problems for one of the policy decision variables while keeping the other two constant.

$$\begin{aligned} \hat{T} &= \text{argmin}_T AC(\hat{Q}, \hat{\mathbf{S}}, T) \\ \hat{\mathbf{S}} &= \text{argmin}_{\mathbf{S}} AC(\hat{Q}, \mathbf{S}, \hat{T}) \\ \hat{Q} &= \text{argmin}_Q AC(Q, \hat{\mathbf{S}}, \hat{T}) \end{aligned}$$

The iterative algorithm starts with a randomly selected $\hat{Q}$ and $\hat{\mathbf{S}}$, and ends either when the same policy parameter values are obtained in two consecutive iterations or the number of iterations reaches 1000. The search space for unimodality investigation consists of $Q \in [Q^{min}, Q^{max}], T \in [T^{min}, T^{max}], S_i \in [S_i^{min}, S_i^{max}]$ for $i = 1, 2, \ldots, N$ with increments of $\Delta_Q = 1, \Delta_T = 0.01, \Delta_{S_i} = 1$ and the boundaries of the search space are given by

$$\begin{aligned}
Q^{min} &= \max(1, Q_m), \quad Q^{max} = \max(10Q_m, 1000) \\
T^{min} &= 0.5Q^{min}/\lambda_0, \quad T^{max} = 1.5Q^{max}/\lambda_0 \\
S_i^{min} &= \min(Q_i, \lceil \lambda_i L_i \rceil), S_i^{max} = Q_i + 10 \lceil \lambda_i L_i \rceil
\end{aligned}$$

These limits for $Q, \mathbf{S}, T$ parameters are chosen so that the search space includes a wide range of parameter sets including extreme values.

In Figure 4.1, we present different realizations of $AC(Q, \mathbf{S}, T)$ with respect to each policy parameter while we keep the other two parameters constant. The figures illustrate that $AC(Q, \mathbf{S}, T)$ is a well behaved function with respect to each parameter and at each iteration of the iterative algorithm described above, the optimal parameter is obtained having searched over a very wide range of the parameter.

In our test problems, we have observed that the solution of the algorithm converged to the same policy parameter values for all 100 starting points. Incidentally, we have never hit the maximum number of iterations. Clearly, this does not guarantee the optimality. However, given the very broad range of the starting points, the optimization search space and the well behaviour of the cost function, the observed convergence can be taken as an experimental indication for unimodality.

Having observed the unimodal property numerically, in the remainder of our numerical studies, we employed an exhaustive search algorithm for finding the optimal parameter values of the proposed policy as outlined below.

For optimization, we employ exhaustive search over a large solution space. The search space consists of $Q \in [Q^{min}, Q^{max}], T \in [T^{min}, T^{max}], S_i \in [S_i^{min}, S_i^{max}]$ for $i = 1, 2, \ldots, N$ with increments of $\Delta_Q = 1, \Delta_T = 0.01, \Delta_{S_i} = 1$ and the

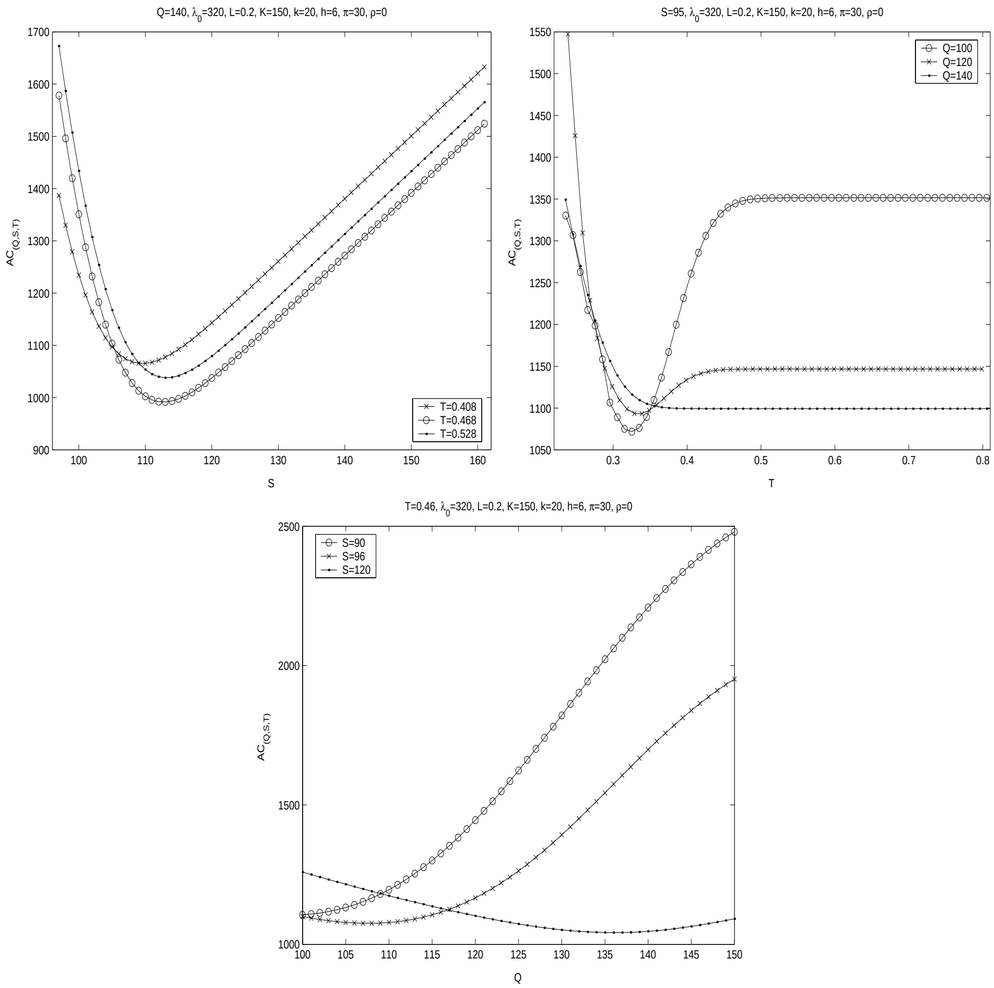

Figure 4.1: Behaviour of $AC(Q, \mathbf{S}, T)$ with respect to each policy parameter

boundaries of the space are given by

$$Q^{min} = \max(1, Q_m), \quad Q^{max} = \max(5Q_m, Q_m + 200),$$
$$T^{min} = 0.5Q^{min}/\lambda_0, \quad T^{max} = 1.5Q^{max}/\lambda_0,$$
$$S_i^{min} = \min(Q_i, \lceil \lambda_i L_i \rceil), \quad S_i^{max} = Q_i + 5\lceil \lambda_i L_i \rceil$$

The employed search algorithm is provided as below:

Search Algorithm:

1.0. Set $Q_m, Q^{min}, Q^{max}, T^{min}, T^{max}$

  2.0. For each $Q \in [Q^{min}, Q^{max}]$ by Δ_Q

    2.1. For each $T \in [T^{min}, T^{max}]$ by Δ_T

      2.1.1. For each item $i \in \{1, 2, \ldots, N\}$

        2.1.1.1. Set $Q_i, S_i^{min}, S_i^{max}$

        2.1.1.2. For each $S_i \in [S_{min}^i, S_{max}^i]$ by Δ_{S_i}

          2.1.1.2.1. Calculate $E[OH_i], E[BO_i], \psi_i$ according to Equations (3.8),(3.9),(3.10).

        2.1.1.3. Set $S_i^* = \text{argmin}\,\{h_i E[OH_i] + \rho_i E[BO_i] + \pi_i \lambda_i \psi_i\}$.

      2.1.2. Compute $E[Y]$ and θ_i for $i = 1, 2, \ldots, N$ according to Equations (3.2)-(3.3).

      2.1.3. Compute $AC(Q, S^*, T)$ with Q given in (2.0) and T given (2.2) according to Equation (3.11).

  3.0. Set $(Q^*, T^*) = \text{argmin} AC(Q, S^*, T)$.

This search algorithm is used to find the optimal policy parameters and optimal cost rate which are presented in the following sections.

4.2 Sensitivity Analysis

In this section, our aim is to illustrate the general behaviour of the optimal policy parameters and the average cost rate of the $(Q, \mathbf{S}, T)$ policy with respect to different cost and system parameters. For the sensitivity analysis, we use an experimental test bed in which $N = 4$ and all items are assumed to be identical in their cost, demand and lead time parameters. Therefore, for the sake of simplicity, we drop the item index i from cost, system and policy parameters. We also consider only unit demands.

Table 4.1 illustrates the experimental set and the results. Notice that the experimental points represent a wide range of parameters from very high service levels with low ordering and high backorder costs to lower service levels with high ordering and low backorder costs. We also note that we consider only one type of backorder costs in our sensitivity analysis.

As explained in the previous chapter, under the proposed policy, there are two reorder trigger mechanisms as discussed above. To better assess the impact and advantage of the time trigger T, of the proposed policy, we also report the probability $\mathcal{H}_T$, that a replenishment order is given by the time trigger. From *Corollary 3.2.1.b*, $\mathcal{H}_T$ can be calculated as follows:

$$\begin{aligned} \mathcal{H}_T &= \sum_{m=1}^{\infty} f_Y(mT) = \sum_{m=1}^{\infty} \phi_0^{m-1}[P_0(Q-1, \lambda_0 T) - \phi_0] \qquad (4.1) \\ &= \frac{1}{1-\phi_0}[P_0(Q-1, \lambda_0 T) - \phi_0] \end{aligned}$$

We present the optimal policy parameters, $(Q^*, \mathbf{S}^*, T^*)$, the optimal cost rate function, AC^*, and $\mathcal{H}_T^*$ calculated at (Q^*, T^*) in Table 4.1 for identical item specific ordering costs, $k_i = 20,\ i = 1, \ldots, 4$.

We observe that the behavior of the policy parameters with respect to system parameters is quite intuitive. We discuss the general findings below in some detail.

The effect of increasing the common ordering cost, K, is to delay the order placement by increasing Q^* and/or T^*, as expected. However, the increase in T^* is usually more pronounced than the change in Q^*. We also note that T^* is, in general, smaller than Q^*/λ_0, which is the average time for Q^* demands to accumulate at the system level. Thus, T^* acts as a proactive trigger. But, as K increases, we lose this property and T^* becomes very close to or larger than Q^*/λ_0. The loss of the proactiveness of the time trigger is also manifest in a decrease in $\mathcal{H}_T^*$ with increasing K. Therefore, the proactive behaviour dominance of placing the orders at the review intervals reduces as K increases. Increasing the common

Parameters				$L = 0.2$					$L = 0.6$				
				Q^*	S^*	T^*	AC^*	$\mathcal{H}_T^*$	Q^*	S^*	T^*	AC^*	$\mathcal{H}_T^*$
$\lambda_0 = 320$	$K = 20$	h=2	$\pi = 30, \rho = 0$	173	75	0.518	463.14	0.703	183	113	0.549	495.50	0.700
			$\pi = 0, \rho = 30$	190	64	0.573	371.20	0.678	198	100	0.597	391.01	0.683
		h=6	$\pi = 30, \rho = 0$	106	55	0.320	866.72	0.626	115	92	0.347	967.83	0.634
			$\pi = 0, \rho = 30$	122	43	0.370	606.42	0.618	129	78	0.390	657.24	0.635
	$K = 50$	h=2	$\pi = 30, \rho = 0$	200	83	0.605	513.21	0.668	207	119	0.625	544.51	0.680
			$\pi = 0, \rho = 30$	216	60	0.658	417.29	0.637	221	105	0.671	435.77	0.670
		h=6	$\pi = 30, \rho = 0$	119	56	0.360	949.11	0.626	127	96	0.384	1045.13	0.633
			$\pi = 0, \rho = 30$	137	46	0.416	679.54	0.599	144	81	0.436	726.69	0.627
	$K = 100$	h=2	$\pi = 30, \rho = 0$	236	90	0.720	583.10	0.626	244	129	0.741	614.02	0.668
			$\pi = 0, \rho = 30$	251	78	0.768	483.58	0.623	260	114	0.790	500.48	0.666
		h=6	$\pi = 30, \rho = 0$	139	63	0.422	1069.96	0.622	150	101	0.456	1158.16	0.621
			$\pi = 0, \rho = 30$	157	50	0.481	785.33	0.587	169	86	0.517	828.01	0.599
	$K = 150$	h=2	$\pi = 30, \rho = 0$	264	97	0.811	645.31	0.601	272	135	0.835	673.21	0.608
			$\pi = 0, \rho = 30$	282	85	0.846	541.20	0.578	291	121	0.901	556.93	0.585
		h=6	$\pi = 30, \rho = 0$	160	68	0.495	1174.19	0.540	166	104	0.509	1258.15	0.586
			$\pi = 0, \rho = 30$	177	54	0.531	877.53	0.520	185	89	0.560	916.96	0.585
$\lambda_0 = 480$	$K = 20$	h=2	$\pi = 30, \rho = 0$	212	97	0.425	566.18	0.703	225	151	0.450	604.33	0.721
			$\pi = 0, \rho = 30$	235	84	0.470	454.55	0.702	243	136	0.486	474.74	0.708
		h=6	$\pi = 30, \rho = 0$	132	72	0.266	1067.17	0.639	142	126	0.284	1188.82	0.676
			$\pi = 0, \rho = 30$	149	58	0.299	743.65	0.636	156	109	0.313	799.07	0.656
	$K = 50$	h=2	$\pi = 30, \rho = 0$	244	104	0.496	628.19	0.641	254	158	0.515	662.74	0.659
			$\pi = 0, \rho = 30$	266	91	0.539	511.29	0.623	274	143	0.552	529.27	0.639
		h=6	$\pi = 30, \rho = 0$	149	76	0.301	1170.13	0.636	156	129	0.316	1284.45	0.641
			$\pi = 0, \rho = 30$	169	62	0.338	833.79	0.608	177	113	0.354	883.75	0.629
	$K = 100$	h=2	$\pi = 30, \rho = 0$	291	116	0.594	716.50	0.628	298	169	0.605	746.93	0.665
			$\pi = 0, \rho = 30$	310	111	0.620	593.00	0.626	315	152	0.642	608.18	0.644
		h=6	$\pi = 30, \rho = 0$	173	83	0.352	1319.24	0.612	182	135	0.368	1423.15	0.647
			$\pi = 0, \rho = 30$	198	68	0.405	964.22	0.593	202	118	0.404	1007.37	0.609
	$K = 150$	h=2	$\pi = 30, \rho = 0$	327	125	0.669	792.16	0.621	335	178	0.670	820.19	0.637
			$\pi = 0, \rho = 30$	353	111	0.723	664.03	0.618	358	162	0.716	677.01	0.602
		h=6	$\pi = 30, \rho = 0$	193	88	0.395	1447.90	0.588	203	140	0.415	1544.54	0.597
			$\pi = 0, \rho = 30$	222	73	0.460	1078.09	0.523	227	123	0.466	1115.97	0.579

Table 4.1: Sensitivity Results with respect to $K, h, \rho, \pi, L, \lambda_0$, $N = 4$ and $k = 20$

ordering cost also results in larger values of S^* so as to avoid stockouts due to the resulting delay in the reordering decision.

As the unit holding cost, h increases, all optimal policy parameters decrease.

That is, reordering decisions are made more frequently. The inventory positions of the items are raised up to lower levels to prevent the increase in the average inventory level resulting from more frequent orders. At the same time, we also observe that $\mathcal{H}_T^*$ generally decreases with h. This implies that the system has a tendency to give the orders which are triggered by the accumulation of Q demands, rather than being made at the time trigger for higher values of h. This is already expected because the average inventory levels increase with the introduction of a finite time trigger and hence the system tries to reduce the effect of increased inventory levels by decreasing T more.

The delivery lead time of the items and the system demand rate (or, equivalently, the individual item demand rates) also have a considerable effect on the optimal policy parameters. As the lead time increases, Q^* and S^* both increase and T^* values usually increase due to increasing Q^*. However, as the system demand rate λ_0 increases, Q^* and $\mathbf{S}^*$ get large but T^* gets smaller. On the other hand, we observe that $\mathcal{H}_T^*$ is generally increasing overall in λ_0 and L. That is, for higher demand rates and/or lead time, the reordering decision is given more frequently by the time trigger. This is also to be expected because longer lead time or larger demand rates increase the risk of stock-outs during lead time, so the proactive option of the policy becomes more desirable and is more often used, which explains the higher values of $\mathcal{H}_T^*$.

When higher service levels are desired, ie. the system works with unit shortage costs, π rather than time weighted backorder costs, ρ, orders are given more frequently and the items are replenished to higher levels, ie. S^* values increase, as expected. We also observe that T^* and Q^* both decrease considerably as the inventory system starts working with unit backorder costs instead of time weighted backorder costs. However, the decrease in T^* is more significant than that of Q^*, so that $\mathcal{H}_T^*$ increases with higher values of service levels. We also observe that, with lower service levels, the decrease in $\mathcal{H}_T^*$ values with increasing K is usually more noticeable. Therefore, it comes out that for the systems working with higher service levels, the proactive behaviour of the proposed policy keeps its importance even with high values of K.

Finally, the optimal cost rate, AC^* is increasing in K, h, π, L and λ_0 across the entire test bed, as expected.

We will next present the performance of the $(Q, \mathbf{S}, T)$ in the following section.

4.3 Comparison with Existing Policies

In this section, we examine the efficacy of the proposed control policy. In particular, we examine the cost improvements achieved by the proposed policy and attempt to identify the operational environments in which it is beneficial to implement the proposed policy in lieu of the existing ones in the literature. Note that all of the available models have been developed only for unit demands and hence this section is also devoted to the comparison of the policies with unit Poisson demands similar to Section 4.2.

For policy comparisons, we introduce the notation below. We let $AC^*_{\mathcal{P}}$ denote the optimal cost rate of a given policy $\mathcal{P}$ where $\mathcal{P}$ can be one of the following: Our proposed $(Q, \mathbf{S}, T)$ policy; $P(\mathbf{s}, \mathbf{S})$ in Viswanathan [79]; $(Q, \mathbf{S})$ in Pantumsinchai [58] (and Ranberg and Planche [60]; the can-order policies, $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$ and $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$, as calculated in Federgruen *et al.* [31] and in Melchiors [53], respectively; and, $Q(\mathbf{s}, \mathbf{S})$ in Nielsen and Larsen [56]. Note that we have excluded the P and MP policies in Atkins and Iyogun [4] since they have previously been shown to be inferior to the aforementioned policies in the literature.

As a measure of the performance of the proposed $(Q, \mathbf{S}, T)$ policy, we use the percentage improvement $\Delta_{\mathcal{P}}\%$ over policy $\mathcal{P}$ as follows:

$$\Delta_{\mathcal{P}}\% \quad = \quad \frac{AC^*_{\mathcal{P}} - AC^*_{(Q,\mathbf{S},T)}}{AC^*_{(Q,\mathbf{S},T)}} \times 100$$

A positive entry for $\%\Delta_{\mathcal{P}}$, by definition, means that the proposed policy dominates policy $\mathcal{P}$. Similarly, a higher value for $\%\Delta_{\mathcal{P}}$ indicates that the $(Q, \mathbf{S}, T)$ policy achieves a higher cost improvement over policy $\mathcal{P}$.

Before we proceed with the results regarding the comparison of the policies, we first clarify some points on how the optimal cost rate of a policy $\mathcal{P}$, $AC^*_{\mathcal{P}}$

values are obtained for each policy.

Among the considered policies, the analysis for the proposed $(Q, \mathbf{S}, T)$ policy and $(Q, \mathbf{S})$ and MP policies are exact. Therefore, the corresponding $AC^*_{\mathcal{P}}$ values are also exact.

For the inventory system operating under the $P(\mathbf{s}, \mathbf{S})$, the model construction is based on the assumption that an order is placed for at least one item at every review interval. Similarly, the model presented in Nielsen and Larsen [56] for $Q(\mathbf{s}, \mathbf{S})$ policy assumes that at least one item is ordered whenever Q demands accumulate on the system. Due to the complicated nature of the can-order policy, the models considered herein, namely, $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$ and $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ are approximations. Hence, the models and the cost functions corresponding to these four policies are only approximations. Consequently, the best policy parameter values for these policies are obtained with the approximate cost functions. An alternative to compute the corresponding true $AC^*_{\mathcal{P}}$ under these policies is to simulate the inventory systems with the given policy parameter values. The simulation results for $AC_{(\mathbf{s},\mathbf{c},\mathbf{S})_F}$ and $AC_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}$ have already been reported in Viswanathan [79] and Melchiors [53], respectively, and were used directly in our numerical study. For the $Q(\mathbf{s}, \mathbf{S}), P(\mathbf{s}, \mathbf{S})$ policies, we solved for the best policy parameters using the approximate cost functions as developed in Viswanathan [79] and in Nielsen and Larsen [56], and then simulated the inventory systems operating under these two policies to obtain the corresponding true $AC^*_{Q(\mathbf{s},\mathbf{S})}$ and $AC^*_{P(\mathbf{s},\mathbf{S})}$. In our simulations, we used a run length of 100,000 ordering instances with a warm-up period of 10,000 order placements, and 100 replications to obtain the corresponding cost figures.

Hence, the $AC^*_{Q(\mathbf{s},\mathbf{S})}$ and $AC^*_{P(\mathbf{s},\mathbf{S})}$ values used for comparison are different from those reported in the corresponding literature. However, we should also mention that in the vast majority of the cases, the difference between the simulated and the approximate cost functions are not discernible.

Our numerical study indicates that the performances of joint replenishment policies and, thereby, the dominance of one over the others depends greatly on the cost and demand rate structures prevalent among the items. Therefore, we present our policy comparisons in two groups.

4.3.1 Atkins-Iyogun and Viswanathan Experimental Test Beds

For the first part of our policy comparisons, we use two test beds. The first one -the Atkins-Iyogun test bed-, consisting of 19 instances, was initially introduced by the authors for their sensitivity study (Atkins and Iyogun [4]) and has subsequently been adopted as a standard test bed for presenting the performance of any proposed stochastic joint replenishment policy. The second one -the Viswanathan test bed [79]- has been developed by the author for comparing the robustness of the $P(\mathbf{s}, \mathbf{S})$ policy against the Atkins-Iyogun policies, and considers a more extensive set of cost parameter combinations (120 instances). These two sets both consider 12 items with the same demand, lead time and item-specific ordering cost values with different combinations of K, h, ρ and π. Note also that the performance of the $Q(\mathbf{s}, \mathbf{S})$ policy over the Viswanathan test set was not reported before in the literature. Hence, the numerical study also provides detailed performance results on this policy for the first time in the literature.

Before we proceed with individual comparisons, we present a summary of our findings over all experiment instances (139 total) in the Atkins-Iyogun and Viswanathan sets. We observed that the proposed $(Q, \mathbf{S}, T)$ policy is the best policy in 100 out of 139 instances with an average improvement of 1.14% and the maximum improvement of 3.55% over the next best policy in these instances. In the remaining 39 cases, $Q(\mathbf{s}, \mathbf{S})$ is the best in 24; $P(\mathbf{s}, \mathbf{S})$ is the best in 8; and, $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ is the best in 7 instances. In the 24 cases where $Q(\mathbf{s}, \mathbf{S})$ is the best, the average improvement over the next best one is 0.86%. The corresponding figures are 0.65% and 0.47% for $P(\mathbf{s}, \mathbf{S})$ and $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ policies, respectively. We also see that MP, $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$ policies are never the best ones over these instances.

Next, we discuss our findings for each test bed separately, beginning with the Atkins-Iyogun test bed. This set consists of 12 items; the items have identical shortage and unit holding costs but differ in their item specific ordering costs, demand rates and delivery lead times. The item-specific costs are as follows: $k_i = \{10, 10, 20, 20, 40, 20, 40, 40, 60, 60, 80, 80\}$, the demand rates are given by $\lambda_i = \{40, 35, 40, 40, 40, 20, 20, 20, 28, 20, 20, 20\}$, and the lead times are taken as $L_i = \{0.2, 0.5, 0.2, 0.1, 0.2, 1.5, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0\}$ for $i = 1, \cdots, 12$. We tabulate the problem parameters common to all items in Table 4.2. In Table 4.2, we also report the corresponding $AC^*_{(Q,\mathbf{S},T)}$ and $\%\Delta_{\mathcal{P}}$ under the five policies considered. Note that, in each experimental instance, the best policy among five is depicted by bold face figures.

Problem Parameters		$AC^*_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}\%$
$\pi = 30, \rho = 0, h = 2$	$K = 50$	1109.90	1.01	5.60	0.27	**-0.09**
	$K = 100$	**1174.21**	0.91	3.22	0.25	3.15
	$K = 150$	1234.12	0.56	1.37	**-0.24**	4.46
	$K = 200$	1282.29	0.22	0.45	**-0.58**	5.69
	$K = 250$	1323.02	0.31	0.00	**-0.46**	6.58
$\pi = 30, \rho = 0, h = 6$	$K = 150$	2279.97	-0.58	1.05	**-1.22**	1.45
$\pi = 0, \rho = 30, h = 2$	$K = 20$	878.91	0.82	8.54	0.36	**-0.80**
	$K = 50$	**928.40**	0.59	5.34	0.07	1.72
	$K = 100$	990.02	0.80	2.62	**-0.40**	4.44
	$K = 150$	1044.04	-0.11	0.76	**-0.77**	5.94
	$K = 200$	1087.17	-0.21	0.02	**-0.84**	6.92
$\pi = 0, \rho = 30, h = 6$	$K = 100$	1635.98	-0.79	2.39	**-1.29**	1.47
	$K = 150$	1717.94	-0.70	0.82	**-1.23**	1.69
	$K = 200$	1786.89	-0.51	0.01	**-1.07**	1.90
$\pi = 0, \rho = 30, h = 20$	$K = 20$	**2294.78**	1.33	10.12	5.93	0.83
	$K = 50$	**2395.45**	1.33	7.70	4.67	1.63
	$K = 100$	**2533.82**	1.10	5.34	3.57	1.46
	$K = 150$	2739.46	**-2.02**	4.13	3.52	-1.35
	$K = 200$	**2721.67**	2.59	0.49	3.48	3.60

Table 4.2: Performance of $(Q, \mathbf{S}, T)$ Policy in the 12-item Atkins-Iyogun Test Bed

We observe that the dominance of the proposed policy is not monotone across the experiment instances. The $(Q, \mathbf{S}, T)$ policy performs better than all other existing policies in 6 out of 19 experiment instances.

For the remaining 13 experiment instances, it is dominated in 10 cases by $Q(\mathbf{s}, \mathbf{S})$, twice by $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$, and once by $P(\mathbf{s}, \mathbf{S})$. We see that the $(Q, \mathbf{S})$ policy is never the best policy. Across the entire Atkins-Iyogun set, the average savings achieved through the implementation of the proposed policy in lieu of each of the existing policies are as follows: 0.35% over $P(\mathbf{s}, \mathbf{S})$, 3.39% over $(Q, \mathbf{S})$ 0.74% over $Q(\mathbf{s}, \mathbf{S})$ and 2.65% over $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$. In the instances where the proposed policy gives the best solution, the average improvement over the next best policy is 1.03% whereas over the instances where the proposed policy is dominated by one of the existing policies, the average deviation from the best solution is −0.77%.

When we examine the experiment instances where $(Q, \mathbf{S}, T)$ and $Q(\mathbf{s}, \mathbf{S})$ are the best two policies, we observe the following: In the instances where $Q(\mathbf{s}, \mathbf{S})$ performs better, the average (under)performance of the $(Q, \mathbf{S}, T)$ policy is −0.66%. In the instances where $(Q, \mathbf{S}, T)$ performs better, its average (over)performance is 2.32%. We also observe that $(Q, \mathbf{S}, T)$ policy achieves lower common ordering, and backorder cost rates than $Q(\mathbf{s}, \mathbf{S})$ policy. On the other hand, the advantage of $Q(\mathbf{s}, \mathbf{S})$ over $(Q, \mathbf{S}, T)$ policy usually results from lower item-specific ordering and holding cost rates both achieved by imposing reorder levels. A similar observation is also true for $P(\mathbf{s}, \mathbf{S})$ policy.

It is interesting to note that the $(Q, \mathbf{S})$ policy performs so poorly with an average underperformance of 3.39% compared to the proposed policy. With the incorporation of the time trigger, *i.e.* increasing the dimensionality by one, we achieve significant improvements. An untabulated observation about the comparison of optimal policy parameters of $(Q, \mathbf{S})$ and $(Q, \mathbf{S}, T)$ policies is that $\mathbf{S}^*$ values of $(Q, \mathbf{S}, T)$ policy are generally smaller than that of $(Q, \mathbf{S})$ policy. Q^* of $(Q, \mathbf{S}, T)$ is also larger than that of $(Q, \mathbf{S})$. On the other hand, the time dimension T adjusts the frequency of reordering decisions. Therefore, $(Q, \mathbf{S}, T)$ policy uses smaller maximum inventory positions by using an effective proactive ordering mechanism. We observe that as K increases, $\Delta_{(Q,\mathbf{S})}\%$ decreases quite significantly. In view of the sensitivity results explained in Section 4.2, this is expected since the system uses the proactive ordering option less with increasing T. It is also interesting to observe that $\Delta_{(Q,\mathbf{S})}\%$ values are generally higher with lower service levels, *ie.* $\pi = 0, \rho = 30, h = 20$. This can be explained by higher values of $\mathcal{H}_T^*$ obtained for higher service levels which results in higher ordering cost rates.

Another interesting (untabulated) observation is that $(Q, \mathbf{S}, T)$ has, in all instances, resulted in a higher optimal system fill rate than the other four policies. In particular, $(Q, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$ policies have resulted in significantly lower optimal system fill-rates. Obviously, this may have important implications for inventory settings with non-linear shortage costs and service level constraints.

The performance of the proposed policy is somewhat mixed over the cost parameter set; a clear dominance region is not discernible. However, a general observation is that the proposed policy performs best for lower shortage, higher holding and lower common ordering costs. Although lower K values correspond to the cases where proactive ordering (*i.e.*, placing the orders at review epochs) becomes the dominant reordering mode, higher holding and lower shortage costs usually correspond to cases where the proactive ordering becomes less important. Therefore, the trade off between the savings in the backorder and holding costs and the increase in the ordering cost rates determines the advantage of the proposed policy. This will be more prominent in the Viswanathan experimental set below.

The second data set used in policy comparison is the one generated by Viswanathan [79]. For this set, the demand rates, lead times and item specific ordering costs are retained as in the standard 12-item problem set of Atkins-Iyogun; and different values are considered for the remaining costs as follows: $\pi = 0$, $K \in \{20, 50, 100, 200, 500\}$, $h \in \{2, 6, 10, 200, 600, 1000\}$, and $\rho \in \{10, 50, 100, 1000, 5000, 10000, 20000\}$. The considered instances and the results are tabulated in Tables 4.3 and 4.4. (We note that comparison with $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ has been made for the 36 instances reported in the study by Melchiors [53] to ensure fairness in comparing simulation-based results for the latter.)

The $(Q, \mathbf{S}, T)$ policy performs better than all other existing policies in 94 out of 120 experiment instances. For the remaining 26 experiment instances, it is dominated in 14 cases by $Q(\mathbf{s}, \mathbf{S})$, 7 times by $P(\mathbf{s}, \mathbf{S})$ and 5 times by $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$. As in the Atkins-Iyogun set, $(Q, \mathbf{S})$ is never the best policy.

K	ρ	h	$AC^*_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}\%$	K	ρ	h	$AC^*_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}\%$
	10	2	**772**	0.79	8.34	0.51	-		10	2	**810**	1.11	5.96	0.61	-
		6	1176	**-0.84**	7.08	1.97	-			6	**1221**	0.07	5.69	1.98	-
		10	1401	**-3.13**	4.70	3.72	-			10	1443	**-1.58**	4.21	3.67	-
	50	2	**905**	2.66	10.66	2.00	-		50	2	**954**	2.61	7.44	1.98	-
		6	1587	-0.57	7.79	**-1.02**	-			6	1669	-0.72	4.84	**-1.20**	-
20		10	**1918**	3.82	12.84	3.56	-	50		10	**2008**	3.92	10.17	3.38	-
	100	2	**965**	1.67	9.79	0.74	-		100	2	**1021**	1.26	6.13	0.70	-
		6	1727	-0.63	8.60	**-0.94**	-			6	1821	-0.49	6.27	**-1.04**	-
		10	**2169**	2.41	12.02	2.16	-			10	**2276**	2.72	8.99	2.20	-
	200	2	**1008**	1.99	10.35	0.70	-		200	2	**1068**	1.39	6.58	0.67	-
		6	1854	-0.54	9.21	**-0.87**	-			6	1955	-0.15	5.74	**-0.83**	-
		10	**2398**	1.55	11.49	1.20	-			10	**2522**	1.82	8.14	1.23	-
	10	2	**863**	0.94	3.45	0.47	-		10	2	**948**	0.74	1.24	0.20	-
		6	1301	**-0.07**	3.12	1.15	-			6	**1418**	0.72	1.51	0.92	-
		10	1523	**-0.91**	2.65	2.57	-			10	**1648**	0.49	1.49	1.51	-
	50	2	**1023**	1.75	4.00	1.18	-		50	2	**1118**	1.61	1.81	0.82	-
		6	1770	-0.73	2.34	**-1.30**	-			6	1933	-0.62	0.22	**-1.29**	-
100		10	**2129**	3.61	7.35	3.11	-	200		10	**2355**	2.37	3.20	1.79	-
	100	2	**1085**	1.37	3.61	0.66	-		100	2	**1181**	1.53	1.63	0.69	-
		6	1932	-0.51	2.57	**-1.20**	-			6	2116	-0.91	0.50	**-1.66**	-
		10	**2406**	2.79	6.59	2.12	-			10	**2635**	2.20	3.02	1.57	-
	200	2	**1137**	1.32	3.65	0.62	-		200	2	**1237**	1.46	1.56	0.49	-
		6	2079	-0.42	2.65	**-1.20**	-			6	2269	-0.62	0.32	**-1.46**	-
		10	**2665**	1.89	5.54	1.08	-			10	**2899**	1.86	2.47	0.94	-
	10	2	**1133**	0.45	0.04	0.00	-		100	2	1396	0.93	0.02	**-0.07**	-
		6	1696	0.25	0.04	**-0.06**	-			6	**2428**	0.82	0.02	0.00	-
500		10	**1966**	0.19	0.00	0.00	-	500		10	**3115**	0.71	0.00	0.00	-
	50	2	**1329**	0.74	0.05	0.00	-		200	2	**1457**	0.96	0.05	0.00	-
		6	**2238**	0.70	0.01	0.00	-			6	**2597**	1.00	0.09	0.00	-
		10	**2806**	0.62	0.00	0.00	-			10	**3389**	0.91	0.01	0.00	-

Table 4.3: Performance of $(Q, \mathbf{S}, T)$ Policy for the 12-item Viswanathan Problem Set

K	ρ	h	$AC^*_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}\%$	K	ρ	h	$AC^*_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(\mathbf{s},\mathbf{c},\mathbf{S})_M}\%$
	1000	200	**18175**	2.56	9.22	2.44	1.96		1000	200	**18627**	2.74	7.87	2.39	-
		600	**34210**	0.01	4.90	2.54	0.30			600	**34597**	0.74	4.46	2.36	-
		1000	44510	-2.36	1.55	1.11	**-2.85**			1000	44964	**-1.67**	1.18	0.77	-
	5000	200	**25501**	3.78	10.02	3.53	2.80		5000	200	**26076**	3.85	8.53	3.43	-
		600	**56828**	1.61	5.56	1.26	0.96			600	**57532**	1.62	4.99	1.34	-
20		1000	**78690**	3.07	5.95	2.79	2.42	50		1000	**79523**	3.11	5.39	2.88	-
	10000	200	**28575**	3.36	9.47	3.11	2.48		10000	200	**29156**	3.48	8.04	3.08	-
		600	**67081**	0.75	4.08	0.31	0.04			600	**67910**	0.70	3.21	0.33	-
		1000	**96323**	1.50	3.92	1.12	0.76			1000	**97329**	1.46	3.40	1.12	-
	20000	200	**31789**	2.01	7.72	1.76	1.09		20000	200	**32436**	2.01	6.38	1.60	-
		600	76249	0.89	3.90	0.29	**-0.04**			600	**77131**	0.86	3.17	0.33	-
		1000	**112118**	1.42	3.54	1.01	0.74			1000	**113019**	1.53	3.27	1.17	-
	1000	200	**19198**	2.47	6.28	2.03	1.69		1000	200	**19999**	2.59	4.83	1.96	-
		600	**35137**	1.32	4.15	2.16	0.67			600	**36059**	1.65	3.47	2.06	-
		1000	45423	-0.99	1.19	0.99	**-1.48**			1000	46270	**-0.11**	1.18	0.93	-
	5000	200	**26781**	3.75	7.08	3.09	2.43		5000	200	**27772**	3.56	5.54	2.89	-
		600	**58589**	1.50	3.63	1.11	0.87			600	**60123**	1.43	2.76	0.95	-
100		1000	**81123**	2.56	4.20	2.21	1.84	200		1000	**82893**	2.61	3.42	2.05	-
	10000	200	**29981**	3.10	6.49	2.54	1.91		10000	200	**31019**	3.20	4.96	2.46	-
		600	**68965**	0.70	2.51	0.25	0.00			600	**70541**	0.64	1.93	0.18	-
		1000	**98678**	1.38	2.91	1.04	0.68			1000	**100627**	1.45	2.05	0.94	-
	20000	200	**33220**	1.94	5.00	1.44	0.54		20000	200	**34323**	2.08	3.73	1.33	-
		600	78346	0.74	2.35	0.21	**-0.02**			600	**80045**	0.64	1.66	0.20	-
		1000	**114420**	1.55	2.88	1.12	0.81			1000	**116708**	1.32	2.13	0.91	-
	1000	200	**21917**	1.83	2.33	1.18	1.40		10000	200	**33390**	2.77	2.86	1.89	1.37
		600	**38526**	1.54	1.96	1.24	1.25			600	74001	0.87	0.86	0.08	**-0.06**
500		1000	**48581**	0.80	1.20	1.02	0.50	500		1000	**104610**	1.49	1.58	0.80	0.52
	5000	200	**30128**	2.80	3.11	1.92	1.54		20000	200	**36815**	1.76	1.96	0.85	0.19
		600	**63345**	1.48	1.57	0.77	0.68			600	**83590**	0.80	0.91	0.18	0.24
		1000	**87051**	1.96	2.28	1.31	1.16			1000	**121234**	1.24	1.43	0.67	0.30

Table 4.4: Performance of $(Q, \mathbf{S}, T)$ Policy for the 12-item Viswanathan Problem Set

Over all the 120 experiment instances, the average savings achieved through the implementation of the proposed policy in lieu of each of the existing policies are as follows: 1.25% over $P(\mathbf{s}, \mathbf{S})$, 4.16% over $(Q, \mathbf{S})$ and 1.07% over $Q(\mathbf{s}, \mathbf{S})$, and 0.82% over $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$.

As in the Atkins-Iyogun set, the dominance of the proposed policy is not monotone across the experiment instances. In the cases where the proposed policy gives the best solution, the improvement over the next best policy is 1.13%. Over these 94 instances, the maximum saving was observed to be 3.55%.

To give a broader view of the policy performances, comparison summaries are presented in two tables: Table 4.5 and Table 4.6. In both tables, we have included summaries of the unreported results on MP and $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$, as well.

Policy	Parameter Dimension	$(Q, \mathbf{S}, T)$	$Q(\mathbf{s}, \mathbf{S})$	$(\mathbf{s}, \mathbf{c}, \mathbf{S})_M^\dagger$	$P(\mathbf{s}, \mathbf{S})$	$(Q, \mathbf{S})$	MP	$(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$
$(Q, \mathbf{S}, T)$	14	-	1.43 (115)	1.85 (47)	1.63 (111)	4.05 (139)	4.94 (138)	10.59 (135)
$Q(\mathbf{s}, \mathbf{S})$	25	0.94 (24)	-	3.57 (17)	0.57 (117)	2.97 (139)	3.80 (139)	10.08 (122)
$(\mathbf{s}, \mathbf{c}, \mathbf{S})_M^\dagger$	36	0.85 (8)	1.01 (38)	-	0.83 (34)	3.39 (46)	5.67 (44)	7.25 (55)
$P(\mathbf{s}, \mathbf{S})$	25	0.85 (28)	2.48 (22)	2.66 (21)	-	3.25 (125)	3.70 (139)	10.78 (117)
$(Q, \mathbf{S})$	13	- (0)	- (0)	3.36 (9)	0.59 (14)	-	2.38 (91)	12.85 (83)
MP	24	0.20 (1)	- (0)	2.91 (11)	- (0)	2.03 (48)	-	12.49 (79)
$(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$	36	1.20 (4)	0.87 (17)	- (0)	0.42 (21)	3.65 (56)	4.12 (60)	-

Table 4.5: The summary comparison of policies over Atkins-Iyogun and Viswanathan sets across pairwise dominated instances. (†) $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ is compared over 55 total instances.

In Table 4.5, we provide a pairwise comparison in a matrix form across instances where one policy dominates the other. The first column lists the polices in the chronological order in which they have been proposed in the literature; the second column reports the number of control parameters that a particular policy employs for the standard test bed of 12 items. Each element of the matrix reports two entities: the average improvement in the expected total cost rate achieved by policy $\mathcal{P}_i$ over policy $\mathcal{P}_j$ in the experiment instances where $\mathcal{P}_i$ dominates $\mathcal{P}_j$; and, the number of such instances in parentheses. The first row of the table gives the performance of the proposed policy in comparison with the other policies. For example, we see that $(Q, \mathbf{S}, T)$ dominates $Q(\mathbf{s}, \mathbf{S})$ in 115 out of 139 considered instances; and, the average improvement in such instances achieved over $Q(\mathbf{s}, \mathbf{S})$

is 1.43%. Similarly, the proposed policy is better than $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ with an average improvement of 1.85% in 47 out of 55 considered instances, so on and so forth.

In Table 4.6, we provide an overall comparison of the *average* performance of the policies. Similar to the previous table, we list the policies in the chronological order that they have appeared in the literature, the dimension of each policy, and present the *average* percentage difference in the expected total cost rate under policy P_i versus P_j. Note that in creating this table, we consider all of the experiment instances, where P_i may or may not dominate P_j. Hence, we have negative averages for certain pairs. For instance, the $(Q, \mathbf{S})$ dominates the $P(\mathbf{s}, \mathbf{S})$ policy in 9 of 139 instances. On the other hand, the optimal cost rate of $(Q, \mathbf{S})$ policy is, on the average, 2.73% higher than that of $P(\mathbf{s}, \mathbf{S})$ policy. A positive entry indicates that policy P_i provides that much average percentage improvement in the cost rate over P_j. A negative entry indicates that the performance of P_i is inferior by that much, on average, in comparison with P_j. The first row gives the performance of the proposed policy $(Q, \mathbf{S}, T)$ with respect to the existing policies. Overall, we see that $(Q, \mathbf{S}, T)$ achieves an improvement of 1.03% over $Q(\mathbf{s}, \mathbf{S})$, 1.46% over $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$, 1.13% over $P(\mathbf{s}, \mathbf{S})$, 4.05% over $(Q, \mathbf{S})$, 4.90% over MP, and 10.25% over $(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$.

Policy	Parameter Dimension	$(Q, \mathbf{S}, T)$	$Q(\mathbf{s}, \mathbf{S})$	$(\mathbf{s}, \mathbf{c}, \mathbf{S})_M^\dagger$	$P(\mathbf{s}, \mathbf{S})$	$(Q, \mathbf{S})$	MP	$(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$
$(Q, \mathbf{S}, T)$	14	-	1.03	1.46	1.13	4.05	4.90	10.25
$Q(\mathbf{s}, \mathbf{S})$	25	-1.00	-	0.32	0.12	2.97	3.80	9.22
$(\mathbf{s}, \mathbf{c}, \mathbf{S})_M^\dagger$	36	-1.41	-0.25	-	-0.27	2.29	3.96	7.25
$P(\mathbf{s}, \mathbf{S})$	25	-1.10	-0.10	0.31	-	2.87	3.70	9.12
$(Q, \mathbf{S})$	13	-3.81	-2.83	-2.12	-2.73	-	0.86	6.20
MP	24	-4.59	-3.62	-3.94	-3.50	-0.77	-	5.32
$(\mathbf{s}, \mathbf{c}, \mathbf{S})_F$	36	-8.65	-7.74	-6.37	-7.66	-4.83	-4.07	-

Table 4.6: The overall average performance of policies over Atkins-Iyogun and Viswanathan sets across all instances. (†) $(\mathbf{s}, \mathbf{c}, \mathbf{S})_M$ is compared over 55 total instances.

When viewing these statistics, we should bear in mind a couple of issues. First, the comparisons are made between policies that have already been demonstrated to perform well. The chronological listing enables one to see the evolution of the performances of the policies studied over time, as well. Second, in multi-item settings, the total system cost are substantial in nominal terms; hence, expressing improvements in percentages inevitably understates their impact. Especially, in operating environments where margins are known to be notoriously low, as in retail industry, an improvement of even a couple of percentage points does have a substantial impact on profitability. (*e.g.* Fisher *et al* [33]. In particular, take the case of a major home-improvement retailer with a pretax profit margin of 5.8% and a return on asset (ROA) of 2.7%. If this company could cut its inventory related costs by just 3%, its pretax profits would increase 37%, and the pretax profit margin would rise to nearly 8%. Therefore, the improvements that the proposed $(Q, \mathbf{S}, T)$ policy achieves over the existing ones are comparably significant. Moreover, the proposed policy attains such performance levels with parsimony - compare $N + 2$ policy parameters of $(Q, \mathbf{S}, T)$ versus $2N + 1$ of $Q(\mathbf{s}, \mathbf{S})$ and $P(\mathbf{s}, \mathbf{S})$ or $3N$ of the can-order policies. The simpler $(Q, \mathbf{S})$ with 13 policy parameters is no match with an average underperformance of 3.81%. This low dimensionality reduces the computational effort in optimization enormously and eases implementation in practice greatly. Viewing the comparisons in this broader perspective, we can conclude that the proposed policy provides significant improvements over the existing policies in terms of cost savings, optimization effort and ease of implementation and that this performance is robust over a broad range of environmental parameters.

Although used as a benchmark testbed, the Atkins-Iyogun and Viswanathan sets exclude an important category of settings in which joint replenishment is commonly practiced - settings where the items have similar ordering cost structures and/or demand rates. Therefore, the impact of the overall system demand rate and of the diversity of demand rates among items is an aspect of stochastic joint replenishment which has not been studied in the literature before. In the next section, we focus on such demand rate affects.

4.3.2 Impact of Demand Rates

To examine the effects of system or item demand rates, we constructed our own test bed with insights from the Atkins-Iyogun set. Since we have identified $Q(\mathbf{s}, \mathbf{S})$ and $P(\mathbf{s}, \mathbf{S})$ policies as the only viable alternatives to our proposed policy in the above comparisons, in this part of our numerical study, we compared $(Q, \mathbf{S}, T)$ with only those two and $(Q, \mathbf{S})$ as a special case to illustrate the advantage of proactive ordering. We begin with the effect of system demand rate on the performance of the control policies.

k	λ	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$
20	20	**1059.86**	4.81	3.65	3.60
	40	**1502.78**	3.79	1.18	1.17
	60	**1858.18**	2.13	1.15	1.15
	80	**2156.14**	1.17	0.06	0.06
40	20	**1250.47**	3.17	1.25	1.24
	40	**1775.37**	2.27	1.05	1.05
	60	**2178.50**	1.64	0.98	0.99
	80	**2523.70**	0.99	0.72	0.72
60	20	**1409.07**	2.17	1.05	1.05
	40	**2012.24**	0.91	0.91	0.92
	60	**2468.62**	0.41	0.18	0.18
	80	**2847.50**	0.24	0.14	0.13

Table 4.7: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Demand Rates and Item-specific Ordering Cost, $N = 8, K = 150, L = 0.2, h = 6, \pi = 30, \rho = 0$

We consider $N = 8$ identical items with $K = 150$, $h_i = h = 6$, $\pi_i = \pi = 30$, $\rho_i = \rho = 0$ and $L_i = L = 0.2$ and $k_i = k = \{0, 20, 40, 60\}$ for all i. With identical item demand rates, we consider the system demand rates as $\lambda_0 = \{160, 320, 480, 640\}$. We present our results for $k = \{20, 40, 60\}$ in Table 4.7 and for $k = 0$ in Figure 4.2.

In all instances, the proposed policy dominates the existing policies. The average savings achieved through the implementation of the proposed policy in lieu of each of the existing policies are as follows: 2.19% over $P(\mathbf{s}, \mathbf{S})$, 1.43% over $(Q, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$. There is not any discernible difference between the performances of $(Q, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$ as also reported in [56]. We observe that the performances of the policies become alike as system demand rate increases. As the demand rate of each item increases, the advantage of each policy somewhat offsets the disadvantage of them. For instance, $(Q, \mathbf{S}, T)$ policy gives lower holding and backorder cost rates than the other policies and as demand rate increases the difference in holding and backorder cost rates across policies increases. On the other hand, as the demand rate increases the proactive ordering behaviour becomes more dominant (See Table 4.1) and hence the ordering cost rates increase and eliminates the advantage coming from the backorder and holding costs. For instance, for $k = 40$ and $\lambda = 20$, holding, backorder and ordering cost rates for $(Q, \mathbf{S}, T)$ policy are 674.11, 120.92, 455.44, respectively. The corresponding figures for $P(\mathbf{s}, \mathbf{S})$ and $(Q, \mathbf{S})$ (or $Q(\mathbf{s}, \mathbf{S})$) policies 681.37,133.50,475.13 and 677.39, 125.90,469.98, respectively. For $\lambda = 80$, the holding, backorder and ordering cost rates are 1362.19, 162.15, 999.36 for $(Q, \mathbf{S}, T)$ policy; 1419.55, 178.14, 950.31 for $P(\mathbf{s}, \mathbf{S})$ policy; 1378.13, 170.16, 970.33 for $(Q, \mathbf{S})$ (or $Q(\mathbf{s}, \mathbf{S})$) policy.

It is also observed that the advantage of the proposed policy decreases with increasing k, possibly resulting from the larger increase in the item specific ordering cost rate.

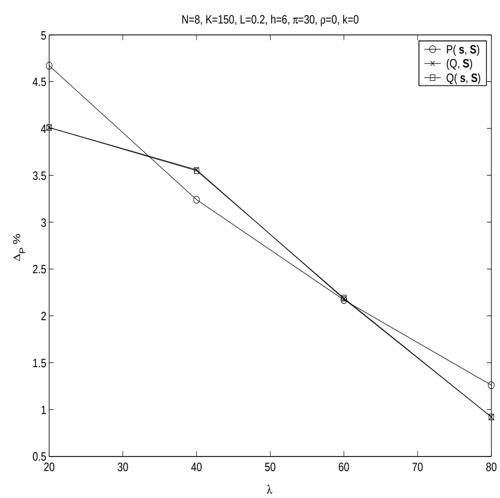

Figure 4.2: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Demand Rates and $N = 8, K = 150, L = 0.2, h = 6, \pi = 30, \rho = 0$

Next, we examine the effect of item demand rates while keeping the system demand rate constant. This is equivalent to examining the effect of number of items that are jointly replenished for a given system demand rate. Hence, we consider the set of N identical items with $\lambda_0 = 320$, $K = 150$, $k_i = k = 20$, $h_i = h = 6$, $\pi_i = \pi = 30$, $\rho_i = \rho = 0$ for all $i = 1, \cdots, N$. We vary the number of items and lead times as $N = 2, 4, 6, 8, 10, 12$ and $L_i = L = 0.2, 0.4, 0.6$. Note that individual demand rates are also equal to each other in this set. The results are presented in Table 4.8 and Figure 4.3.

L	N	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$
	2	**1130.21**	5.87	4.60	4.60
	4	**1319.60**	4.85	3.26	3.26
0.4	6	**1492.23**	4.65	2.94	2.93
	8	**1550.21**	4.20	2.83	2.82
	10	**1760.81**	1.14	0.38	0.38
	12	**1918.50**	0.29	0.09	0.08
	2	**1193.20**	5.59	4.67	4.67
	4	**1271.01**	5.03	3.39	3.38
0.6	6	**1524.10**	4.30	3.01	3.00
	8	**1608.83**	3.47	2.40	2.40
	10	**1799.12**	1.44	0.83	0.82
	12	**2009.97**	0.41	0.30	0.30

Table 4.8: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Lead-time and Number of Items, $\lambda_0 = 320, K = 150, h = 6, \pi = 30$

In all cases, the proposed policy dominates the other policies. The average savings achieved through the implementation of the proposed policy in lieu of each of the existing policies are as follows: 3.56% over $P(\mathbf{s}, \mathbf{S})$, 2.29% over $(Q, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$. As also presented in Table 4.7, a peculiarity of the $Q(\mathbf{s}, \mathbf{S})$ policy strikes out immediately: Incorporation of individual trigger levels $\mathbf{s}$ does not improve the much simpler $(Q, \mathbf{S})$ policy noticeably in the case of identical items. In comparison with the $(Q, \mathbf{S})$ policy, we observe, however, that introduction of a time trigger in the $(Q, \mathbf{S}, T)$ policy provides significant savings. The savings under the proposed policy are much pronounced for small number of items. As, N grows large, the difference between the performances of the policies starts diminishing; however, $P(\mathbf{s}, \mathbf{S})$ is much slower in this respect. We observe that the effect of lead time is not monotone.

The last issue we investigate is the impact of demand rate diversity among the items on the policy performances. In Table 4.9, we report a representative case of $N = 4$ items with identical cots parameters of $K = 150$, $\pi = 30$, $\rho = 0$, $k = 20$,

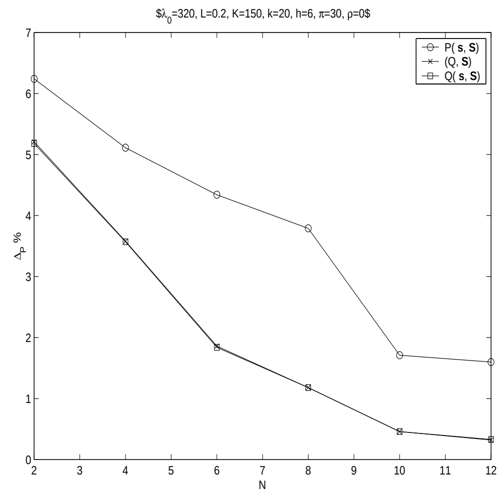

Figure 4.3: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Number of Items and $\lambda_0 = 320, K = 150, L = 0.2, k = 20, h = 6, \pi = 30, \rho = 0$

$h = 6$ and identical lead times of $L = 0.2$ and a system demand rate of $\lambda_0 = 320$. As tabulated, we consider various groupings of demand rates among the items. In the first block (instance 1), all items have equal demand rates and it may be viewed as a reference instance. The rest of the instances attempt to generate groupings of differing demand rate diversity among the items, producing 'lop-sided' spread of demands. In the second and third blocks (instances 2 through 5 and 10 through 13), three items are identical, one is different. In the third block (instances 6 through 9), items are grouped into two identical pairs. In the last block (instances 14 through 18), all four items have different demand rates.

The proposed policy dominates the existing policies in this set, as well. The average savings achieved through the implementation of the proposed policy in lieu of each of the existing policies are as follows: 3.76% over $P(\mathbf{s}, \mathbf{S})$, 3.88% over $(Q, \mathbf{S})$ and 2.66% over $Q(\mathbf{s}, \mathbf{S})$. The average savings over the next best policy is 2.60%.

For the improvement achieved through the proposed policy we make the following observations. For the first 13 instances, we see that, as diversity among item demand rates increases, the savings of the proposed policy also increases due to the demand diversification structure. For the last block, the opposite is true when we compare the proposed policy against $P(\mathbf{s}, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$. Furthermore, for the last block, the improvement is significantly smaller than those for other instances. We also note that as item demand rates become more dissimilar, performances of $P(\mathbf{s}, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$ approach that of the proposed policy. Considering the next best policy, we observe an interesting change of dominance among the policies. In the first 13 instances, performances of $Q(\mathbf{s}, \mathbf{S})$ and $(Q, \mathbf{S})$ are almost identical and they both dominate $P(\mathbf{s}, \mathbf{S})$. However, in the last block (instances 14 through 18), when items have identical unit cost structures but are dissimilar greatly in their individual demand rates, we observe a shift so that $Q(\mathbf{s}, \mathbf{S})$ performs significantly better than $(Q, \mathbf{S})$. Furthermore, in the same region, $P(\mathbf{s}, \mathbf{S})$ starts to dominate $Q(\mathbf{s}, \mathbf{S})$, albeit by a small margin. Such changes in dominance is not only of interest for theory but of importance for practice of supply chain design and management. It would be interesting to investigate the joint location-allocation-replenishment problem in a supply chain.

λ_1	λ_2	λ_3	λ_4	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{P(\mathbf{s},\mathbf{S})}\%$	$\Delta_{(Q,\mathbf{S})}\%$	$\Delta_{Q(\mathbf{s},\mathbf{S})}\%$
80	80	80	80	**1191.08**	5.11	3.58	3.58
70	70	70	110	**1080.65**	4.75	2.30	2.30
60	60	60	140	**1049.18**	5.12	2.45	2.45
50	50	50	170	**1018.61**	6.24	2.78	2.78
40	40	40	200	**998.95**	6.40	3.12	3.12
70	70	90	90	**1131.88**	4.12	3.79	3.78
60	60	100	100	**1109.69**	4.56	3.98	3.98
50	50	110	110	**1087.92**	5.09	4.21	4.21
40	40	120	120	**1066.60**	5.40	4.34	4.34
70	83.33	83.33	83.33	**1166.47**	3.27	2.66	2.66
60	86.67	86.67	86.67	**1154.93**	4.02	2.77	2.77
50	90	90	90	**1143.49**	4.30	2.99	2.99
40	93.33	93.33	93.33	**1132.17**	4.55	3.21	3.21
70	60	100	90	**1170.55**	1.58	5.27	1.78
70	50	110	90	**1173.22**	1.31	5.36	1.56
70	40	120	90	**1177.38**	0.92	5.49	1.01
70	30	130	90	**1179.07**	0.55	5.61	0.77
70	20	140	90	**1179.53**	0.34	5.93	0.56

Table 4.9: Performance of $(Q, \mathbf{S}, T)$ Policy for Non-Identical Items-Additional Set, $K = 150, k = 20, h = 6, \pi = 30, \rho = 0$

In summary, we conclude that the performance of the proposed policy is influenced by the structure of the demand rates within the system, as expected.

4.3.3 Impact of Fill Rate Constraints

In this part of the numerical study, we solve the optimization problem under fill rate constraints for the items instead of using time-weighted and/or unit backorder costs. Therefore, the optimization problem can be stated as:

$$\min_{Q,\mathbf{S},T} AC(Q,\mathbf{S},T) = \frac{K + \sum_{i=1}^{N} k_i \theta_i}{E[Y]} + \sum_{i=1}^{N} h_i E[OH_i]$$
$$\text{subject to}$$
$$\psi_i \geq \overline{\psi}_i \qquad i = 1, 2, \ldots, N$$

where $\overline{\psi}_i$ is the target fill rate for item i.

The $(Q, \mathbf{S}, T)$ policy under fill rate constraints is also solved using the search algorithm in Section 4.1 except that in step 2.1.1.3 we have $S_i^* = \min\{S_i : S_i \geq 1, \psi_i \geq \overline{\psi}_i\}$ and we do not construct the limits, S_i^{min} and S_i^{max}. For $(Q, \mathbf{S})$ policy, the optimal policy parameters are solved by using the search algorithm in Section 4.1 with $T \to \infty$ and the modification for step 2.1.1.3. For $P(\mathbf{s}, \mathbf{S})$ and $Q(\mathbf{s}, \mathbf{S})$ policies, we enumerate a large number of policy parameters and cost rates and the optimal policy parameters are found by selecting the minimum cost rate function satisfying the target fill rate constraint.

We present the performance of the proposed policy for different values of $\overline{\psi}_i$ in Table 4.10. We only included the $P(\mathbf{s}, \mathbf{S})$, $(Q, \mathbf{S})$, and $Q(\mathbf{s}, \mathbf{S})$ for numerical comparison under fill rate constraints since these policies are previously shown to dominate the other policies.

P	$(Q, \mathbf{S}, T)$			$P(\mathbf{s}, \mathbf{S})$			$(Q, \mathbf{S})$			$Q(\mathbf{s}, \mathbf{S})$		
$\overline{\psi}$	$AC_{(Q,\mathbf{S},T)}$	OC	HC	$\Delta_{P(\mathbf{s},\mathbf{S})}$	OC	HC	$\Delta_{(Q,\mathbf{S})}$	OC	HC	$\Delta_{Q(\mathbf{s},\mathbf{S})}$	OC	HC
0.99	**1050.94**	567.11	550.11	8.16	451.52	685.18	6.03	553.27	561.08	6.02	553.13	561.08
0.95	**1002.18**	526.81	475.37	7.23	432.21	642.23	4.18	518.27	525.62	4.19	518.27	525.62
0.90	**975.26**	510.23	465.03	6.15	403.19	632.05	3.25	499.93	507.02	3.24	499.84	507.02
0.85	**946.53**	492.19	454.34	5.14	394.98	600.20	2.56	481.96	488.80	2.55	481.87	488.80
0.80	**910.11**	469.18	440.93	4.13	378.17	569.53	2.18	461.70	468.25	2.19	461.70	468.25

Table 4.10: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Fill Rates, $\lambda_0 = 320, L = 0.2, N = 2, K = 150, h = 6$

In our results, we assumed $\lambda_0 = 320, N = 2, L = 0.2, K = 150, k = 20, h = 6$, We have numerically observed that the performance of the proposed policy is more significant with higher fill rate constraints. When compared with $(Q, \mathbf{S})$ or

$Q(\mathbf{s}, \mathbf{S})$ policies, this may be explained by the higher stock that the other policies should have in order to achieve the target fill rate constraints and hence the difference of holding cost rate between different policies increase. As the fill rate constraints get more relaxed, the difference in the holding cost rates is smaller so that the $\Delta_{\mathcal{P}}\%$ values decrease. Moreover, the $(Q, \mathbf{S}, T)$ policy achieves the target fill rate consraints in a more tight way, *i.e.* for $\overline{\psi} = 0.99$, $\psi = 0.9903$ whereas $\psi = 0.9912$ for both $(Q, \mathbf{S})$ or $Q(\mathbf{s}, \mathbf{S})$ policies. Similarly, for $\overline{\psi} = 0.80$, we have $\psi = 0.8005$ for $(Q, \mathbf{S}, T)$ policy and the corresponding value is 0.8019 for $(Q, \mathbf{S})$ or $Q(\mathbf{s}, \mathbf{S})$ policies. This can be explained by the continuous time dimension of the proposed policy. When compared with $P(\mathbf{s}, \mathbf{S})$ policy, the advantage of $(Q, \mathbf{S}, T)$ policy also comes from the lower holding costs. However, there is not a monotone behaviour for the fill rates achieved by $P(\mathbf{s}, \mathbf{S})$ policy when compared with $(Q, \mathbf{S}, T)$. ψ values obtained are 0.9906 and 0.8002 for $\overline{\psi} = 0.99$ and 0.80, respectively.

4.4 Batch Demand

Finally, we study the impact of batch demand arrivals. In the previous chapter, we presented an exact methodology to analyze the proposed $(Q, \mathbf{S}, T)$ policy under compound Poisson demand. However, it is very difficult to carry out a numerical analysis with compound Poisson demand if the batch size distribution of the items is not closed under convolution. Because, in that case it becomes almost impossible to calculate $w_{\Theta}(q, k)$ especially for higher values of q, k and/or N from computational point of view.

Therefore, to illustrate the performance of $(Q, \mathbf{S}, T)$ policy, we consider the case where all items are identical in demand rate and cost parameters. The demand is assumed to follow a compound Poisson process with an overall rate λ_0/N and a geometrically distributed demand size with parameter p for all items. Therefore, the k'th convolution of the batch size follows a negative binomial distribution with parameters k and p. Since the rest of the policies have not been generalized for compound demand processes, we can only report the comparison between $(Q, \mathbf{S}, T)$ and $(Q, \mathbf{S})$ which is its special case. To make a fair comparison across different demand size parameters, we fix the average number of units demanded per time, $\lambda_0/p = \alpha$. We set $\alpha = 320$, $L = 0.2$, $K = 150$, $k = 20$, $h = 6$, $\pi = 30$, $\rho = 0$. We vary $p \in \{0.2, 0.5\}$ and $N \in \{2, 4, 6, 8, 10, 12\}$. We use σ_d^2 to denote the variance of demand size.

Batch Dist.	Unit ($\sigma_d = 0$)		Geo($p = 0.5$),($\sigma_d = 2$)		Geo($p = 0.2$),($\sigma_d = 20$)	
N	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{(Q,\mathbf{S})}\%$	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{(Q,\mathbf{S})}\%$	$AC_{(Q,\mathbf{S},T)}$	$\Delta_{(Q,\mathbf{S})}\%$
2	**1058.91**	5.21	**1082.34**	6.14	**1105.18**	6.54
4	**1191.08**	3.58	**1225.67**	4.13	**1276.10**	4.72
6	**1358.01**	1.86	**1398.98**	2.15	**1453.01**	3.19
8	**1502.78**	1.18	**1576.14**	1.29	**1604.26**	1.76
10	**1681.29**	0.46	**1723.09**	1.02	**1775.20**	1.24
12	**1861.19**	0.32	**1903.21**	0.68	**1952.15**	0.95

Table 4.11: Performance of $(Q, \mathbf{S}, T)$ Policy for Identical Items with Different Number of Items and Compound Poisson Demand, $\alpha = 320, L = 0.2K = 150, k = 20, h = 6, \pi = 30, \rho = 0$

We present the performance of $(Q, \mathbf{S}, T)$ policy in the presence of batch demands in Table 4.11. We observe the intuitive finding that as the variance of the demand size increases, the savings due to the introduction of a time trigger also increase. As in the unit demand case, the savings decrease with the number of items, N, but at a slightly slower rate.

Chapter 5

SJRP in a Two-Echelon Inventory System

In the previous two chapters, we have presented a new policy for the Stochastic Joint Replenishment Problem in a single-location, multi-item setting and compared the performance of the proposed policy with those of the existing ones in the literature. The single-location, multi-item model also corresponds to a two-echelon supply chain where the upper echelon employs cross docking. In this chapter, we extend our model to incorporate a single-item, multi-location setting where the upper echelon also holds inventory.

In Section 5.1, we present the assumptions of our model and introduce the ordering policies for both echelons. The modeling methodology presented here is based on the development of the ordering process by the lower echelon and provides an analytical tool to investigate various joint replenishment policies under a more general policy class. In Sections 5.2, we explain the proposed framework for the analysis of the warehouse and the retailers. Note that the proposed methodology is not specific to a particular policy but is applicable to any policy that satisfies the characteristics of the considered class.

5.1 Model Assumptions

We consider a single-item, divergent two-echelon inventory system with a single warehouse and N retailers. The retailers face stationary and independent unit Poisson demands with rates λ_i $i = 1, 2, \ldots, N$ and all unmet demands are backordered at every installation. Provided that the ordered amount is available, the lead time from the warehouse to retailer i, L_i is constant. The warehouse gives orders to an external supplier with ample stock and lead time for deliveries is a constant, L_0 (See Figure 5.1).

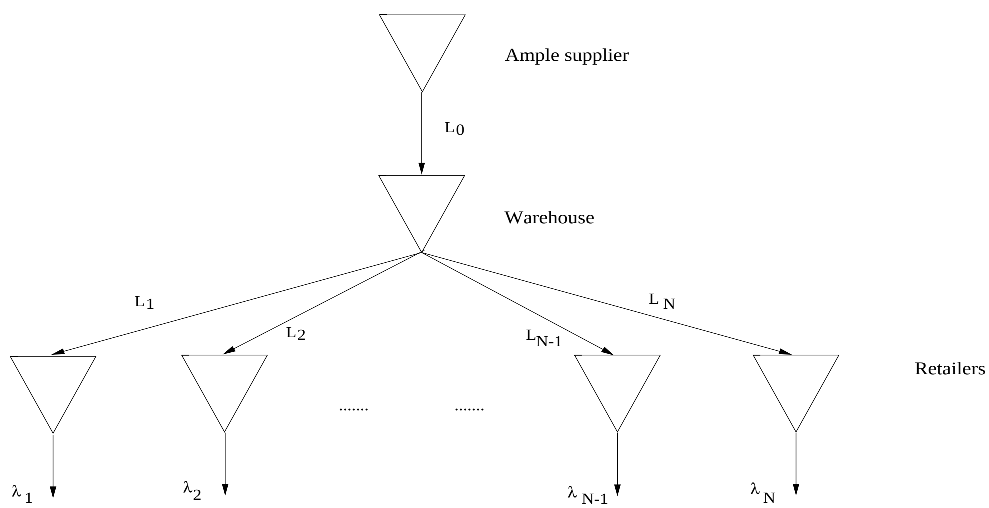

Figure 5.1: Illustration of a Divergent Two-Echelon Inventory System

The system is continuously reviewed and all the information regarding the last replenishment epochs at each installation; the time elapsed since then; the total demand that has arrived at the system subsequent to an order placed is available. The records related with the timing of the orders also enable us to review the system periodically.

The ordering costs associated with the inventory system are the warehouse fixed ordering cost K_0 and a common fixed cost K associated with a retailer order. To enhance the impact of joint replenishment and fully utilize the economies of scale in replenishments, we assume that the retailer specific ordering costs are negligible.

Holding costs per unit per time are charged at every installation with rates h_0 and h_i for the warehouse and retailer i, respectively. Moreover, h_0 includes the holding cost of the items from the time the order is released from the outside supplier until it reaches the warehouse. Also note that in the sequel, we assume that the inventory is charged at the retailer's expense from the instance it leaves the warehouse. Backorder costs are difficult to measure and determine and hence backorder costs at the retailers are handled implicitly by considering fill rate constraints, ie. the proportion of demand delivered immediately from the retailers' stock. Furthermore, we assume that the cost of monitoring the inventory system continuously is negligible.

While satisfying the orders at the warehouse, the following is assumed:

1. The integrality of the orders placed by the retailers at the warehouse is preserved. This means that, if an order arrives and the existing on-hand inventory is not sufficient to satisfy the order, then the existing inventory is kept at the warehouse while the entire order waits until an inventory of an adequate size accumulates at the warehouse as outstanding orders. This is also referred to as the no-lot-splitting assumption.

2. The orders of the retailers are satisfied in the sequence they were placed, ie. order crossing is not allowed. That is, even if there is enough stock on-hand at the warehouse to satisfy an order, that order will be backordered if there is a previously placed order for which the existing on-hand inventory is reserved.

Under the assumed cost structure described above, our objective is to minimize the expected total cost per unit time subject to fill rate constraints at the retailers. With this objective, we consider a system in which the warehouse employs the following continuous review (s, S) ordering policy:

Policy of the warehouse (s_0, S_0)**:** *When the inventory position of the warehouse crosses* s_0*, a replenishment order is placed at the outside supplier to raise it up to* S_0*.*

The ordering cost structure of the retailers presents an opportunity to exploit the economies of scale in the replenishment. Therefore, instead of installation stock policies, we propose that all retailers are included in every possible replenishment opportunity, *e.g. a retailer triggers an order because its inventory position drops to its reorder level or the number of demands accumulated in the system reaches a truckload size*, to take full advantage of savings in the ordering costs. We consider four different policies which are briefly described as below:

1. $(Q, \mathbf{S}, T)$ **Policy:** *Monitor all inventory positions continuously, and raise the inventory positions of the items up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$
   i) whenever a total of Q *demands accumulate for the items or*
   ii) at time kT *if at least one demand occurs in* $((k-1)T, kT]$ *with no demand arrivals in* $(0, (k-1)T]$,
   whichever occurs first. This policy has been recently proposed by Özkaya *et al.* [57] in a single location, multi-item setting and in Chapter 3 of this thesis.

2. $(Q, \mathbf{S})$ **Policy:** *Raise the inventory positions of the retailers up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$ *whenever a total of* Q *demands accumulate for the retailers.* This policy was previously studied by Cheung and Lee [23] for a two-echelon inventory system.

3. $(Q, \mathbf{S}|T)$ **Policy:** *Monitor all inventory positions every* T *time units, and raise the inventory positions of the retailers up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$ *if a total of at least* Q *demands have accumulated in the system.*
   This policy was previously studied by Cachon [17] in a single-location and multi-item inventory system considering shelf space and truck capacities.

4. $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ **Policy:** *Whenever the inventory position of a retailer* i *drops to its reorder level,* s_i*, raise the inventory positions of the retailers up to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$.
   The policy is a special case of $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ can-order policy where $\mathbf{c} = \mathbf{S} - \mathbf{1}$

which were previously studied by Silver [67] with two items and zero lead time and Van Eijs [77] in a more general multi-item setting.

For each of the above four policies, the instances at which the orders are placed at the warehouse are regenerative points for the retailers. This follows since the unit demand process is Poisson and at each order trigger instance, the inventory position of the retailers are at their order-up-to levels. These policies only differ in how/when the orders are placed, *ie.* how the ordering instances are generated.

The regenerative structure of these policies at the ordering instances enable us to develop a framework for the analysis of them under a more general policy class $\mathcal{P}$ described as below:

Policy of the Retailers ($\mathcal{P}$): *At each ordering instance, raise the inventory positions of the retailers to* $\mathbf{S} = (S_1, S_2, \ldots, S_N)$.

The structure of policy class $\mathcal{P}$ makes use of the idea of joint replenishment. Under the joint replenishment policy class $\mathcal{P}$, when a retailer is taken in isolation, it experiences exogenously generated opportunities of replenishment with no additional fixed ordering costs. In the presence of opportunities of replenishment at no additional ordering cost, it is intuitive that a retailer may choose to reorder at these opportunity arrivals since this would reduce the ordering costs in the system. Therefore, each retailer i whose inventory position is below S_i at an opportunity arrival instance chooses to use the replenishment opportunity and raises its inventory position to order-up-to level, S_i. If the retailer specific ordering costs were positive the order-up-to structure of the considered policy class would not be cost-effective. We also note that although the policy class $\mathcal{P}$ is quite general, the MP policy of Atkins and Iyogun [4] does not belong to class $\mathcal{P}$ since the retailers are not regenerative due to retailer specific review intervals.

Figure 5.2 illustrates the ordering process of the retailers under the $(Q, \mathbf{S}, T)$ policy within the class $\mathcal{P}$. Let $IP_i(t)$ denote the inventory position of retailer i at time t. We present the time sequence of events and the decisions taken in terms of the retailer ordering process. We have $N = 2, S_1 = 9, S_2 = 6$ $Q = 8$ and some

$T > 0$ as the policy parameters and initially both retailers are at their maximum stocking levels. Until time $t_1 = T$, 3 and 2 demands arrive for retailers 1 and 2, respectively. Since T time units have passed before Q demands have accumulated at the system, an order is placed at $t = t_1$ which brings the inventory position of retailer 1 to S_1 and of retailer 2 to S_2. At time $t_2 = 2T$, a total of T time units have elapsed since the last order was placed; therefore, an order is placed as triggered by the policy. The order size is one and only retailer 2 is included in this order since no demand has arrived for retailer 1 between $t = t_1$ and $t = t_2$. At time $t = t_3(< t_2 + T)$, Q^{th}demand after the last order was placed arrives and hence an order is triggered. This order is composed of 5 and 3 units for retailers 1 and 2, respectively. The process goes on further in the same manner.

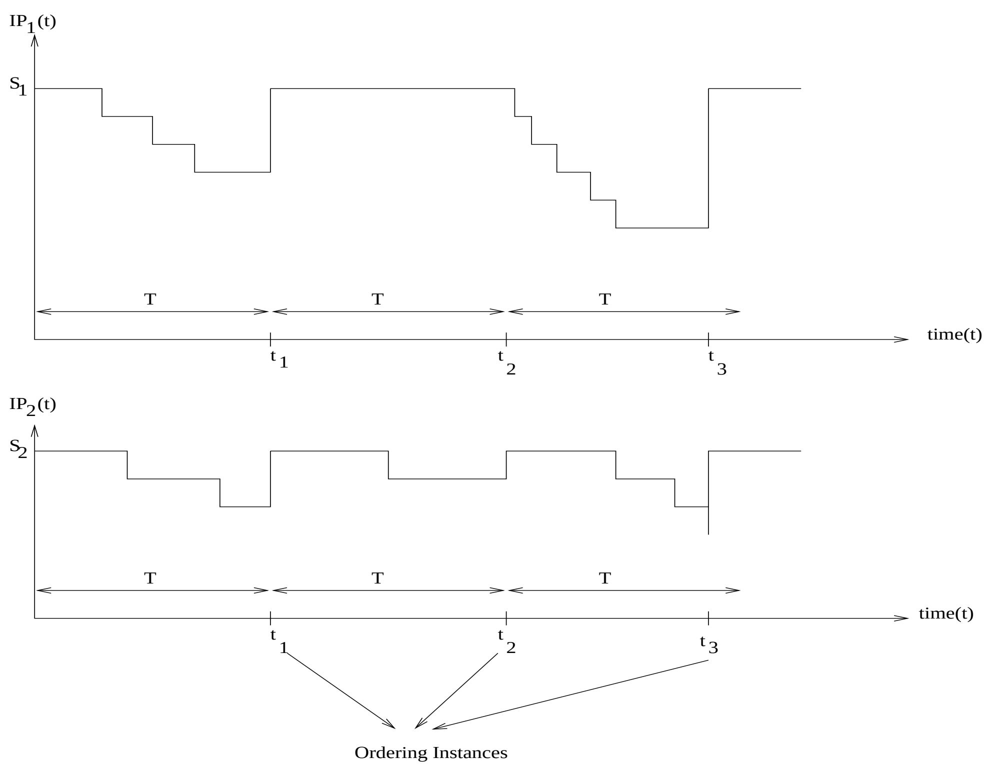

Figure 5.2: Illustration of Ordering Process of Retailers under Policy Class $\mathcal{P}$

We note that, under all the suggested policies within the class $\mathcal{P}$, the order size, the inter-order time, the number of retailers included in each order and the order quantity of each retailer in an order are all random. These will be explained in more detail in Chapter 6. Since ordering instances are regenerative points at the retailers' level, the mentioned characteristics of consecutive orders are independent of each other.

Now, consider an order of size $Q_0 = q$ placed at the warehouse and let $R_i(q)$ be the random variable representing the order quantity of retailer i in this order with probability mass function, $P_{R_i(q)}(m_i)$. $R_i(q) = m_i$ implies that the inventory position of retailer i just before the order of interest is placed is $S_i - m_i$ and we clearly have $\sum_{i=1}^{N} R_i(q) = q$.

Another important feature of our model is that the size of an order is not independent of the corresponding inter-order time. Therefore, the cumulative demand at the warehouse constitutes a compound renewal process where the inter-order time, Y and order quantity, Q_0 have a joint density, $f_{Y,Q_0}(y, q)$. Also, let $P_{Q_0}()$ and $f_Y()$ be the probability mass function of Q_0 and probability density function of Y with corresponding cumulative distribution functions $F_{Q_0}()$ and $F_Y()$.

Now, suppose that at time $t = 0$, the inventory positions of the retailers are at $S = (S_1, S_2, \ldots, S_N)$ and let $(Y_1, Q_1) = (y_1, q_1), (Y_2, Q_2) = (y_2, q_2), \ldots$ be the inter-order time and order quantity of the consecutive orders after time $t = 0$. Let $X_1, X_2, ..., X_n$ be independent and identically distributed replicants of a random variable, X. The n^{th} convolution of X is denoted by $X^{(n)} = \sum_{i=1}^{n} X_i$ with the convention that $X^{(0)} = 0$. Since, the convolutions of Y and Q_0 will be used quite frequently in the sequel, we also let $F_{Y^{(n)},Q_0^{(n)}}(y, q) = P(Y^{(n)} \leq y, Q_0^{(n)} = q)$. Notice that, although F usually represents distribution functions, by an abuse of notation, we will let F denote the sub-distribution function of $Y^{(n)}$ and $Q_0^{(n)}$. Then, we have

$$F_{Y^{(n)},Q_0^{(n)}}(y, q) = \int_{t=0}^{y} f_{Y^{(n)},Q_0^{(n)}}(t, q)dt \quad \text{if} \quad n > 0, y > 0, q > 0 \qquad (5.1)$$

Here, $f_{Y^{(n)},Q_0^{(n)}}(y, q)$ is the joint density of $Y^{(n)}$ and $Q_0^{(n)}$ and given by

$$f_{Y^{(n)},Q_0^{(n)}}(y, q) = \sum_{\mathcal{A}(n,y,q)} \prod_{i=1}^{n} f_{Y,Q_0}(y_i, q_i) \quad \text{if} \quad n > 0, y > 0, q > 0$$

where $\mathcal{A}(n,y,q) = \{(y_1,q_1),(y_2,q_2),\ldots,(y_n,q_n) : \sum_{i=1}^n y_i = y, \sum_{i=1}^n q_i = q\}$. If $n = 0, y \geq 0, q = 0$, we have $F_{Y^{(n)},Q_0^{(n)}}(y,q) = 1$.

Let $D_0(t_1,t_2]$, $D_0[t_1,t_2)$ and $D_0(t_1,t_2)$ be the total number of units demanded from the warehouse in half-closed intervals $(t_1,t_2]$, $[t_1,t_2)$ and (t_1,t_2), respectively. Next, we obtain the probability mass function of $D_0(0,t]$.

Lemma 5.1.1 *Let $\varphi(t,k)$ denote the probability that $D_0(0,t]$ is k. Then,*

$$\varphi(t,k) = \begin{cases} \overline{F}_Y(t) & \text{if } k = 0 \\ \sum_{n=1}^k F_{Y^{(n)},Q_0^{(n)}}(t,k) - \int_{y=0}^t F_{Y^{(n)},Q_0^{(n)}}(t-y,k)dF_Y(y) & \text{if } k > 0 \end{cases}$$

Proof: *Observe that the event $\{D_0(0,t] = k\} \equiv \{Y_1 > t\}$ if $k = 0$ and $\{D_0(0,t] = k\} \equiv \{\sum_{i=1}^n Y_i \leq t < \sum_{i=1}^{n+1} Y_i, \sum_{i=1}^n Q_i = k\}$ for $k > 0$.*
Hence, for $k = 0$, $\varphi(0,t) = \overline{F}_Y(t)$ and for $k > 0$

$$\begin{aligned} \varphi(t,k) &= P(Y^{(n)} \leq t, Q_0^{(n)} = k) - P(Y^{(n+1)} \leq t, Q_0^{(n)} = k) \\ &= F_{Y^{(n)},Q_0^{(n)}}(t,k) - \int_{y=0}^t P(Y^{(n)} + Y_{n+1} \leq t, Q_0^{(n)} = k | Y_{n+1} = y)dF_Y(y) \\ &= F_{Y^{(n)},Q_0^{(n)}}(t,k) - \int_{y=0}^t F_{Y^{(n)},Q_0^{(n)}}(t-y,k)dF_Y(y) \end{aligned}$$

where the last equation follows from the independence of Y_{n+1} and $(\sum_{i=1}^n Y_i, \sum_{i=1}^n Q_i)$.

5.2 Proposed Framework

In this section, we will present a general framework to analyze the operating characteristics of the inventory policies within the policy class $\mathcal{P}$.

5.2.1 Analysis at the Warehouse

We first consider the analysis at the warehouse level and derive the steady-state distribution of the inventory position, the waiting time distribution of an order placed at the warehouse and the on-hand inventory of the warehouse and the order placement rate at the warehouse.

As explained above, the warehouse faces a compound renewal demand where the inter-order time Y and the order quantity Q_0 have a joint density given by $f_{Y,Q_0}(y,q)$. The warehouse employs a continuous review (s_0, S_0) ordering policy where s_0 and S_0 are the reorder and order-up-to levels of the warehouse, respectively. Each instance at which a retailer order is placed at the warehouse is a regeneration point for the retailers. Since each warehouse order is triggered by a retailer order and raises the inventory position of the warehouse up to S_0, the warehouse ordering epochs are regenerative for the warehouse. Hence, we know that the steady-state distribution of the inventory position of the warehouse exists (See Stidham [72]).

For $t > 0$, define the two dimensional stochastic process, $\xi(t) = \{IP_0(t), Z(t)\}$, where $IP_0(t)$ denotes the inventory position at time t and $Z(t)$ denotes the time between t and the last order arrival at the warehouse. The states of $\xi(t)$ will be denoted by (i, z) where $(i,z) \in [s_0+1, s_0+2, \ldots, S_0] \times [0, \infty)$. Let $g(t, i, z)$ denote the probability density function of $\xi(t)$. Assuming a steady state distribution exists, we have the following result:

Lemma 5.2.1 (a) *The steady state p.d.f., denoted by $g(i,z)$ is given by*

$$g(i,z) \quad = \quad C_i \overline{F}_Y(z) \quad \text{for} \quad (i,z) \in [s_0+1, s_0+2, \ldots, S_0] \times [0,\infty) \tag{5.2}$$

where C_i are normalizing constants and obtained by solving (5.3) and (5.4) below simultaneously.

$$C_i = \begin{cases} \sum_{j=s_0+1}^{S_0} \sum_{q=j-s_0}^{\infty} C_j P_{Q_0}(q) & \text{if} \quad i = S_0 \\ \\ \sum_{j=i+1}^{S_0} C_j P_{Q_0}(j-i) & \text{if} \quad i \in [s_0+1, s_0+2, \ldots, S_0-1] \end{cases} \tag{5.3}$$

$$\sum_{j=s_0+1}^{S_0} C_j = 1/E[Y] \tag{5.4}$$

(b) *Let IP_0 correspond to the inventory position of the warehouse at steady state. Then, the distribution of IP_0 is given by*

$$\pi_i \quad = \quad P(IP_0 = i) = C_i E[Y] \qquad i \in [s_0+1, s_0+2, \ldots, S_0]$$

(c) *Let $f_{Z(t)}(z)$ denote the probability density function of $Z(t)$. Then, the steady-state p.d.f. of $Z(t)$, denoted by f_Z is given by*

$$f_Z(z) \quad = \quad \overline{F}_Y(z)/E[Y] \quad \text{for} \quad z \geq 0$$

Proof: *See Appendix.*

A common approach in analyzing a multi-echelon inventory system is to separate the levels so that each level can be modeled as a single location inventory system whose parameters relate to some characteristics or performance of the upper and lower echelons. There are several ways to express the dependency of the upper and the lower echelons. To analyze the operating characteristics at the lower echelon, we need the waiting time of an order at the upper echelon which will determine the effective lead time of an order that a retailer faces. Waiting time of an order is an important performance measure of a warehouse for the backorders and it provides a linkage between the warehouse and the retailers.

Consider the time instances where the orders are placed at the warehouse. Suppose that at time t an order of size q has just been placed at the warehouse and we are interested in the distribution of $W_0(t,q)$, the waiting time of the order of size q which is placed at the warehouse at time t. $W_0(t,q)$ is the time that elapses between the arrival of the order at the warehouse, t, and the release of it from the warehouse (See Figure 5.3).

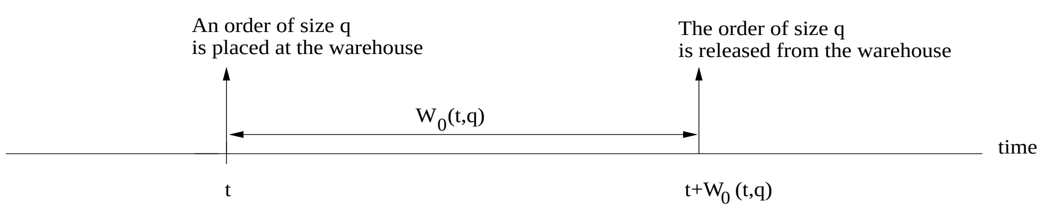

Figure 5.3: Illustration of waiting time of an order at the warehouse

We next present the following result which will be frequently used for the derivation of the waiting time distribution of an order.

Lemma 5.2.2 *As $t \to \infty$, $IP_0(t)$ and $D_0(t, t+\tau]$ are independent $\forall\ \tau > 0$.*
Proof: *See Appendix.*

Lemma 5.2.2 is evident for unit Poisson/compound Poisson demand due to the memoryless property. However, the mentioned independence is not obvious for a demand structure where the inter-demand time and demand size have a bivariate distribution. Sahin [63] has provided the proof for continuous demands. Based on Sahin [63], Kruse [49] has argued that the independence result also holds for discrete renewal demand where the inter-arrival times are independent of the demand sizes but has not provided any proof.

In inventory literature, there are only a few studies on waiting time distribution of orders in an (s, S) inventory system. To the best of our knowledge, Kruse [49] is the only study that derives the waiting time distribution of an order of random size with renewal discrete demand. Kruse [49] identifies each unit in a demand of size q by an index j $(j = 1, 2, \ldots, q)$ and derives the waiting time distribution of each unit in the demand. We will use a similar approach to that of Kruse [49] to compute the steady-state distribution of the waiting time of an order of size q. However, we will also allow s_0 to be negative and consider the joint distribution of Y and Q_0 whereas Kruse [49] only assumed non-negative s_0 values and independence of the demand sizes and the inter-demand time. Recently, Kiesmüller and de Kok [48] also considered the waiting time of an order arriving according to a compound renewal process (approximated as mixed Erlang distribution) under (s, Q) policy where $s \geq 0$.

Before we go on with the derivation of the waiting time distribution, we would like to point out some remarks on how retailer orders are satisfied at the warehouse.

i) There may be more than one warehouse order which is used to satisfy a retailer order.

ii) Although the retailer orders are placed in batches, it is more convenient to think of the items in a batch as if they have arrived one after the other in a chronological order.

iii) In conjunction with observation (*ii*) above, since partial shipments are not allowed, waiting time of an order is the waiting time of the last unit in a particular order.

iv) Due to the FIFO assumption while satisfying the orders, we have the following main observation: k^{th} unit of a batch order placed by the retailer is satisfied by the first warehouse order placed at or after a retailer order which covers the S_0^{th} item before the k^{th} unit in the batch order at the retailer, counted backwards in time. This observation will become more clear with the example illustrated below.

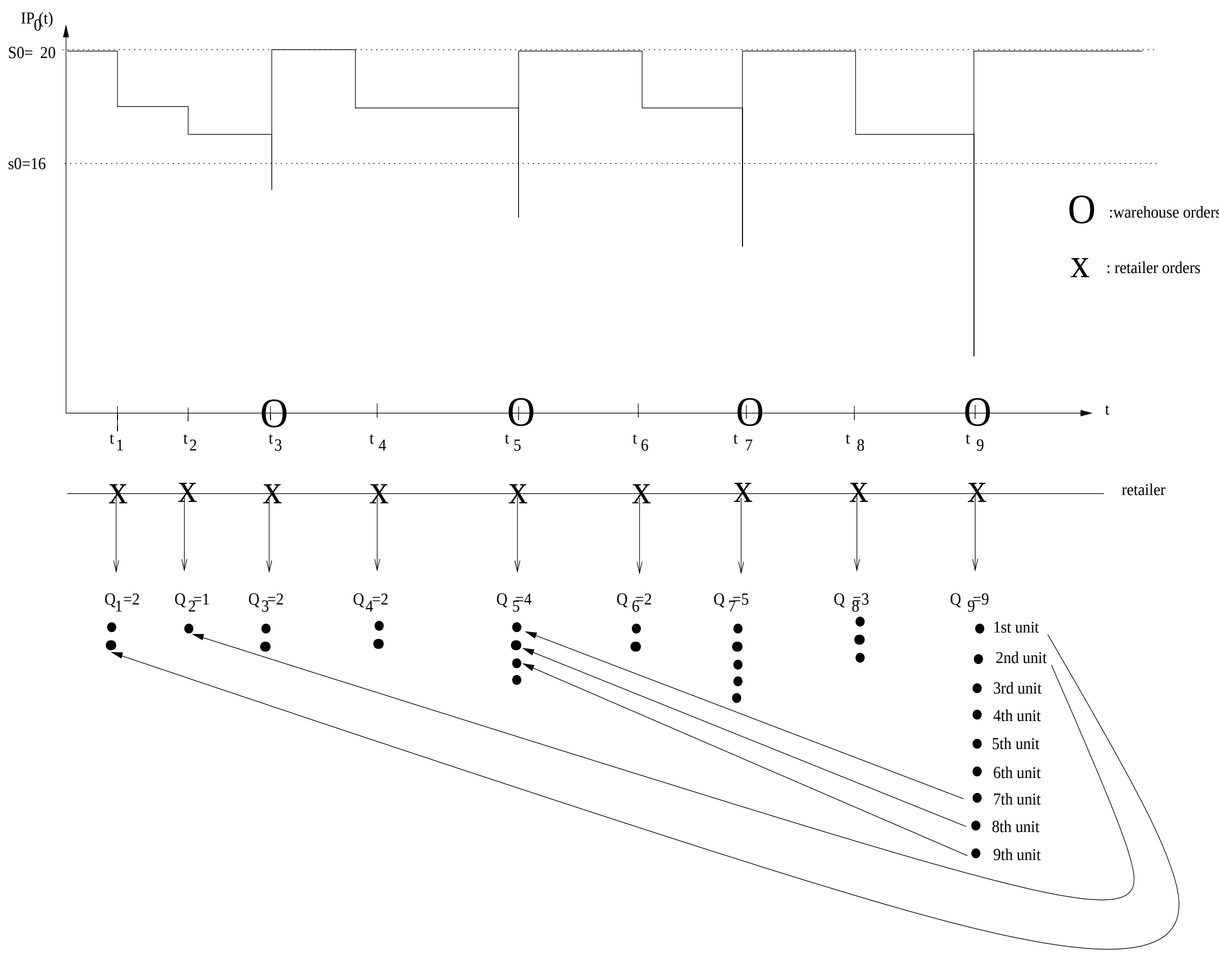

Figure 5.4: Illustration of satisfying an order at the warehouse - Example

We depict a graph of consecutive retailer and warehouse orders in Figure 5.4. We consider 9 retailer orders arriving at the times $t_1, t_2, \ldots, t_9$, respectively with order sizes $Q_i, i = 1, 2, \ldots, 9$ indicated in the graph. We assume that $s_0 = 16$ and $S_0 = 20$ and the inventory position of the warehouse just before time t_1 is $IP_0(t_1^-) = 20$. With the given retailer orders, the warehouse orders are placed at times t_3, t_5, t_7, t_9. We are interested in how the order of size $Q_0 = 9$, placed at time $t = t_9$ is satisfied.

Firstly, we think of this order of 9 units, as if it is composed of 9 unit demands that have arrived one after the other. The retailer order covering the 20^{th} (that is the S_0^{th}) item before the first unit in the order at time $t = t_9$ is placed at time t_1 since $\sum_{k=2}^{8} Q_k = 19 < 20$ and $\sum_{k=1}^{8} Q_k = 21 \geq 20$. The retailer order covering the 20^{th} item before the second unit of the order at time $t = t_9$ is placed at time t_2 because $\sum_{k=3}^{8} Q_k + 1 = 19 < 20$ and $\sum_{k=2}^{8} Q_k + 1 = 20 \geq 20$. In a similar manner, for this illustration, the retailer order covering the S_0^{th} item before the j^{th} unit in the order is given at t_k if $\sum_{i=k+1}^{8} Q_i + j - 1 < S_0$ and $\sum_{i=k}^{8} Q_i + j - 1 \geq S_0$. Therefore, the retailer order that covers the 20^{th} (S_0^{th}) item before the 9^{th} unit of the order placed at t_9 is placed at time t_4.

The first warehouse order placed after or at the retailer orders arriving at t_1, t_2, t_3 is given at time t_3 and hence the first four units in the retailer order placed at time t_9 are satisfied by this warehouse order. Similarly, the first warehouse order after the retailer orders placed at t_4 and t_5 is given at time t_5. Therefore, the last five units are satisfied by this warehouse order placed at t_5. For this example, the order of size $q = 9$ is satisfied by two warehouse orders and since we do not allow partial shipments the retailer order is totally satisfied whenever the warehouse order placed at time t_5 arrives. Therefore, the waiting time of the order is the waiting time of the $q = 9^{th}$ unit in the order. The retailer and warehouse orders corresponding to each unit in the order given at time $t = t_9$ are presented in Figure 5.5.

Since we do not allow partial shipments, an order of size q can be satisfied without delay if there are at least q units on-hand and none of them has been tagged to a previous order. In view of the remarks given above, we next consider the different cases for $W_0(t, q)$ with respect to the values of s_0 and $IP_0(t^-)$, which is the inventory position of the warehouse just before the order arrives.

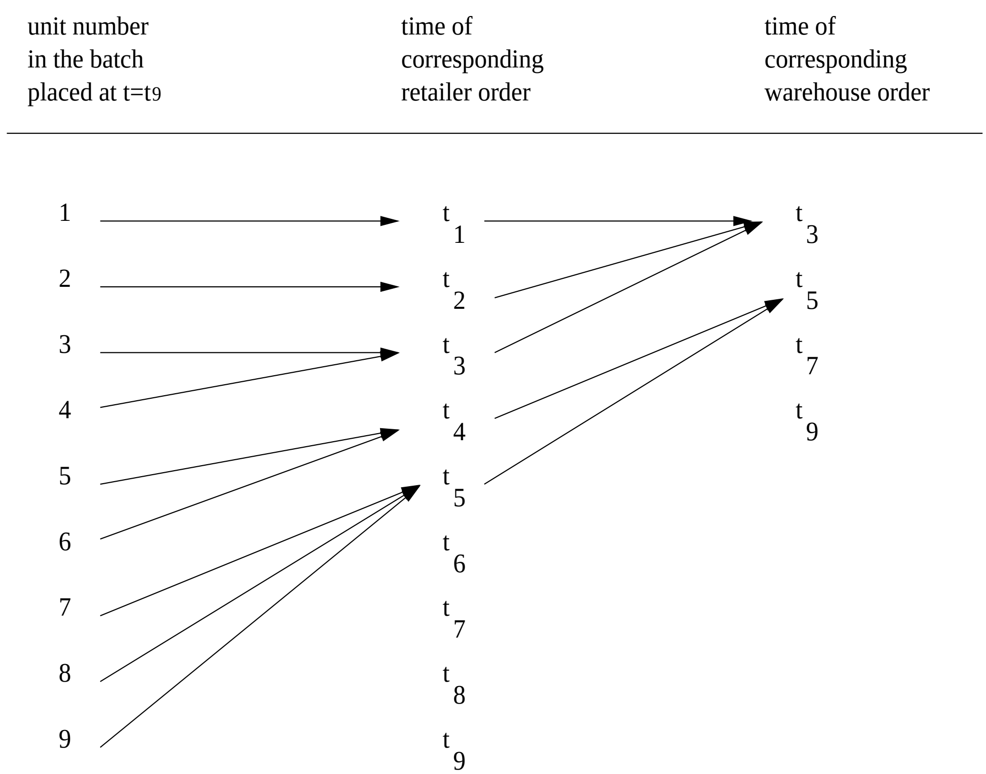

Figure 5.5: Retailer and warehouse order corresponding to units in the order - Example

i) Suppose $s_0 \geq 0$. An arriving order is satisfied either by a warehouse order which has been placed previously or by the order triggered by itself. Therefore, $0 \leq W_0(t,q) \leq L_0$.

- Suppose that an arriving order of size q finds $IP_0(t^-) \geq q$. The order will wait at most $\tau (\tau < L_0)$ units of time if and only if the q^{th} unit of the order is satisfied by one of the items of $IP_0(t+\tau-L_0)$, which will be available as on-hand inventory by time $t+\tau$. Therefore,

$$\{W_0(t,q) \leq \tau\} \Leftrightarrow \{IP_0(t+\tau-L_0) - D_0[t+\tau-L_0,t) \geq q\} \quad \text{for} \quad \tau < L_0 \tag{5.5}$$

- Now, suppose that the order finds $IP_0(t^-) < q$. Then, this order is satisfied by the warehouse order that is triggered by itself. Hence, the waiting time of the order is L_0.

$$\{W_0(t,q) = L_0\} \Leftrightarrow \{IP_0(t^-) < q\} \tag{5.6}$$

ii) Now, suppose $s_0 < 0$. In this case, the order of interest may have to wait for the next warehouse order placed, we may also have $W_0(t,q) \geq L_0$.

- If the arriving order finds $IP_0(t^-) \geq q$, we have the same scenario as $s_0 \geq 0$. Therefore, we can write

$$\{W_0(t,q) \leq \tau\} \Leftrightarrow \{IP_0(t+\tau-L_0) - D_0[t+\tau-L_0,t) \geq q\} \quad \text{for} \quad \tau < L_0 \tag{5.7}$$

- If the order finds $IP_0(t^-) \leq q + s_0$, then this order itself triggers an order placement. The waiting time of the order is therefore L_0.

$$\{W_0(t,q) = L_0\} \Leftrightarrow \{IP_0(t^-) \leq s_0 + q\} \tag{5.8}$$

- Now, suppose that the order of size q arriving at time t finds $s_0 + q < IP_0(t^-) < q$. Then, this order can neither be satisfied by a previously placed warehouse order nor can trigger a warehouse order itself. Therefore, it will wait for the arrival of the next warehouse order. If a warehouse order is triggered in $(t, t+\tau-L_0]$ it will be available by time $t+\tau$ and hence $L_0 < W_0(t,q) \le \tau$.

$$\{L_0 < W_0(t,q) \le \tau\} \Leftrightarrow \{IP_0(t^-) - D_0[t, t+\tau-L_0] \le s_0\} \quad \text{for} \quad \tau > L_0 \tag{5.9}$$

In view of the arguments presented above, we next present the steady-state distribution of the waiting time of an order.

Lemma 5.2.3 *The steady-state distribution of the waiting time of an order of size q, $F_{W_0(q)}(\tau)$ is given as follows:*
(a) For $s_0 \ge 0$,

$$F_{W_0(q)}(\tau) = \begin{cases} \sum_{i=\max(s_0+1,q)}^{S_0} \sum_{k=0}^{i-q} \pi_i \varphi(L_0-\tau,k) & \text{if} \quad \tau < L_0 \\ 1 & \text{if} \quad \tau \ge L_0 \end{cases}$$

(b) For $s_0 < 0$,

$$F_{W_0(q)}(\tau) = \begin{cases} \sum_{i=q}^{S_0} \sum_{k=0}^{i-q} \pi_i \varphi(L_0-\tau,k) & \tau < L_0 \\ 1 - \sum_{i=s_0+q+1}^{q-1} \pi_i & \tau = L_0 \\ 1 - \sum_{i=s_0+q+1}^{q-1} \sum_{k=0}^{i-q-s_0-1} \pi_i \varphi(\tau-L_0,k) & \tau > L_0 \end{cases}$$

Proof: *See Appendix.*

In light of the above discussions, an order placed at the warehouse waits for a random amount of time if the entire order quantity is not immediately available on at the shelf. This results in an *effective* lead time, $T_i(q)$ for an order of size q for retailer i which is composed of the lead time L_i for retailer i and the order waiting time at the warehouse, $W_0(q)$. Hence, we have $T_i(q) = L_i + W_0(q)$.

Next, we consider $OH_0(t)$, the on-hand inventory level at time t. To compute the on-hand inventory level at the warehouse at any time, we employ the standard argument of Hadley and Whitin [42] and consider the system at the time instances t and $t+L_0$, where L_0 is the constant replenishment time of the warehouse. With this choice of the time interval, we observe that all outstanding orders at time t and no orders placed afterward have arrived by time $t+L_0$.

In the standard argument of Hadley and Whitin [42] with unit Poisson demands single retailer or single item, the inventory position at time t and the demand during lead time are sufficient to find the on-hand inventory or backorder level at time t, together with the above observation. However, in the present setting of joint replenishment with more than one retailer and bivariate behaviour of (Y, Q_0), the inter-order time and the order quantity, as well as the integrality assumption regarding order release policy at the warehouse, similar information is not sufficient to determine the on-hand inventory level. To illustrate this complication, we consider the example below:

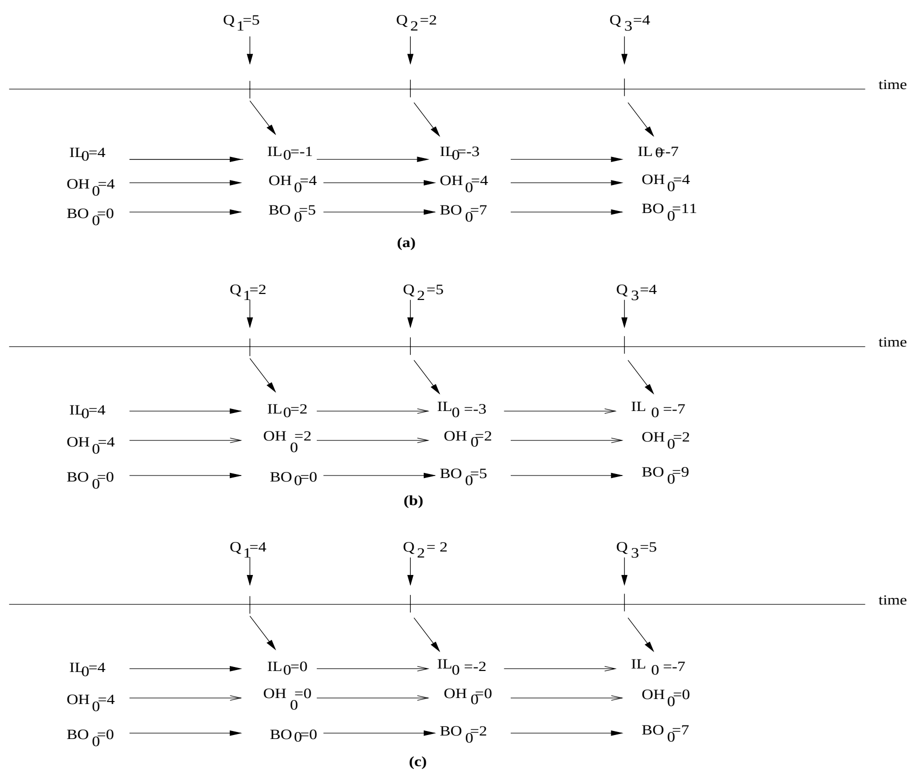

Figure 5.6: Effect of order sequences on on-hand inventory at the warehouse

Suppose that the initial inventory level, IL_0, on-hand inventory, OH_0 and backorder level, BO_0 of the warehouse are 4, 4 and 0, respectively. Suppose also that three orders have generated a total of 11 units with order sizes 5, 2 and 4 in a fixed time interval. We also assume that initially there are no outstanding orders at the warehouse and no warehouse orders are triggered during the time interval that we consider. Figure 5.6(a) presents IL_0, OH_0 and BO_0 values after each order when the sequence of the orders is 5, 2 and 4. Since the initial on-hand inventory ($OH_0 = 4$) is not enough to satisfy the first order of size 5, it is backordered and the existing 4 units in stock are reserved for the first order. Therefore, neither the second nor the third order can be satisfied directly from the existing on hand inventory. After the third order, there are four units in stock which are still reserved for the first order with a total of 11 backorders and inventory level of -7. Figure 5.6(b) and Figure 5.6(c) correspond to the alternative cases in which the sequence of the orders are switched to 2, 5, 4 and 4, 2, 5, respectively. In both of these cases, the units of the first order are satisfied directly from the existing inventory on-hand. In Figure 5.6(b), there are 2 units left in stock after the first order is satisfied but they are not sufficient to satisfy any further orders, resulting in 9 backorders and 2 units on hand. In Figure 5.6(c), no units are left after the first order is satisfied and, hence, we have 7 backorders after the third order arrival. Notice that the inventory levels at the end of the third order are the same for all realizations whereas the on-hand inventory and backorder levels are different and determined by the sequence of the orders.

As the above discussion indicates, in order to find the on-hand inventory, we need the sequence of the order quantities as well as the inter-order times in addition to the information in the Hadley and Whitin [42] setting. Suppose that at time t, the state of the system is $\xi(t) = (IP_0(t), Z(t)) = (i, z)$. We next consider the possible cases in detail:

1. If the total number of units demanded during the interval t and $t + L_0$, $D_0(t, t+L_0] < i$ then $OH_0(t+L_0) = i - D_0(t, t+L_0]$. Here, $D_0(t, t+L_0] = k$ if a total of $n \leq k$ orders are placed, *ie.* $\sum_{i=1}^{n} Y_i \leq z + L_0, \sum_{i=1}^{n+1} Y_i > z + L_0$ and a total of k units are demanded in n orders, $\sum_{i=1}^{n} Q_i = k$.

2. If $D_0(t, t + L_0) \geq i$, there may or may not be on-hand inventory at time $t + L_0$ due to the restriction on the order release policy at the warehouse.

Therefore, the inventory level at time $t + L_0$ can not be determined solely by (i, z) and $D_0(t, t + L_0]$. Rather, the sequence of the size of the orders arriving during $(t, t + L_0]$ is needed to find $OH_0(t + L_0)$ as also illustrated in an example in Figure 5.6.

Suppose that the total number of units demanded between time t and time $t + L_0$ exceeds i at the $(n+1)$'st order for the first time. That is $\sum_{k=1}^{n+1} Y_k \leq z + L_0, \sum_{k=1}^{n} Q_k < i$ and $\sum_{k=1}^{n+1} Q_k \geq i$ hold. Due to order integrality on delivery, the $(n + 1)$'st order can not be satisfied immediately from stock even though $OH_0(t + L_0) = i - \sum_{k=1}^{n} Q_k \geq 0$. Hence, Q_{n+1} units will be backordered as well as all the orders arriving after the $n + 1$'st order until time $t + L_0$ (if there are any) although there are still $i - \sum_{k=1}^{n} Q_k$ sitting in inventory at time $t + L_0$.

In view of these scenarios, given $(IP_0(t), Z_0(t)) = (i, z)$, we can write $OH_0(t+L_0)$ as follows:

$$OH_0(t + L_0) = \begin{cases} i & \text{if } \{Y_1 > z + L_0\} \text{ or } \{Y_1 \leq L_0 + z, Q_1 > i\} \\ \\ i - k & \text{if } \sum_{i=1}^{n} Y_i \leq z + L_0, \sum_{i=1}^{n} Q_i = k, \sum_{i=1}^{n+1} Q_i > i \\ & \text{for } 0 < k \leq i, n \geq 1 \end{cases} \tag{5.10}$$

Lemma 5.2.4 *Let OH_0 denote the steady-state on-hand inventory of the warehouse. Then, the distribution of OH_0 is given by*

$$\vartheta_i = P(OH_0 = i) = C_i \int_{z=0}^{\infty} \left[\overline{F}_Y(L_0 + z) + \sum_{j=i+1}^{\infty} [F_{Y,Q_0}(L_0 + z, j) - F_{Y,Q_0}(z, j)] \right] dz$$
$$+ \overline{F}_{Q_0}(i) \sum_{k=1}^{S_0 - i} \sum_{n=1}^{\infty} \sum_{j=1}^{k} C_{i+k} \int_{z=0}^{\infty} \int_{t=z}^{L_0+z} F_{Y^{(n-1)}, Q_0^{(n-1)}}(L_0 + z - t, k - j) dF_{Y,Q_0}(t, j) dz$$
$$i = \max(s_0 + 1, 1), \ldots, S_0$$

Proof: *See Appendix.*

Using Lemma 5.2.4, we can find the steady-state expected on-hand inventory as $E[OH_0] = \sum_{i=\max(s_0+1,1)}^{S_0} i\vartheta_i$.

Since the inventory position of the warehouse is S_0 at each warehouse order, the ordering instance is a regeneration point of the warehouse and we define a warehouse cycle as the time between two consecutive warehouse ordering instances. Next, we will derive $E[Y_0]$, expected warehouse cycle time.

Let $\eta(q) = \min\{n : \sum_{i=1}^{n} Q_i \geq q\}$. Then, Y_0 can be written as:

$$Y_0 = \sum_{i=1}^{\eta(S_0 - s_0)} Y_i$$

Since $Y_1, Y_2, \ldots$ are independent and identically distributed random variables with finite $E[Y]$, and $\eta(S_0 - s_0)$ is a stopping time for $Y_1, Y_2, \ldots$ such that $E[\eta(S_0 - s_0)] < \infty$, we can use Wald's equation [62] to find $E[Y_0]$ as follows:

$$E[Y_0] \quad = \quad E[\eta(S_0 - s_0)]E[Y] \tag{5.11}$$

where as proved in the Appendix, $E[\eta(S_0 - s_0)]$ is given as:

$$E[\eta(S_0 - s_0)] \quad = \quad \sum_{k=0}^{S_0 - s_0 - 1} F_{Q_0}^{(k)}(S_0 - s_0 - 1) \tag{5.12}$$

In each warehouse cycle, the warehouse ordering cost is incurred once. Hence the ordering cost rate at the warehouse is simply $K_0/E[Y_0]$.

As we indicated before, we assume that it is the warehouse who owns the units during their transportation from the outside supplier to the warehouse. Therefore, every time an order is placed by the warehouse, a unit holding cost of h_0 is incurred for Q_w units during L_0 time units where Q_w is the warehouse order size. Finally, we derive $E[Q_w]$, expected warehouse order size.

Let $\kappa(q) = \min\{n : n \geq q, \sum_{i=1}^{k} Q_i < q, \sum_{i=1}^{k} Q_i = n, k \leq n\}$. Then, Q_w can be written as $Q_w = \kappa(S_0 - s_0)$. As we prove in the Appendix, $E[Q_w]$ is given by the following expression:

$$E[Q_w] \quad = \quad \sum_{n=S_0 - s_0}^{\infty} n \left[\sum_{k=1}^{n} \sum_{j=k-1}^{S_0 - s_0 - 1} P_{Q_0}^{(k-1)}(j) P_{Q_0}(n-j) \right] \tag{5.13}$$

In a warehouse cycle, the inventory carrying cost of a warehouse order during the transportation is given by $h_0 E[Q_w] L_0$.

5.2.2 Analysis at the Retailers

In the previous section, we have determined the steady-state distribution of the waiting time of an order of a given size placed at the warehouse. That is the only information needed at the warehouse level to characterize the operating characteristics at the retailers. At the lower echelon, we base our analysis on the calculation of the cost that an order incurs at each retailer.

Suppose that an order of size q is placed at the warehouse and this order consists of individual retailer order quantities $R_i(q) = m_i, i = 1, 2, \dots, N$. Then, at retailer i, the first of the m_i units in the order is used to satisfy the $(S_i - m_i + 1)$'th demand following the order since the inventory position of retailer i just before the order is placed at the warehouse is $S_i - m_i$. Similarly, the second of the m_i units in the order is tagged to the $(S_i - m_i + 2)$'th demand following the order. In general, the jth of the m_i units in that order is used to satisfy the $(S_i - m_i + j)$'th future demand of retailer i.

In order to calculate the holding cost at the retailers, our approach is based on computing the age of an order at the retailers. We denote the *age* of the m_i units allocated to retailer i in an order of size q by $AR_i(m_i, q)$ and define it as the sum of the times that m_i units spend at the retailer i until they satisfy a demand.

If a unit in the order does not spend any time at retailer i and is immediately used to fulfill a demand that has been waiting at retailer i, then this means that it is used to satisfy a backordered demand. Let $B_i(m_i, q)$ be the number of items of m_i units which are used to satisfy backordered demands at retailer i. $B_i(m_i, q)$ is the number of demands which are backordered and eventually satisfied by the units allocated to retailer i in the order of size q.

From the previous section, we know that the effective lead time of an order of size q for retailer i is $T_i(q) = L_i + W_0(q)$. Therefore, the j'th demand following the order $(0 \leq j \leq S_i - m_i)$ will be satisfied by an item which has spent $[X_j^i - T_i(q)]^+$ units of time at retailer i, where $[x]^+$ is $\max(x, 0)$ and X_j^i denotes the arrival time of the j'th demand at retailer i after the order of size q has been placed, which has an Erlang distribution with parameters j and λ_i. Also, observe that if $X^i_j < T_i(q)$ then the demand of interest will be backordered . In view of these observations, regarding the age of an order and the number of units satisfying the backordered demands, we have the following:

i) Suppose $m_i = 0$. Then retailer i is not included in the order. Hence, $AR_i(m_i, q) = B_i(m_i, q) = 0$

ii) $m_i > 0, m_i \leq S_i$. Since the inventory position at the order instance $S_i - m_i \geq 0$ is non-negative, all m_i units will be used to satisfy future demands at retailer i. We can write $AR_i(m_i, q)$ as follows:

$$AR_i(m_i, q) = \sum_{j=1}^{m_i} \left[X^i_{S_i-j} - T_i(q)\right] I(X^i_{S_i-m_i+j} > T_i(q)) \tag{5.14}$$

Here, the term $m_i L_i$ represents the total age of m_i items when they arrive at retailer i. Recall that the retailer possesses the items during their transportation from the warehouse to the retailer.

Similarly, $B_i(m_i, q)$ is written as:

$$B_i(m_i, q) = \begin{cases} 0 & \text{if } T_i(q) \leq X^i_{S_i-m_i+1} \\ m_i - k & \text{if } X^i_{S_i-k} < T_i(q) < X^i_{S_i-k+1}, 1 \leq k \leq m_i - 1 \\ m_i & \text{if } X^i_{S_i} < T_i(q) \end{cases}$$

iii) $m_i > 0, m_i > S_i$. Since the inventory position at the order instance $S_i - m_i < 0$ is negative, the first $(m_i - S_i)$ units in the order will be used to satisfy the demands which arrive and is backordered before the order of interest is placed at the warehouse. The first unit in the order will satisfy the first backordered demand, ie. the demand which arrives when the inventory position of retailer i is 0. The second unit in the order will satisfy the demand which arrives when the inventory position of retailer i is -1. Similarly, the $(m_i - S_i)$'th demand will satisfy the demand which arrives when the inventory position of retailer i is $S_i - m_i + 1$. Since these $(m_i - S_i)$ units are used as soon as they arrive at the retailer, $AR_i(m_i, q)$ can be written as follows:

$$AR_i(m_i, q) = \sum_{j=m_i-S_i+1}^{m_i} \left[X^i_{S_i-j} - T_i(q)\right] I(X^i_{S_i-m_i+j} > T_i(q)) \tag{5.15}$$

Since $(m_i - S_i)$ units will be used to satisfy the backordered demands, we can write $B_i(x_i, q)$ as:

$$B_i(m_i, q) = (m_i - S_i) + \begin{cases} 0 & \text{if } T_i(q) \leq X^i_1 \\ k & \text{if } X^i_k < T_i(q) \leq X^i_{k+1}, k = 1, 2, \ldots, S_i \\ S_i & \text{if } X^i_{S_i} < T_i(q) \end{cases}$$

Lemma 5.2.5 *Let $F(x, k, \lambda)$ be the c.d.f. of an Erlang random variable with shape and scale parameters, k and λ, respectively. Then, given m_i and q, $E[AR_i(m_i, q)]$ and $E[B_i(m_i, q)]$ are given by:*

$$E[AR_i(m_i,q)] = \begin{cases} 0 & \text{if } m_i = 0 \\ \sum_{j=1}^{m_i} \int_0^\infty \left[\frac{S_i-m_i+j}{\lambda_i}\overline{F}(L_i+w, S_i-m_i+j+1, \lambda_i)\right] dF_{W_0(q)}(w) \\ \quad - \sum_{j=1}^{m_i} \int_0^\infty \left[(L_i+w)\overline{F}(L_i+w, S_i-m_i+j, \lambda_i)\right] dF_{W_0(q)}(w) & \text{if } S_i \geq m_i > 0 \\ \sum_{j=m_i-S_i+1}^{m_i} \int_0^{L_0} \left[\frac{S_i-m_i+j}{\lambda_i}\overline{F}(L_i+w, S_i-m_i+j+1, \lambda_i)\right] dF_{W_0(q)}(w) \\ \quad - \sum_{j=m_i-S_i+1}^{m_i} \int_0^{L_0} \left[(L_i+w)\overline{F}(L_i+w, S_i-m_i+j, \lambda_i)\right] dF_{W_0(q)}(w) & \text{if } m_i > 0, S_i < m_i \end{cases}$$

and

$$E[B_i(m_i,q)] = \begin{cases} 0 & \text{if } m_i = 0 \\ \sum_{j=1}^{m_i} \int_0^\infty F(L_i+w, S_i-m_i+j, \lambda_i) dF_{W_0(q)}(w) & \text{if } m_i > 0, S_i \geq m_i \\ (m_i - S_i) + \sum_{j=m_i-S_i+1}^{m_i} \int_0^\infty F(L_i+w, S_i-m_i+j, \lambda_i) dF_{W_0(q)}(w) & \text{if } m_i > 0, S_i < m_i \end{cases}$$

Proof: *See Appendix.*

5.2.3 Optimization Problem

In this section, we construct the cost rate function and define the optimization problem.

Recall that holding and ordering costs are incurred at the retailers and backorder costs are handled implicitly by considering retailer fill rates. Let $TC(q)$ denote the expected cost of an order of size q incurred by the retailers. The common ordering cost of the retailers, K is incurred for each order placed at the warehouse. Based on the analysis described in the previous section, $TC(q)$ can be written as:

$$TC(q) = K + \sum_{i=1}^{N} \sum_{m_i=0}^{q} h_i E[AR_i(m_i, q)] P_{R_i(q)}(m_i)$$

Let C_R denote the expected cost rate of the retailers. Since each ordering instance is a regenerative point for the retailers, in view of renewal reward theorem [62], C_R is given by

$$C_R = \sum_{q=1}^{\infty} \frac{P_{Q_0}(q) TC(q)}{E[Y]} + \sum_{q=1}^{\infty} \frac{P_{Q_0}(q) \sum_{i=1}^{N} \sum_{m_i=0}^{q} h_i L_i m_i P_{R_i(q)}(m_i)}{E[Y]} \quad (5.16)$$

Let $E[BT_i]$ be the expected number of backorders given by retailer i per time unit. Using renewal reward theorem [62], we can write $E[BT_i]$ as:

$$E[BT_i] = \sum_{q=1}^{\infty} \sum_{m_i=0}^{q} \frac{P_{Q_0}(q) E[B_i(m_i, q)] P_{R_i(q)}(m_i)}{E[Y]} \quad (5.17)$$

Then, modified fill rate of retailer i [78], γ_i, can be written as :

$$\gamma_i = 1 - E[BT_i]/\lambda_i$$

At the warehouse level, recall that K_0 is incurred only once in a warehouse cycle. Since the warehouse takes in charge of the holding of the items during the transportation time from the outside supplier to the warehouse (L_0 time units), the expected cost rate of the warehouse, C_W can be written as:

$$C_W = \frac{K_0 + h_0 E[Q_w] L_0}{E[Y_0]} + h_0 E[OH_0] \quad (5.18)$$

Finally, we are ready to formulate the objective function, expected total cost rate, $AC_{\mathcal{P}}$ under policy $\mathcal{P}$.

$$AC_{\mathcal{P}} = C_W + C_R \tag{5.19}$$

Since the system is a no-lost sales system, rewriting the pipeline inventory, Equation (5.19) can be written as follows:

$$AC_{\mathcal{P}} = \frac{K_0}{E[Y_0]} + h_0 E[OH_0] + \sum_{q=1}^{\infty} \frac{P_{Q_0}(q) TC(q)}{E[Y]} + h_0 \lambda_0 L_0 + \sum_{i=1}^{N} h_i \lambda_i L_i \tag{5.20}$$

Using a fill rate constraint, the optimization problem can be stated as follows:

$$\begin{gathered} \min AC_{\mathcal{P}} \\ \text{s.t.} \\ \gamma_i \geq \overline{\gamma_i} \quad i = 1, 2, \ldots, N \end{gathered}$$

where $\overline{\gamma_i}$ is the target fill rate of retailer i.

Note that the methodology developed herein is solely based on the policy class $\mathcal{P}$ and uses the times between orders, Y and size of corresponding orders, Q_0 and is not specific to any of the policies defined in Section 5.1. Also, notice that the policy class $\mathcal{P}$ is not restricted to the cited policies and any policy in which the retailers are regenerative at the order instances can be studied under the policy class $\mathcal{P}$.

In the next chapter, we will explain the four mentioned policies within the policy class, $\mathcal{P}$ with all the particulars and provide detailed expressions of the operating characteristics under each of these policies.

Chapter 6

Joint Replenishment Policies within Class $\mathcal{P}$

In the previous chapter, we have introduced four joint replenishment policies for a two echelon inventory system. The common feature of these policies allowed us to present a framework to solve them under a single policy class $\mathcal{P}$. Recall that the framework did not include any specifics of these policies but rather it was based on the joint distribution of (Y, Q_0) and the convolutions.

Similar to a multi-item, single-location setting, a joint replenishment policy employed by multiple retailers for a single-item is the generator of the opportunity arrival process. By choosing a particular policy to employ, we also choose a particular generation mechanism for the opportunities. The overall costs incurred by the inventory system depend greatly on how these opportunities arrive at the system. Similarly, the policies within class $\mathcal{P}$ differ in how the ordering instances or the ordering opportunities are generated. In this chapter, we will restate the individual policies and explain the opportunity generation and joint order mechanism for each policy in detail. We will mainly present expressions for the convolution of (Y, Q_0) if closed forms exist and provide approximations whenever necessary. We will also provide sample realizations for the key operating characteristics of these policies, steady-state distribution of warehouse inventory position and waiting time distributions. In Section 6.1 and

6.2, we provide a detailed analysis of $(Q, \mathbf{S})$ and $(Q, \mathbf{S}, T)$ policies. Sections 6.3 and 6.4 examine $(Q, \mathbf{S}|T)$ and $(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$ policies, respectively. We note that the policy statements were given in Section 5.1 and hence we will not repeat them in this chapter.

6.1 $(Q, \mathbf{S})$ Policy

As indicated before, the $(Q, \mathbf{S})$ policy was originally proposed by Renberg and Planche [60] and analyzed by Pantumsinchai [58] in a single location, multi-item inventory system. In a more recent study, the $(Q, \mathbf{S})$ policy was studied in a single-item, two-echelon inventory system by Cheung and Lee [23]. Assuming a (Q, R) policy at the warehouse, they presented an exact approach for the model using the results of Chen and Zheng [22].

In this section, we present how we implement the proposed framework for the $(Q, \mathbf{S})$ policy.

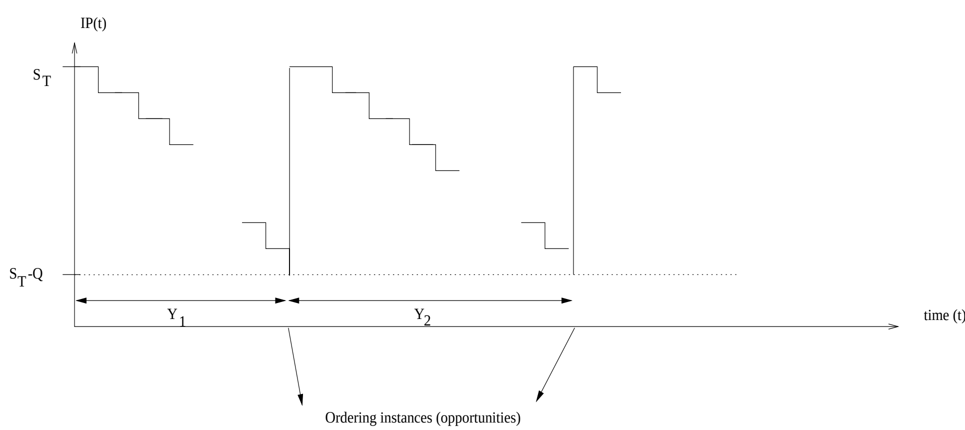

Figure 6.1: Illustration of Ordering Instances for $(Q, \mathbf{S})$ Policy

As illustrated in Figure 6.1, the ordering opportunities for the retailers arise whenever Q demands accumulate in the system, ie. the total inventory position of the retailers, $IP(t) = \sum_{i=1}^{N} IP_i(t)$ drops to $S_T - Q$ where $S_T = \sum_{i=1}^{N} S_i$. Every time an opportunity arises in the system, an order is placed and hence the time between consecutive orders have $Erlang_Q$ distribution.

$$f_{Y,Q_0}(y,q) \;=\; f(y,Q,\lambda_0) \quad \text{if} \quad q = Q, y > 0 \tag{6.1}$$

and the marginals have the following distributions:

$$P_{Q_0}(Q) = 1, \quad f_Y(y) = f(y,Q,\lambda_0) \quad \text{for} \quad y \geq 0 \tag{6.2}$$

Since the orders always have a size of Q and arrive according to an $Erlang_Q$ distribution, it is obvious that

$$F_{Y^{(n)},Q_0^{(n)}}(y,q) \;=\; \begin{cases} 1 & \text{if} \quad n = 0, y \geq 0, q = 0 \\ F(y,nQ,\lambda_0) & \text{if} \quad n > 0, y > 0, q = nQ \end{cases} \tag{6.3}$$

Corollary 6.1.1 *Under $(Q,\mathbf{S})$ policy, $\varphi(t,k) = P(D_0(0,t] = k)$ is given by:*

$$\varphi(t,k) \;=\; \begin{cases} \overline{F}(t,Q,\lambda_0) & \text{if} \quad k = 0 \\ F(t,k,\lambda_0) - F(t,k+Q,\lambda_0) & \text{if} \quad k = nQ, n > 0 \end{cases}$$

Proof: *Using Lemma 5.1.1 and Equations (6.1) and (6.3) , we can write*

$$\varphi(t,k) \;=\; \begin{cases} \overline{F}(t,Q,\lambda_0) & \text{if} \quad k = 0 \\ F(t,k,\lambda_0) - \int_{y=0}^{t} F(t-y,k,\lambda_0) f(y,Q,\lambda_0)dy & \text{if} \quad k = nQ, n > 0 \end{cases}$$

The result follows after observing that $\int_{y=0}^{t} F(t-y,k,\lambda_0)f(y,Q,\lambda_0)dy = F(t,k+Q,\lambda_0)$.

Since we do not allow partial shipments, it is obvious that the optimal (s_0, S_0) values will be integer multiples of Q. Therefore, during the remaining part of this section, we will restrict ourselves to (s_0, S_0) values which are integer multiples of Q. Also note that the (Q,R) policy of the warehouse in Cheung and Lee [23] and (s_0, S_0) policy assumed in this study are equivalent with $s_0 = R$ and $S_0 = Q + R$.

Next, we consider the steady-state distribution of warehouse inventory position. Cheung and Lee [23] conjectures that IP_0 is uniformly distributed over $[s_0 + Q, \ldots, S_0 - Q, S_0]$. Therefore,

$$g(i,z) \;=\; \frac{\lambda_0}{Q\zeta_{\Delta_0}} \overline{F}(z,Q,\lambda_0) \quad \text{and} \quad \pi_i = \frac{1}{\zeta_{\Delta_0}} \quad \text{for} \quad i \in [s_0 + Q, \ldots, S_0 - Q, S_0], z \geq 0$$

where $\zeta_k = k/Q$ and $\Delta_0 = S_0 - s_0$. It can also be easily shown that $g(i,z)$ and π_i given above satisfy Equations (5.2)-(5.4).

Corollary 6.1.2 *Under the $(Q, \mathbf{S})$ policy, $F_{W_0(Q)}(w)$ is given by*
(a) For $s_0 \geq 0$,

$$F_{W_0(q)}(\tau) = \begin{cases} 1 - \frac{1}{\zeta_{\Delta_0}} \sum_{i=1}^{\zeta_{\Delta_0}} F(L_0 - \tau, s_0 + iQ, \lambda_0) & \text{if} \quad \tau < L_0 \\ 1 & \text{if} \quad \tau \geq L_0 \end{cases}$$

(b) For $s_0 < 0$,

$$F_{W_0(q)}(\tau) = \begin{cases} \frac{\zeta_{S_0}}{\zeta_{\Delta_0}} - \frac{1}{\zeta_{\Delta_0}} \sum_{i=1}^{\zeta_{S_0}} F(L_0 - \tau, iQ, \lambda_0) & \tau < L_0 \\ 1 - \frac{\zeta_{-s_0-Q}}{\zeta_{\Delta_0}} & \tau = L_0 \\ 1 - \frac{\zeta_{-s_0-Q}}{\zeta_{\Delta_0}} + \frac{1}{\zeta_{\Delta_0}} \sum_{k=1}^{\zeta_{-s_0-Q}} F(\tau - L_0, kQ, \lambda_0) & \tau > L_0 \end{cases}$$

Proof: *See Appendix.*

Sample realizations for the waiting time distribution of an order are illustrated in Figure 6.2.

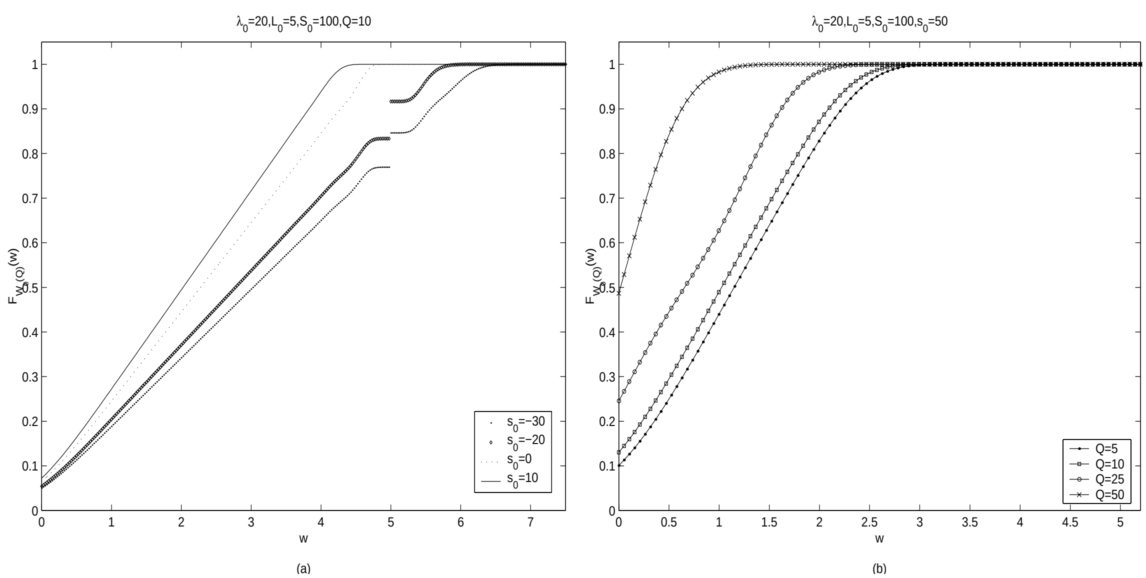

Figure 6.2: Realizations for $F_{W_0(Q)}$ under $(Q, \mathbf{S})$ Policy

Figure 6.2(a) depicts waiting time distribution of an order with different s_0 values. As expected, waiting time of an order with smaller s_0 values is stochastically larger than the waiting time of an order with higher s_0 values. If $s_0 < 0$, then the waiting time of an order has a probability mass at $w = 0$ and $w = L_0$. On the other hand, if $s_0 \geq 0$, the waiting time of an order has a probability mass at only $w = 0$ since the inventory position can not fall below the order quantity Q.

Under the same (s_0, S_0) values, one intuitively expects that an order with a larger size will have to wait longer at the warehouse. However, Figure 6.2(b) illustrates just an opposite case. For instance, for $Q = 5$, the probability that the order does not wait at the warehouse is 0.1008 whereas the probability of a zero waiting time of an order of size 10 is 0.1302. The corresponding figures are 0.2453 and 0.4867 for $Q = 25$ and $Q = 50$, respectively. Although this is counterintuitive, we interpret this as a result of the joint distribution of the inter-order time and order size. Recall that the inter-order time of an order of size Q has an Erlang_Q distribution and the inter-order time of an order with a larger size is stochastically smaller than the inter-order time of an order with a smaller size. Therefore, on the average, it takes a longer time for a larger order to arrive at the warehouse. The longer inter-order times allow the warehouse to accumulate the necessary stock. Although we can not prove in general, we observe that $W_0(5) >_{st} W_0(10) >_{st} W_0(25) >_{st} W_0(50)$ for this specific realization.

Finally, the distribution of the retailer order quantities are given as below which follows from the Poisson demand arrivals.

$$P_{R_i(Q)}(m_i) = \binom{Q}{m_i} r_i^{m_i}(1-r_i)^{Q-m_i} \quad \text{if} \quad 0 \leq m_i \leq Q$$

where r_i is, as defined before, the probability an arriving demand is for retailer i and given by $r_i = \lambda_i/\lambda_0$.

6.2 $(Q, \mathbf{S}, T)$ Policy

In Chapter 3, we proposed and analyzed the $(Q, \mathbf{S}, T)$ policy in a single location, multi-item setting.

As illustrated in Figure 6.3, the ordering opportunities may arise in two ways under the $(Q, \mathbf{S}, T)$ policy. Suppose for example that a total of Q demands have arrived before T time units have elapsed since the last ordering opportunity arrival; then, an order is placed at the instance of the Qth demand arrival. Suppose alternatively that T time units have elapsed before a total of Q demands have arrived. At this opportunity arrival, the inventory review may or may not result in an order placement. If at least one demand has arrived within the last T units of time, an order will be placed, otherwise nothing will be ordered. Therefore, opportunities are used if there is at least one retailer whose inventory position is below its order-up-to level.

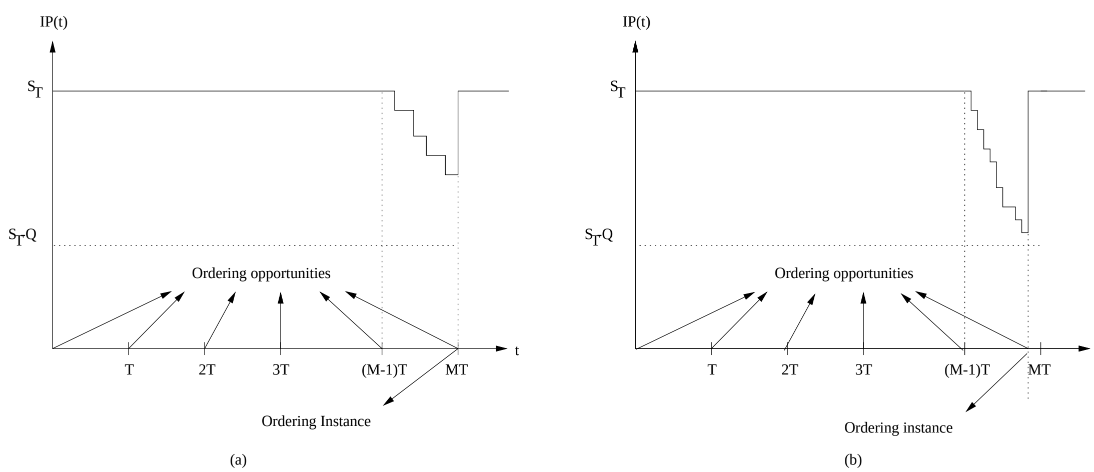

Figure 6.3: Illustration of Ordering Instances for $(Q, \mathbf{S}, T)$ Policy

Next, we consider the joint density of the convolutions of Y and Q_0, $f_{Y^{(n)},Q_0^{(n)}}$. However, $f_{Y,Q_0}(y,q)$ given in Lemma 3.2.1, has a very difficult structure to find the convolution and does not allow to calculate $f_{Y^{(n)},Q_0^{(n)}}$ directly. The difficulty encountered while finding $f_{Y^{(n)},Q_0^{(n)}}$ from $f_{Y^{(n)},Q_0^{(n)}}$ will be more clear with the example illustrated below.

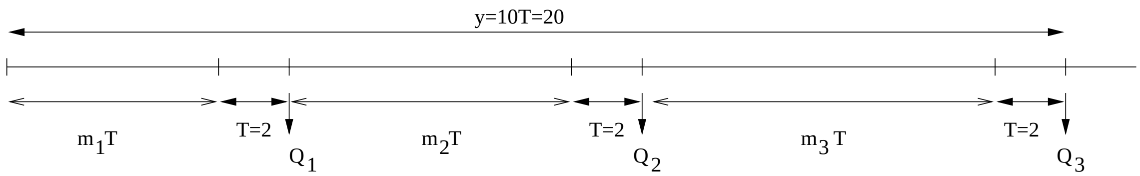

Figure 6.4: Illustration of 3-fold convolution of (Y, Q_0) - Example

Suppose that $Q = 3, T = 2$ and we consider $f_{Y^{(3)},Q_0^{(3)}}(y, q)$, the joint density of the arrival time of the third order and the total size of three orders where $y = 20, q = 4$. Since the size of an order should at least be one unit, none of the three orders of interest will have a size of $Q = 3$ units and hence all three orders will be triggered at review intervals. The inter-order time of each of the three orders will be a multiple of T. Then, in view of Equation (5.2) and Figure 6.4, we can write:

$$f_{Y^{(3)},Q_0^{(3)}}(20,4) = \sum_{\left\{\begin{array}{c} m_1+m_2+m_3=7, m_i \geq 0 \\ q_1+q_2+q_3=4, q_i \geq 1 \end{array}\right\}} \prod_{i=1}^{3} f_{Y,Q_0}(m_i T + T, q_i)$$

For this specific example, the total number of m_i's $\sum_{i=1}^{3} m_i = 7$ is 36 whereas the number of q_i's that satisfy $\sum_{i=1}^{3} q_i = 4$ is 3. Although n is taken as 3 and we have a very special case of the order compositions, *ie.* the order sizes are all less than Q units, we have to carry out a summation over 108 terms in order to find $f_{Y^{(3)},Q_0^{(3)}}(20, 4)$. This number increases extremely for larger values of n because we face a combinatorial problem in which we try to find all possible values of y_i's and q_i's such that $\sum_{i=1}^{n} y_i = y, \sum_{i=1}^{n} q_i = q$ as given in Equation (5.2). We next present an easier and a more compact approach in order to find $f_{Y^{(n)},Q_0^{(n)}}()$.

In order to use in our new approach, we first define a new function, $P'_{Q_0}(q)$ as:

$$P'_{Q_0}(q) = \begin{cases} P_{Q_0}(q) & \text{if} \quad 1 \leq q < Q \\ 0 & \text{o.w.} \end{cases}$$

We also define an additional random variable, Y_d which corresponds to the time since last decision epoch until an order is given. Recall that a decision epoch may correspond to an order placement or only a review instance. In the following corollary, we present the joint p.d.f. of Y_d and Q_0 as well as the conditional distribution of Y_d given $Q_0 = q$.

Corollary 6.2.1

a) Let $f_{Y_d,Q_0}(y,q)$ denote the joint probability density function of Y_d and Q_0. Then, $f_{Y_d,Q_0}(y,q)$ is given by

$$f_{Y_d,Q_0}(y,q) = \begin{cases} p_0(q,\lambda_0 T)/(1-\phi_0) & \text{if } y=T, 0<q<Q \\ \\ f(y,Q,\lambda_0)/(1-\phi_0) & \text{if } 0<y<T, q=Q \end{cases}$$

b) The conditional p.d.f. of Y_d given $Q_0 = q$, $f_{Y_d|Q_0}(y|q)$ is given by the following expression:

$$f_{Y_d|Q_0}(y|q) = \begin{cases} 1 & \text{if} \quad y=T, 0<q<Q \\ \\ f_{Te}(y,T,Q,\lambda_0) & \text{if} \quad 0<y<T, q=Q \end{cases}$$

where $f_{T_e}(x,T,Q,\lambda_0)$ corresponds to p.d.f. of a Truncated $Erlang_Q$ (at T) random variable given by

$$f_{T_e}(x,T,Q,\lambda_0) = f(x,Q,\lambda_0)/F(T,Q,\lambda_0) \quad \text{if} \quad 0<x<T$$

Proof: *See Appendix.*

In view of Corollary 6.2.1, we see that if the order has size Q, the time elapsed since the last decision epoch until this order has a truncated $Erlang_Q$ distribution. If the order has a size less than Q, the time elapsed since the last decision epoch until the order is placed is obviously T.

We next present the p.d.f. of the n^{th} convolution of (Y_d, Q_0) which is used to find the p.d.f. of the n^{th} convolution of (Y, Q_0).

Corollary 6.2.2

a) Under $(Q,\mathbf{S},T)$ policy, $f_{Y_d^{(n)},Q_0^{(n)}}(y,q)$ is given by the following expression:

$$f_{Y_d^{(n)},Q_0^{(n)}}(y,q) = \begin{cases} f_{T_e}^{(n)}(y,T,Q,\lambda_0)P_{Q_0}(Q)^n & \text{if} \quad 0<y<nT, q=nQ \\ \\ \sum_{m=1}^{n} C(n,m)P_{Q_0}(Q)^{n-m} P_{Q_0}'^{(m)}(q-(n-m)Q) f_{T_e}^{(n-m)}(y-mT,T,Q,\lambda_0) \\ \qquad \text{if} \quad 0<y\le nT, n\le q<nQ \end{cases}$$

where $P'^{(n)}_{Q_0}$ *is the* n^{th} *convolution of* P'_{Q_0}.
b) Under $(Q, \mathbf{S}, T)$ *policy,* $f_{Y(n),Q_0^{(n)}}(y,q)$ *is given by the following expression:*

$$f_{Y(n),Q_0^{(n)}}(y,q) = \sum_{k=\lceil y/T-n\rceil}^{\lfloor y/T\rfloor} C(n+k-1,k)\phi_0^k(1-\phi_0)^n f_{Y_d^{(n)},Q_0^{(n)}}(y-kT,q) \quad \text{if} \quad y>0, n\le q\le nQ$$

where $\lfloor x\rfloor$ *is the largest integer less than or equal to* x *and* $\lceil x\rceil$ *is the smallest integer larger than or equal to* x.

Proof: *See Appendix.*

Using Corollary 6.2.2, we can immediately write:

$$F_{Y(n),Q_0^{(n)}}(y,q) = \sum_{k=\lceil y/T-n\rceil}^{\lfloor y/T\rfloor} C(n+k-1,k)\phi_0^k(1-\phi_0)^n F_{Y_d^{(n)},Q_0^{(n)}}(y-kT,q) \quad \text{if} \quad y>0, n\le q\le nQ \tag{6.4}$$

where

$$F_{Y_d^{(n)},Q_0^{(n)}}(y,q) = \begin{cases} F_{T_e}^{(n)}(y,T,Q,\lambda_0)P_{Q_0}(Q)^n & \text{if} \quad 0<y<nT, q=nQ \\ \sum_{m=1}^{n} C(n,m)P_{Q_0}(Q)^{n-m}P'^{(m)}_{Q_0}(q-(n-m)Q)F_{T_e}^{(n-m)}(y-mT,T,Q,\lambda_0) & \text{if} \quad 0<y\le nT, n\le q<nQ \\ 1 & \text{if} \quad y\ge nT, n\le q\le nQ \end{cases}$$

The complex structure of $P_{Q_0}(q)$ and $F_{Y(n),Q_0^{(n)}}(y,q)$ given in Corollary (3.2.1) and Equation (6.4) illustrates that under $(Q,\mathbf{S},T)$ policy it is not possible to obtain closed form expressions for π_i's and $\varphi(t,k)$. Therefore, these quantities can be calculated only numerically. We next point out some remarks regarding the numerical computation of $f_{Y(n),Q_0^{(n)}}(y,q)$

Recall that we defined a function $P'_{Q_0}(q)$ which stands for the probability mass function of Q_0 where $Q_0<Q$: $P'_{Q_0}(q) = p_0(q,\lambda_0 T)/(1-\phi_0)$ for $1\le q<Q$. The truncated structure does not make it possible to obtain a closed form expression

for $P'^{(n)}_{Q_0}(q)$. Therefore, $P'^{(n)}_{Q_0}(q)$ can be calculated through an iterative procedure. as follows:

$$P'^{(n)}_{Q_0}(q) = \sum_{k=1}^{Q-1} P'^{(n-1)}_{Q_0}(q-k) P_{Q_0}(k) \quad \text{for} \quad n \le q \le n(Q-1)$$

As seen in Corollary 6.2.2, we need the convolution of truncated Erlang random variables. Although Erlang distribution is closed under convolution, the same does not apply for truncated Erlang distribution. Hence, it may become quite difficult to find the distribution of the convolution of a truncated Erlang random variable as demonstrated below with the 2 and 3-fold convolutions, the proof of which can be found in Appendix.

$$F^{(2)}_{T_e}(t,T,Q,\lambda_0) = \begin{cases} F(t,2Q,\lambda_0)/F(T,Q,\lambda_0)^2 & \text{if } 0 \le t < T \\ F(t-T,Q,\lambda_0)/F(T,Q,\lambda_0) + & \text{if } T \le t < 2T \\ \left[\int_{y=t-T}^{T} F(t-y,Q,\lambda_0) f(y,Q,\lambda_0) dy\right] / F(T,Q,\lambda_0)^2 & \\ 1 & \text{if } t \ge 2T \end{cases} \tag{6.5}$$

and

$$F^{(3)}_{T_e}(t,T,Q,\lambda_0) = \begin{cases} F(t,3Q,\lambda_0)/F(T,Q,\lambda_0)^3 & \text{if } 0 \le t < T \\ \left[\int_{y=0}^{T} F^{(2)}_{T_e}(t-y,T,Q,\lambda_0) f(y,Q,\lambda_0 dy)\right] / F(T,Q,\lambda_0) & \\ & \text{if } T \le t < 2T \\ F(t-2T,Q,\lambda_0)/F(T,Q,\lambda_0) + & \text{if } 2T \le t < 3T \\ \left[\int_{y=t-2T}^{T} F^{(2)}_{T_e}(t-y,T,Q,\lambda_0) f(y,Q,\lambda_0) dy\right] / F(T,Q,\lambda_0) & \\ 1 & \text{if } t \ge 3T \end{cases} \tag{6.6}$$

From Equations (6.5) and (6.6), we see that the convolution of a truncated Erlang random variable has a piecewise structure and requires an iterative procedure since it includes the convolution with a smaller degree. These equations indicate that it may be quite difficult to find the the distribution of the n^{th} convolution of a truncated Erlang random variable, even for small values of n especially from computational point of view. Therefore, it is necessary to use an approximate distribution to avoid the computational burden.

One option to approximate the distribution of a convolution is to use Normal distribution. Therefore, as an alternative for $F_{T_e}^{(n)}(t, T, Q, \lambda_0)$, one can use $\phi(t, n\mu_{T_e}, n\sigma_{T_e}^2)$ where $\phi(t, \mu, \sigma^2)$ represents the distribution of a Normal random variable with mean μ and variance σ^2. Here, μ_{T_e} and $\sigma_{T_e}^2$ corresponds to the mean and variance of the truncated Erlang random variable of interest and is given by the following:

$$\mu_{T_e} = \frac{QF(T, Q+1, \lambda_0)}{\lambda_0 F(T, Q, \lambda_0)}, \quad \sigma_{T_e}^2 = \frac{Q(Q+1)F(T, Q+2, \lambda_0)}{\lambda_0^2 F(T, Q, \lambda_0)} - \frac{Q^2 F(T, Q+1, \lambda_0)^2}{\lambda_0^2 F(T, Q, \lambda_0)^2}$$

In Figure 6.5, we present some examples to illustrate the performance of the Normal approximation. Figure 6.5(a) presents the exact distribution function and the normal approximation for a 2-fold convolution of a truncated Erlang random variable with $T = 0.3, Q = 10, \lambda_0 = 20$. This corresponds to a case where the truncation value, T, has a quite important effect in determining $F_{T_e}()$, *ie.* $F(T, Q, \lambda_0) = 0.0839$. For this specific example, the maximum absolute difference between the exact distribution function and the normal approximation is 0.0563 with a corresponding percentage difference of 10.59%. As T increases, the deviation of the approximation from the exact distribution reduces as expected. For instance, for $T = 0.9, Q = 10, \lambda_0 = 20$ where $F(T, Q, \lambda_0) = 0.9846$ (illustrated in Figure 6.5(b), the maximum absolute difference is 0.0199 with a corresponding percentage difference of 4.44%.

Figures 6.5(c) and 6.5(d) present the exact and approximate $F_{T_e}^{(n)}(t, T, Q, \lambda_0)$ with $T = 0.3, Q = 10, \lambda_0 = 20$ for $n = 3$ and $n = 4$, respectively. The maximum absolute difference for $n = 3$ is 0.0456 with a percentage difference of 8.35%. The corresponding figures with $n = 4$ are 0.0360 and 7.43%. As n increases the performance of the normal distribution to approximate $F_{T_e}^{(n)}()$ increases as expected. Moreover, the effect of the mentioned differences on the operating characteristics of the system is quite negligible as will be presented in the subsequent parts in this chapter and in the next chapter.

In Figure 6.6, we present different realizations for the steady-state distribution

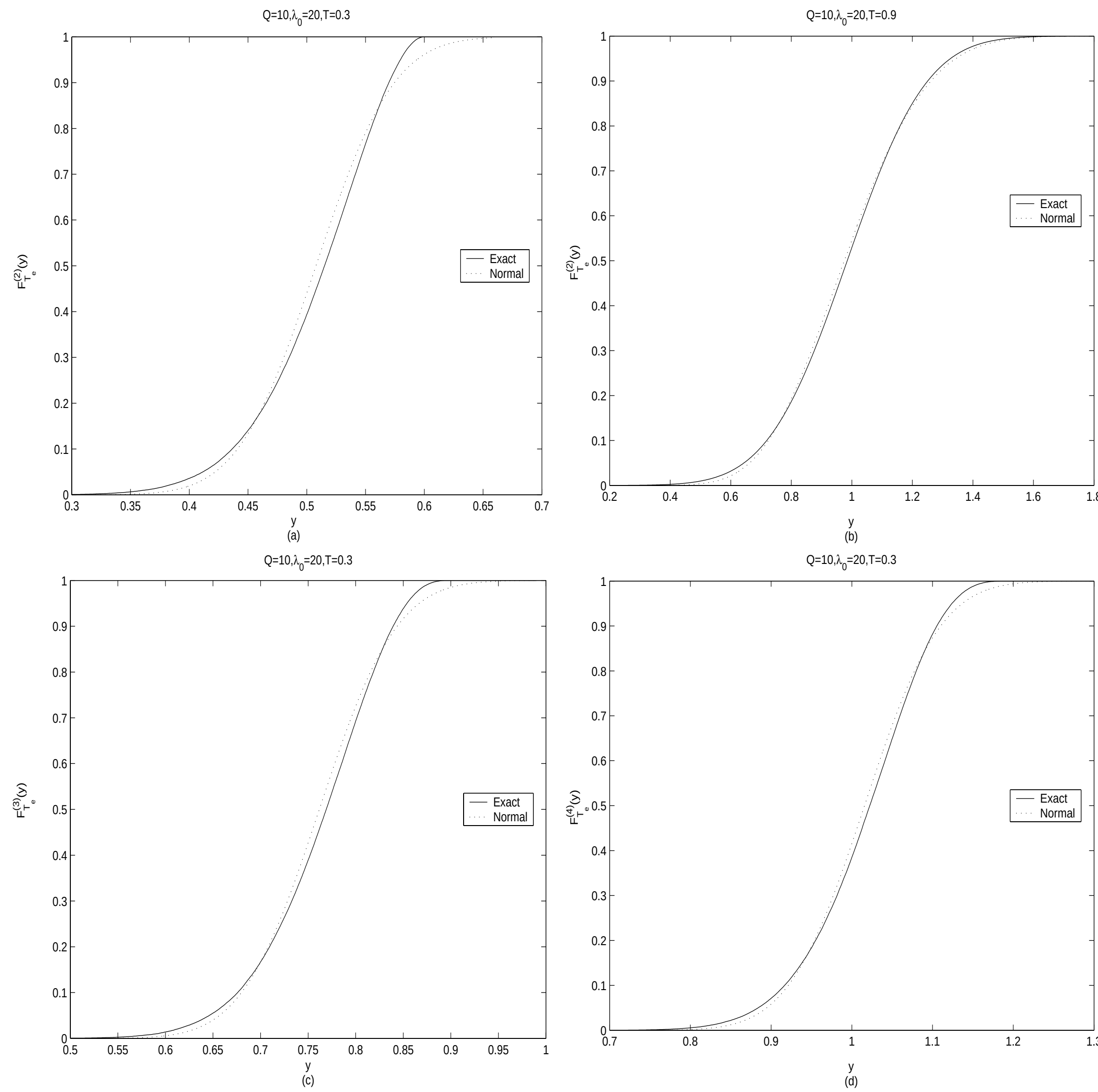

Figure 6.5: Comparison of a Normal approximation with the convolution of a Truncated Erlang random variable

of IP_0 under $(Q, \mathbf{S}, T)$ Policy. These realizations illustrate that the behaviour of π_i depends on Q and T values. If Q/λ_0 value is larger than T, IP_0 behaves uniformly in $[s_0+1, S_0-Q]$ and the behavior in $[S_0-Q+1, S_0]$ is more complex (See $T = 0.4$ in Figure 6.6(a)). Keeping Q constant, as T increases the value of $P_{Q_0}(Q)$ increases and therefore generally $\pi_{S_0}, \pi_{S_0-Q}, \ldots$ values increase and other π_i values decrease.This is to be expected because as $T \to \infty$, $(Q, \mathbf{S}, T)$ policy reduces to $(Q, \mathbf{S})$ policy and the steady-state inventory position of the warehouse becomes uniformly distributed in $[s_0+Q, ..., S_0-Q, S_0]$. Figure 6.6(b) presents π_i values with respect to different Q values keeping T constant. For larger Q values the distribution of IP_0 has a more smooth structure over $[s_0+1, s_0+2, \ldots, S_0-Q]$. We also observe that although there is not a general shape and behaviour of the distribution of IP_0, π_i's tend to behave in a similar manner in clusters of Q points. In Figure 6.6(b), see the behaviour of IP_0 for $Q = 4$ over $[47, 50]$, $[43, 46]$. Similarly, observe π_i's for $T = 1.2$ or $T = 0.8$ with $Q = 8$ in Figure 6.6(a).

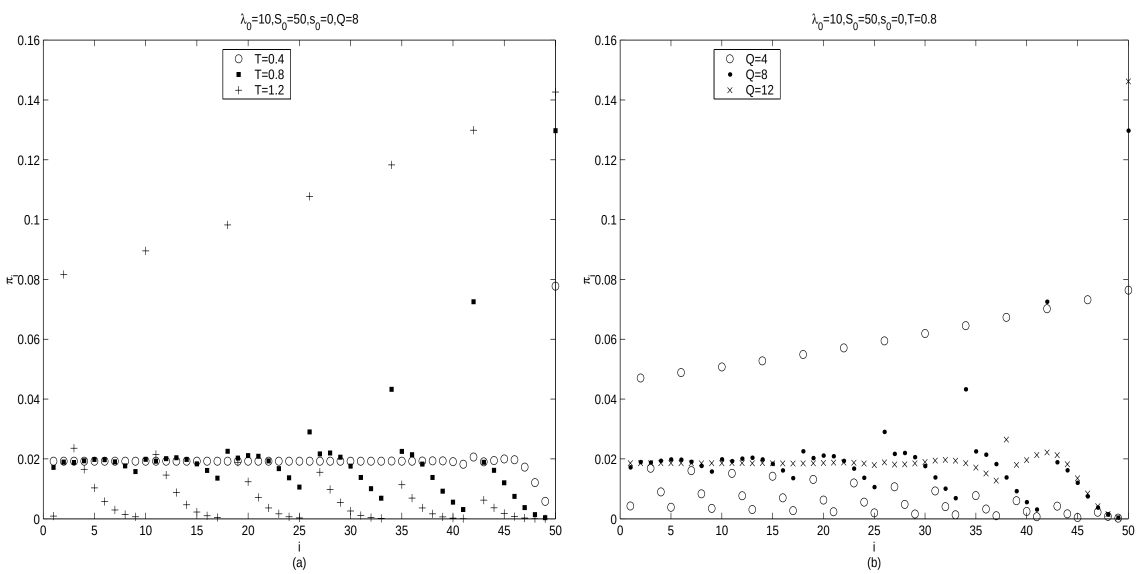

Figure 6.6: Realizations for Steady-State Distribution of IP_0 under $(Q, \mathbf{S}, T)$ Policy

Figure 6.7(a) presents waiting time distribution of an order with different sizes under the $(Q, \mathbf{S}, T)$ policy. Unlike the $(Q, \mathbf{S})$ policy, the waiting time of an order with a larger size is observed to be stochastically larger than the waiting time of an order with a smaller size. Under the $(Q, \mathbf{S}, T)$ policy, this is expected, because given an order of size $q < Q$, the inter-order time has the same distribution for all q values and a larger order has to wait more for the necessary stock to accumulate at the warehouse. An order of size Q has an inter-order time which is stochastically smaller than that of an order with size less than Q. The order of size Q, on the average, arrives earlier and waits more for the sufficient stock to exist on the shelf since it has a larger size. Hence, the order of size Q has a tendency to wait more at the warehouse. In Figure 6.7(b), we present $F_{W_0}(Q)(w)$ for $Q = 10, \lambda_0 = 20, L_0 = 5, S_0 = 100, s_0 = -30$ with different T values. An important observation is that the waiting time with smaller T values is stochastically smaller than that of waiting time with larger T values. Hence, introducing an effective time trigger to $(Q, \mathbf{S})$ policy reduces the waiting time distribution of an order at the warehouse and hence it decreases the effective lead time of an order placed by the retailers.

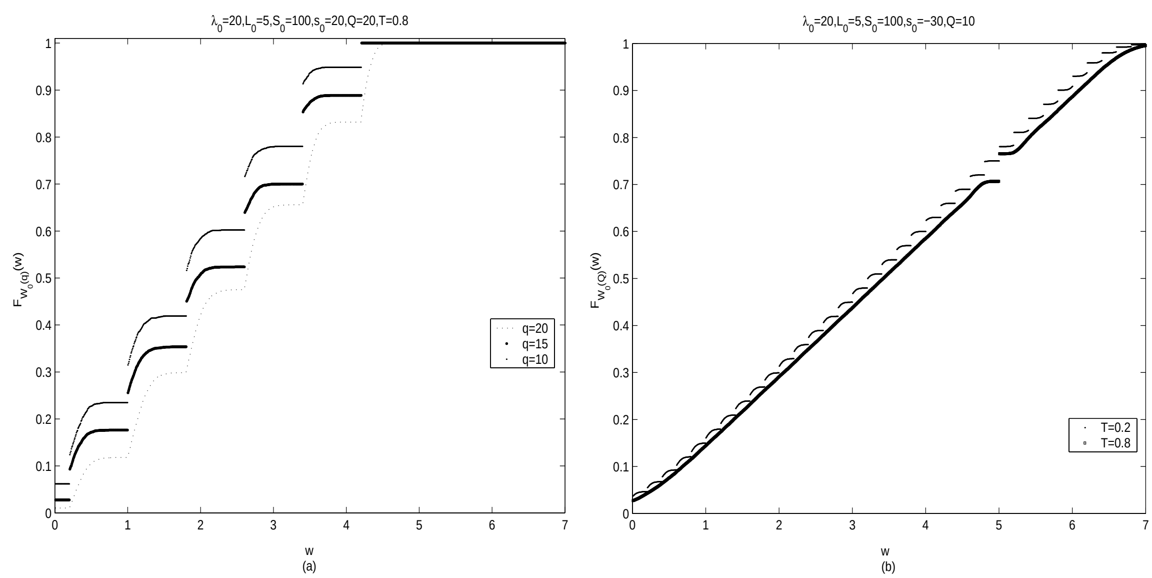

Figure 6.7: Realizations for $F_{W_{0(q)}}$ under $(Q, \mathbf{S}, T)$ Policy

For the calculation of the waiting time distributions depicted in Figure 6.7 we use Normal approximation for the truncated Erlang random variable as explained above. In order to examine the effect of the approximation on the waiting time distribution, we simulated the inventory system under $(Q, \mathbf{S}, T)$ policy to obtain the true waiting time distributions. In our simulations, we used a run length of 20,000 warehouse ordering instances with a warm-up period of 1,000 order placements, and 10 replications to obtain the corresponding waiting time distributions. For $q = 20$, the simulated expected waiting time is found to be 2.4126 whereas the approximate expected value is calculated as 2.4098 (0.12% percentage deviation from the true value). For $q = 15$ and $q = 10$, the simulated $E[W_0(q)]$ values are 2.1062 and 1.8343, respectively. The corresponding approximate values are 2.1104 and 1.8311 with 0.19% and 0.17% percentage deviation, respectively.

We last present the retailer order quantity distribution. The $(Q, \mathbf{S}, T)$ policy imposes a limit both on the time between ordering opportunities as T and the order quantities as Q. Unlike the $(Q, \mathbf{S})$ policy, the order quantity for the $(Q, \mathbf{S}, T)$ policy is a random variable and hence the conditional distribution of the retailer order quantity in an order of size q can be written as:

$$P_{R_i(q)}(m_i) = \frac{\begin{pmatrix} q \\ m_i \end{pmatrix} r_i^{m_i}(1-r_i)^{q-m_i}}{P_{Q_0}(q)} \quad \text{if} \quad 0 \leq m_i \leq q, 1 \leq q \leq Q$$

6.3 $(Q, \mathbf{S}|T)$ Policy

As indicated in the previous chapter, the minimum quantity periodic review $(Q, \mathbf{S}|T)$ policy was originally studied by Cachon [17] in a multi-item and single-location inventory system within the context of shelfspace and truck capacities.

Unlike the other policies, $(Q, \mathbf{S}|T)$ policy is a periodic review policy and hence the ordering opportunities arise at the end of each period, *ie.* every T time units. Figure 6.8 illustrates the ordering mechanism under $(Q, \mathbf{S}|T)$ policy. An ordering opportunity that arises at the end of each period is used if at least Q demands have accumulated for the retailers since the last ordering instance, ie. the total inventory position of the retailers, $IP(t)$ is below $S_T - Q$ with $S_T = \sum_{i=1}^{N} S_i$.

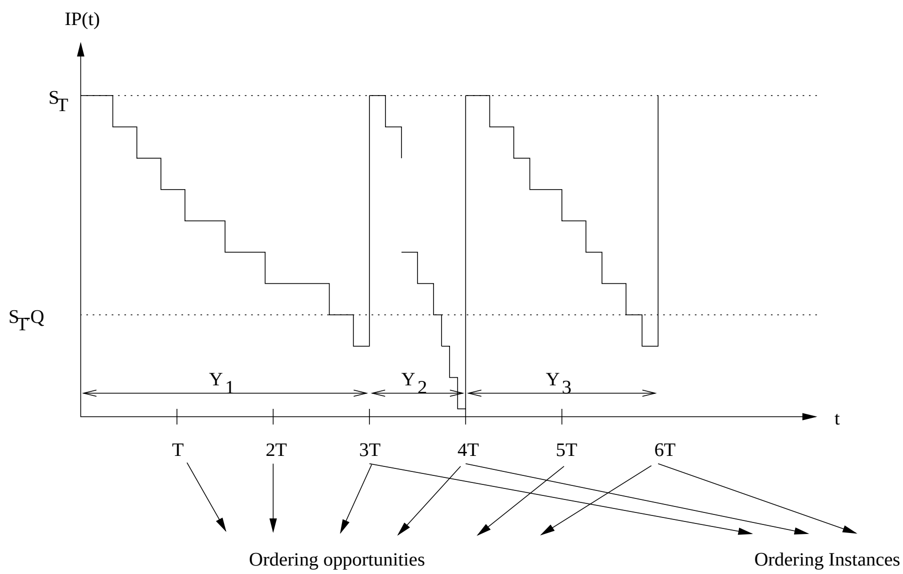

Figure 6.8: Illustration of Ordering Instances for $(Q, \mathbf{S}|T)$ Policy

Otherwise, only a review is carried out at an ordering opportunity. Therefore, the inter-order time and order sizes take values from the sets $\{T, 2T, 3T, \ldots\}$ and $\{Q, Q+1, Q+2, \ldots\}$, respectively. We next present the probability mass function of Y and Q_0.

Lemma 6.3.1 *The joint probability mass function of Y and Q_0 is given by*

$$f_{Y,Q_0}(y,q) = \begin{cases} p_0(q, \lambda_0 T) & \text{if} \quad q \geq Q, y = T \\ \\ p_0(q, \lambda_0 mT)B(Q-1, q, 1-1/m) & \text{if} \quad q \geq Q, y = mT, m > 1 \end{cases}$$

where $B(k,n,p)$ is the Binomial cumulative distribution function with parameters n and p.

Proof: *See Appendix.*

Using the above lemma, we can find the marginals as given below:

Corollary 6.3.1

(a) The probability mass function $P_{Q_0}(q) = P(Q_0 = q)$ of Q_0 is given by:

$$P_{Q_0}(q) = p_0(q, \lambda_0 T) + \sum_{m=2}^{\infty} p_0(q, \lambda_0 mT)B(Q-1, q, 1-1/m) \quad \text{if} \quad q \geq Q$$

(b) The p.m.f., $f_Y(y)$, of Y is given by:

$$f_Y(y) = \begin{cases} \overline{P}_0(Q-1, \lambda_0 T) & \text{if} \quad y = T \\ \\ P_0(Q-1, \lambda_0(m-1)T) - P_0(Q-1, \lambda_0 mT) & \text{if} \quad y = mT, m > 1 \end{cases}$$

The complex structure of $f_{Y,Q_0}(y,q)$ as well as $P_{Q_0}(q)$ and $f_Y(y)$ does not make it possible to obtain a closed form expression for the convolutions. Therefore, the convolutions under $(Q, \mathbf{S}|T)$ policy can be calculated only numerically. The discrete structure of both Y and Q_0 allows us to calculate $f_{Y^{(n)},Q_0^{(n)}}(y,q)$ iteratively as follows:

$$f_{Y^{(n)},Q_0^{(n)}}(mT,q) = \sum_{\left\{ \substack{m_1+m_2=m,\, m_1 \geq n-1,\, m_2 \geq 1 \\ q_1+q_2=q,\, q_1 \geq (n-1)Q,\, q_2 \geq Q} \right\}} f_{Y^{(n-1)},Q_0^{(n-1)}}(m_1 T, q_1) f_{Y,Q_0}(m_2 T, q_2)$$

$$\text{for} \quad m \geq n, q \geq nQ \tag{6.7}$$

and

$$F_{Y^{(n)},Q_0^{(n)}}(y,q) = \begin{cases} 0 & \text{if} \quad y < nT, q \geq nQ \\ \sum_{j=n}^{\lfloor y/T \rfloor} f_{Y^{(n)},Q_0^{(n)}}(jT,q) & \text{if} \quad y \geq nT, q \geq nQ \end{cases} \tag{6.8}$$

Corollary 6.3.2 *Under $(Q, \mathbf{S}|T)$ policy, $\varphi(t,k) = P(D_0(0,t] = k)$ is given by:*

$$\varphi(t,k) = \begin{cases} 1 & \text{if} \quad k = 0, t < T \\ \\ P_0(Q-1, \lfloor t/T \rfloor \lambda_0 T) & \text{if} \quad k = 0, t \geq T \\ \\ \sum_{n=1}^{\lfloor k/Q \rfloor} \left[\sum_{j=n}^{\lfloor t/T \rfloor} f_{Y^{(n)},Q_0^{(n)}}(jT,k) - \sum_{m=1}^{\lfloor (t-nT)/T \rfloor} \sum_{j=n}^{\lfloor (t-mT)/T \rfloor} f_{Y^{(n)},Q_0^{(n)}}(jT,k) f_y(mT) \right] \\ & \text{if} \quad k \geq Q, t \geq T \end{cases}$$

Proof: *See Appendix.*

Similar to the $(Q, \mathbf{S}, T)$ policy, $P_{Q_0}(q)$ under $(Q, \mathbf{S}|T)$ policy does not provide a closed form expression for π_i's. Next we present sample realizations for the steady-state distribution of the warehouse inventory position under $(Q, \mathbf{S}|T)$ policy in

Figure 6.9. We observe that the distribution of the warehouse inventory position is not very sensitive to the choice of Q and T values. Only when T is small compared to Q/λ_0 or Q is large compared to $\lambda_0 T$, the behaviour of π_i is somehow observed to be oscillatory. Noting that we use the same policy parameter values, (Q, T, s_0, S_0) as in Figure 6.6, we observe that the distribution of IP_0 has a more smooth structure when compared with the $(Q, \mathbf{S}, T)$ policy.

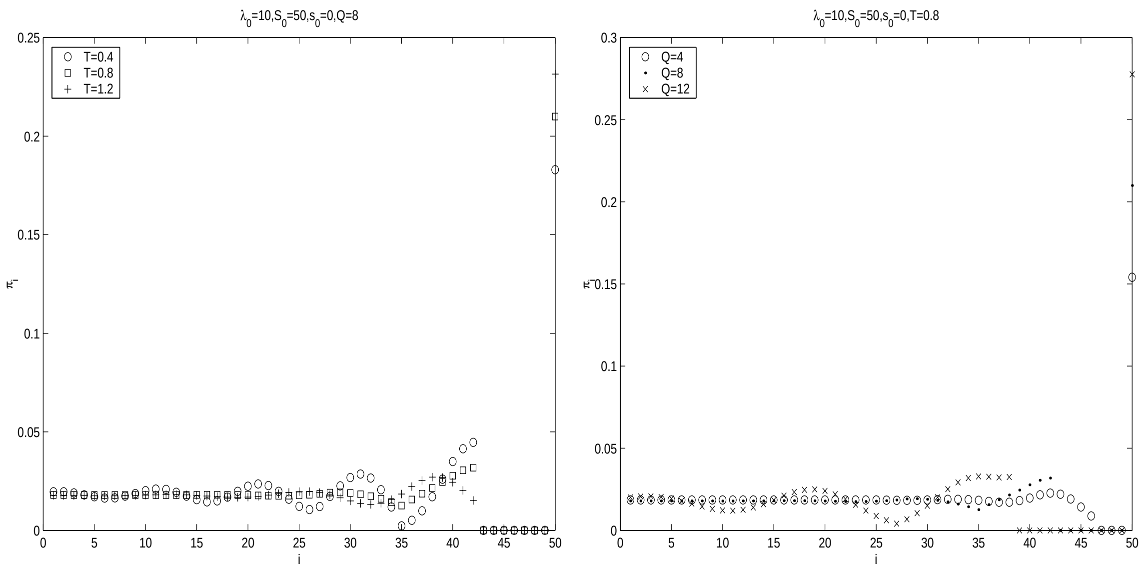

Figure 6.9: Realizations for Steady-State Distribution of IP_0 under $(Q, \mathbf{S}|T)$ Policy

Since $(Q, \mathbf{S}|T)$ is a periodic review policy, the orders are placed at the end of the periods, *ie.* the warehouse faces demands at the end of the periods. Therefore warehouse orders are also placed at the end of the periods and the warehouse employs a periodic review (s_0, S_0) policy with review interval T. Unlike the other policies, waiting time of an order at the warehouse has always a discrete structure. $W_0(q)$ takes values from the set $\{\ldots, L_0 + 2T, L_0 + T, L_0, L_0 - T, L_0 - 2T, \ldots, 0\}$.

In Figure 6.10, we present different realizations regarding the distribution of the waiting time of an order at the warehouse. Figure 6.10(a) illustrates $F_{W_0(q)}(w)$ for different values of q. We observe that there is not a particular pattern for the behaviour of the waiting time distribution with different order sizes. Figure

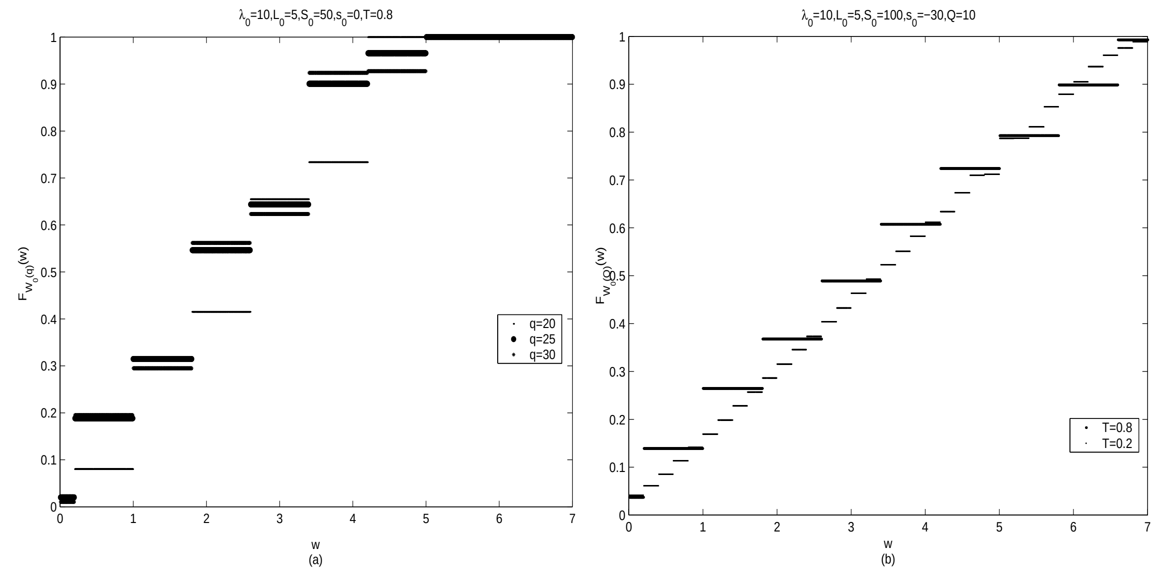

Figure 6.10: Realizations for $F_{W_0(q)}$ under $(Q, \mathbf{S}, |T)$ Policy

6.10(b) illustrates the waiting time of an order of a fixed size for different values of T. We observe that when $w \leq L_0$, the waiting time with $T = 0.2$ is stochastically larger than that of $T = 0.8$ whereas we may observe the opposite for $w > L_0$. Hence, when $s_0 < 0$, a smaller review time may be beneficial to obtain smaller effective lead times.

Similar to $(Q, \mathbf{S})$ and $(Q, \mathbf{S}, T)$ policies, since the demand is Poisson the conditional distribution of the retailer order quantity in an order of size q can be written as:

$$P_{R_i(q)}(m_i) = \frac{\begin{pmatrix} q \\ m_i \end{pmatrix} r_i^{m_i}(1-r_i)^{q-m_i}}{P_{Q_0}(q)} \quad \text{if} \quad 0 \leq m_i \leq q, q \geq Q$$

6.4 $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ Policy

The $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is a special case of $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ can-order policy with $\mathbf{c} = \mathbf{S}-\mathbf{1}$ which were previously studied by Silver [67] in a 2-item inventory system with zero lead time and Van Eijs [77] in a general multi-item setting. For the two-echelon inventory system, it is formally stated as follows:

$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ **Policy:** *Whenever the inventory position of a retailer i drops to its reorder level, s_i, raise the inventory positions of the retailers up to $\mathbf{S} = (S_1, S_2, \ldots, S_N)$.*

In $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy, the ordering opportunities for the retailers arise whenever the inventory position of a retailer drops to its reorder level. Observe that this is the only policy among the three that bases its ordering decision on the individual inventory positions of the retailers. Recall that under three policies explained above, the ordering opportunities arrive either on the basis of the total demands accumulated in the system and/or the time elapsed. Similar to $(Q, \mathbf{S})$ policy, every ordering opportunity is used and all retailers are replenished to their order-up-to levels at each ordering opportunity.

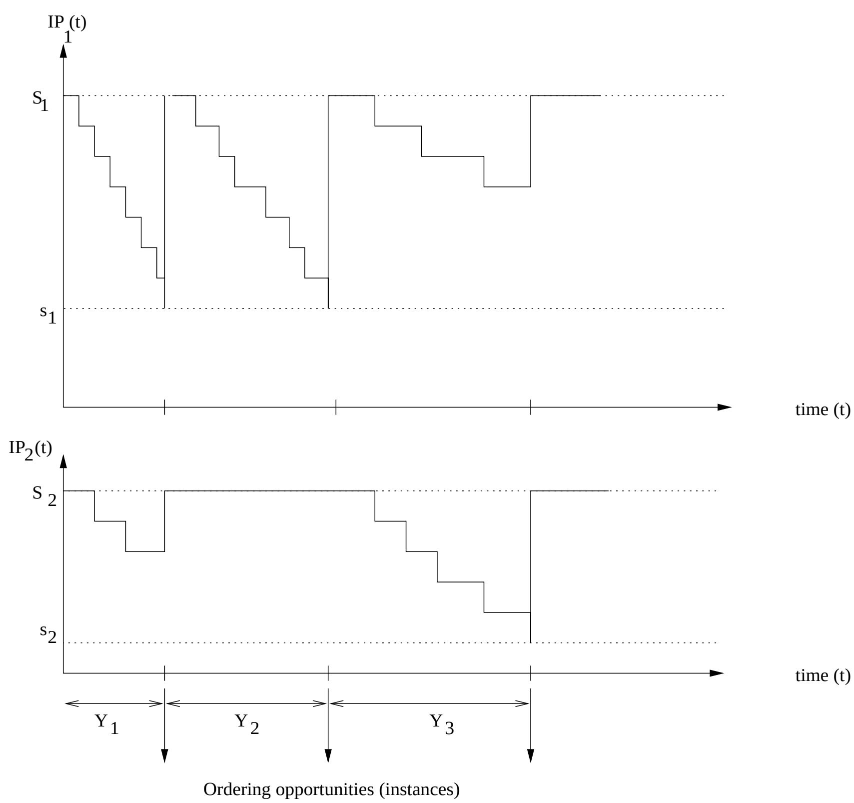

Figure 6.11: Illustration of Ordering Instances for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ Policy

We illustrate a sample realization of the ordering behaviour of the retailers under $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy in Figure 6.11. Suppose that there are $N = 2$ retailers and the inventory position of each retailer i is at its maximum level, S_i at time $t = 0$.The inventory position of retailer 1 drops to its reorder level before the second retailer and hence an ordering opportunity also arises for the retailer 2. Since the inventory position of retailer 2 is below S_2, both retailers are replenished to $\mathbf{S}$ values. Similarly, second ordering opportunity is also triggered by retailer 1. However, since no demands have arrived for retailer 2 since the first order, the opportunity is not used by retailer 2 and only retailer 1 is replenished. Suppose that the third order opportunity is triggered by retailer 2, *ie.* the inventory position of retailer 2 drops to s_2 before. Since the inventory position of retailer 1 is below S_1, retailer 1 is also included in the third replenishment.

Now, let $\Delta_i = S_i - s_i$ be the maximum quantity that retailer i can order. Then, the minimum size of an order is $\underline{Q}_0 = \min_i \Delta_i$ which corresponds to the case where only the retailer with the smallest Δ_i value is included in the order, *ie.* the inventory positions of the other retailers are at their order-up-to levels at the opportunity arrival. The maximum size of an order is $\overline{Q}_0 = \sum_i \Delta_i - (N-1)$ which occurs when the inventory position of each retailer $i, i = 1, 2, \ldots, N$ is $s_i + 1$ and a demand arrives which generates an ordering instance. We next present the probability mass function of Y and Q_0.

Lemma 6.4.1 *The joint probability density function of Y and Q_0 is given by*

$$f_{Y,Q_0}(y,q) = f(y,q,\lambda_0) \sum_{i=1}^{N} \sum_{\left\{ \substack{\sum_{j=1}^{N} x_j = q,\, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \ \text{for} \ j \neq i} \right\}} (q-1)! \frac{r_i^{x_i}}{(x_i - 1)!} \left[\prod_{j \neq i} \frac{r_j^{x_j}}{x_j!} \right] I(q \ge \Delta_i)$$

$$\text{if} \quad y \ge 0, \underline{Q}_0 \le q \le \overline{Q}_0$$

where $I()$ is the indicator function of its argument.

Proof: *See Appendix.*

The above lemma is used to find the marginals as below:

Corollary 6.4.1

(a) The probability mass function $P_{Q_0}(q) = P(Q_0 = q)$ of Q_0 is given by:

$$P_{Q_0}(q) = \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{l} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \text{ for } j \neq i \end{array} \right\}} (q-1)! \frac{r_i^{x_i}}{(x_i - 1)!} \left[\prod_{j \neq i} \frac{r_j^{x_j}}{x_j!} \right] I(q \ge \Delta_i) \quad \text{if} \quad \underline{Q}_0 \le q \le \overline{Q}_0$$

(b) The p.d.f., $f_Y(y)$, of Y is given by:

$$f_Y(y) = \sum_{q=\underline{Q}_0}^{\overline{Q}_0} f(y, q, \lambda_0) P_{Q_0}(q)$$

Proof: *See Appendix.*

Using (a) of Corollary 6.4.1, we see that $f_{Y,Q_0}(y,q)$ can also be written as:

$$f_{Y,Q_0}(y,q) = f(y,q,\lambda_0) P_{Q_0}(q) \quad \text{if} \quad y \ge 0, \underline{Q}_0 \le q \le \overline{Q}_0$$

This leads to an important result which enables the analysis of this policy in a more compact way: Given an order of size q, the inter-order time has an $Erlang_q$ distribution.

$$f_{Y|Q_0}(y|q) = f(y,q,\lambda_0) \quad \text{if} \quad y \ge 0, \underline{Q}_0 \le q \le \overline{Q}_0 \tag{6.9}$$

Using Equation (6.9), we can write

$$f_{Y^{(n)},Q_0^{(n)}}(y,q) = f(y,q,\lambda_0) P_{Q_0}^{(n)}(q) \quad \text{if} \quad y \ge 0, n\underline{Q}_0 \le q \le n\overline{Q}_0$$

and

$$F_{Y^{(n)},Q_0^{(n)}}(y,q) = F(y,q,\lambda_0) P_{Q_0}^{(n)}(q) \quad \text{if} \quad y \ge 0, n\underline{Q}_0 \le q \le n\overline{Q}_0 \tag{6.10}$$

Then, we are ready to give an expression for the distribution of $D_0(0,t]$.

Corollary 6.4.2 *Under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policy, $\varphi(t,k) = P(D_0(0,t] = k)$ is given by:*

$$\varphi(t,k) = \begin{cases} \sum_{q=\underline{Q}_0}^{\overline{Q}_0} \overline{F}(t,q,\lambda_0) P_{Q_0}(q) & \text{if} \quad k = 0 \\ \sum_{n=\lfloor k/\overline{Q}_0 \rfloor}^{\lfloor k/\underline{Q}_0 \rfloor} P_{Q_0}^{(n)}(k) \left[F(y,k,\lambda_0) - \sum_{q=\underline{Q}_0}^{\overline{Q}_0} P_{Q_0}(q) F(t, k+q, \lambda_0) \right] & \text{if} \quad k \ge \underline{Q}_0 \end{cases}$$

Proof: *See Appendix.*

The crucial quantity in the analysis of the $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is the probability mass function of Q_0, which has a quite complex structure as given in Corollary (6.4.1). Only for $N = 2$, we can provide a closed form expression for $P_{Q_0}(q)$.

$$P_{Q_0}(q) = \sum_{i=1}^{2} \binom{q-1}{\Delta_i - 1} r_i^{\Delta_i}(1-r_i)^{q-\Delta_i} I(q \geq \Delta_i) \text{ for } \min_i \Delta_i \leq q \leq \sum_{i=1}^{2} \Delta_i - 1$$

For $N > 2$, when a retailer is triggered by a retailer i, if the order quantities of the other retailers were unrestricted, the order quantity of retailer j will have a Binomial distribution with parameters $q - \Delta_i$ and $\lambda_j/(\lambda_0 - \lambda_i)$. However, the order quantities of the retailers are no longer unrestricted and hence we have to sum over all possible values of these order quantities, which is computationally very difficult especially for larger values of N. Therefore, an approximation for $N > 2$ is necessary to carry out the numerical calculations.

Before we present the proposed approximation for $P_{Q_0}(q)$, we introduce some notation. Let $r'_{j,i}$ with $(i \neq j)$ be the probability that a demand arrives at retailer j given that it does not arrive at retailer j and is given by $\lambda_j/(\lambda_0 - \lambda_i)$. Let $\mu_{i,j}$ and $\sigma^2_{i,j}$ with $i \neq j$ be the expected value and variance of the order quantity of retailer j in an order triggered by retailer i. Also let $\mu_i = \sum_{j\neq i} \mu_{j,i}$ and $\sigma_i^2 = \sum_{j \neq i} \sigma^2_{j,i}$. and for any quantity X, we let $\widehat{X}$ denote the approximation for X.

Now, we propose the following $\widehat{P_{Q_0}(q)}$:

$$\widehat{P_{Q_0}(q)} = \sum_{i=1}^{N} C_{Q_0} \binom{q-1}{\Delta_i - 1} r_i^{\Delta_i}(1-r_i)^{q-\Delta_i} \left[\begin{array}{l} \phi(q - \Delta_i + 0.5, \widehat{\mu_i}, \widehat{\sigma_i^2}) \\ -\phi(q - \Delta_i - 0.5, \widehat{\mu_i}, \widehat{\sigma_i^2}) \end{array} \right] \quad (6.11)$$

where C_{Q_0} is the normalizing constant such that $\sum_{q=\underline{Q}_0}^{\overline{Q}_0} \widehat{P_{Q_0}(q)} = 1$ and $\widehat{\mu_{j,i}}$ and $\widehat{\sigma^2_{j,i}}$ are calculated by

$$\widehat{\mu_i} = \sum_{j \neq i} \widehat{\mu_{j,i}} = \sum_{j\neq i} \sum_{k=0}^{\Delta_j - 1} k b(k, q - \Delta_i, r'_{j,i}) \quad (6.12)$$

$$\widehat{\sigma_i^2} = \sum_{j \neq i} \widehat{\sigma^2_{j,i}} = \sum_{j\neq i} \sum_{k=0}^{\Delta_j - 1} k^2 b(k, q - \Delta_i, r'_{j,i}) - \sum_{j \neq i} \widehat{\mu_{j,i}}^2 \quad (6.13)$$

where $b(k, n, p)$ represents the probability mass function of a Binomial random variable with parameters n and p.

The idea behind Equation (6.11) is to approximate the probability that the sum of the order quantities of the retailers which do not trigger the order is $q-\Delta_i$ if retailer i triggers the order. As explained above, the retailer order quantities are restricted with $\Delta_j - 1$ where $j \neq i$ and hence, we first find the approximate mean and variance of the retailer order quantities in Equations (6.12) and (6.13), respectively. The probability mass function of the sum of the order quantities of the retailers which do not trigger the order is then approximated by a Normal distribution with continuity correction. We use the proposed approximate $\widehat{P_{Q_0}(q)}$ throughout our numerical experiments with $N > 2$ and as explained below the approximation works quite well even with small values of N.

Similar to $(Q, \mathbf{S}, T)$ and $(Q, \mathbf{S}|T)$ policies, the steady-state distribution of IP_0 under $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy can be computed only numerically. In Figure 6.12, we present different realizations of π_i's. Recall that the policy bases the ordering decisions on the individual inventory positions rather than the total inventory position and hence we will investigate the effect of number of retailers, individual demand rates and the retailer maximum order quantities on the behaviour of π_i's.

Figure 6.12(a) illustrates the behaviour of π_i for $N = 2$ and $N = 4$ keeping λ_i fixed. We observe that for smaller values of the inventory position value i, π_i is not sensitive to the number of retailers in the system. The behaviour of π_i's is also observed to be oscillatory. In Figure 6.12(b), we present π_i values for $N = 2$ and $N = 4$ keeping the system demand rate, λ_0 fixed. A similar behaviour as in Figure 6.12(a) is observed for π_i's.

For the calculation of π_i values for $N = 4$ given in Figures 6.12(a)-(b), we use the proposed normal approximation for $P_{Q_0}(q)$ given in Equation (6.11). To illustrate the effect of the approximation on the steady-state inventory position distribution, we simulated the inventory system under $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy to obtain the true π_i values. In our simulations, we used a run length of 20,000 warehouse ordering instances with a warm-up period of 1,000 order placements, and 10 replications. The maximum percentage deviation of the approximate π_i values from the true values is found to be 5.14% with an average of 1.02% deviation. As will be presented in the next chapter, the effect of the

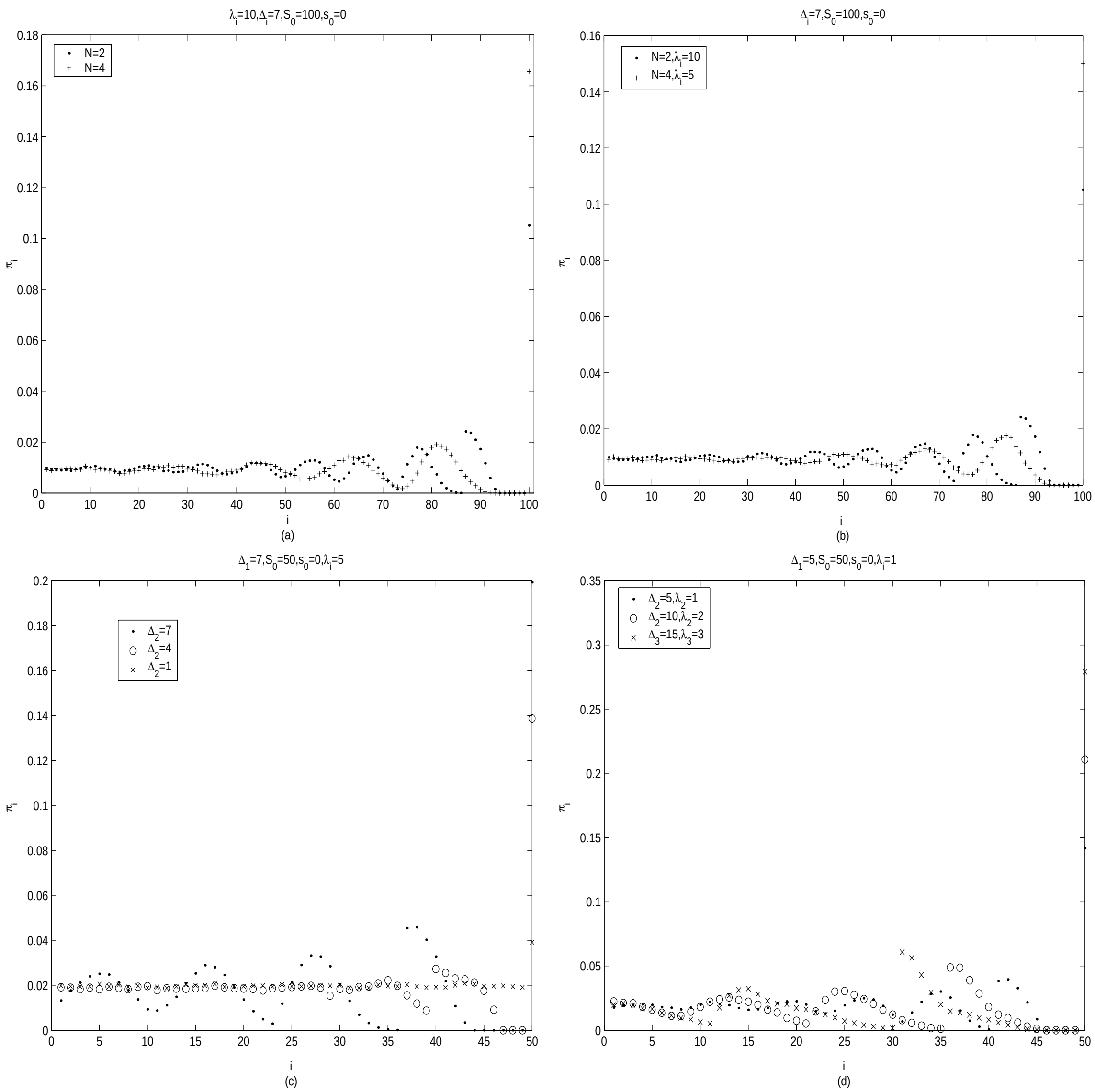

Figure 6.12: Realizations for Steady-State Distribution of IP_0 under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ Policy

approximation on the operating characteristics and cost rate function is negligible.

Figure 6.12(c)-(d) present realizations for the steady-state Distribution of IP_0 with different retailer parameters. Figure 6.12(c) demonstrates the behaviour of π_i for different Δ_2 values keeping Δ_1 and λ_i fixed. As Δ_2 decreases, the system starts to behave as if there is only retailer 2 in the system and π_i's have a more smooth structure, ie. for $\Delta_2 = 1$, the distribution of the inventory position is quite similar to a uniform distribution. Figure 6.12(d) gives different realizations of π_i for different Δ_2 and λ_2 values while keeping $\Delta_2/\lambda_2 = \Delta_1/\lambda_1$ fixed. We observe that there is not a particular pattern for the behaviour of π_i.

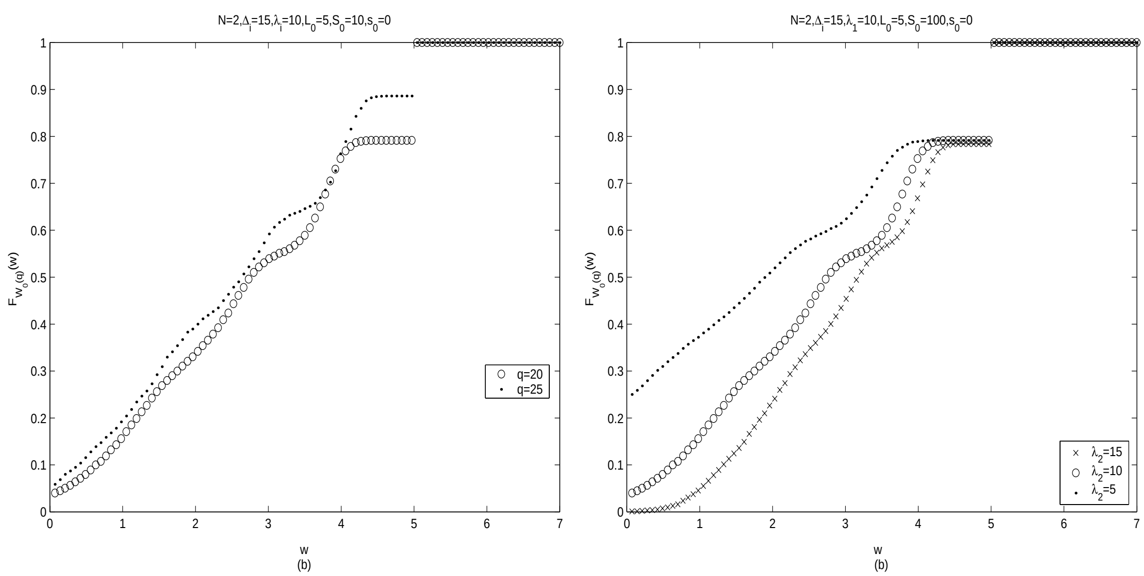

Figure 6.13: Realizations for $F_{W_0(q)}$ under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ Policy

Figure 6.13 presents realizations for the waiting time distribution under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policy. The inter-order time has a similar structure to that of $(Q, \mathbf{S})$ policy, *ie.* the inter-order time of an order of size q has an $Erlang_q$ distribution. Therefore, we observe that the waiting time distribution under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policy behaves similar to $(Q, \mathbf{S})$. In Figure 6.13(a), we present $F_{W_0(q)}(w)$ for two different values of q. It is observed that for higher values of q, $P(W_0(q) = 0)$ for $q = 25$ is higher than that of $q = 20$, which may be attributed to the stochastically larger inter-order time as in $(Q, \mathbf{S})$ policy. Figure 6.13(b) presents different realizations for the waiting time distribution when the demand rates are differentiated. For a larger demand rate, the waiting time of an order is stochastically larger.

Unlike the previous three policies, there are limitations on the retailer order quantities for an order of size q. Because, if retailer i triggers the order, then $R_i(q) = \Delta_i$ and $R_j(q) < \Delta_j$ for $j \neq i$. Then,

$$P_{R_i(q)}(m_i) = \begin{cases} \sum_{\left\{ \begin{array}{c} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \leq x_j < \Delta_j \text{ for } j \neq i \end{array} \right\}} \frac{(q-1)!}{x_1! \ldots (\Delta_i - 1)! \ldots x_N!} r_1^{x_1} \ldots r_i^{m_i} \ldots r_n^{x_N} / P_{Q_0}(q) & \\ \qquad\qquad \text{if} \quad m_i = \Delta_i, m_i \leq q & \\ \sum_{k \neq i} \sum_{\left\{ \begin{array}{c} \sum_{j=1}^{N} x_j = q, x_k = \Delta_k \\ x_i = m_i, 0 \leq x_j < \Delta_j \text{ for } j \neq k, \end{array} \right\}} \frac{(q-1)!}{x_1! \ldots m_i! \ldots x_N!} r_1^{x_1} \ldots r_i^{m_i} \ldots r_n^{x_N} / P_{Q_0}(q) & \\ \qquad\qquad \text{if} \quad 0 \leq m_i < \Delta_i, m_i \leq q & \end{cases}$$

Chapter 7

Numerical Results for Policies in Class $\mathcal{P}$

In the previous two chapters, we have studied different joint replenishment policies in a two-echelon, single item inventory system. In this chapter, our aim is to discuss the computational results regarding these policies. Although the policies can be analyzed under a single policy class, they each choose a particular generation mechanism for the opportunities/orders. Therefore, the overall costs incurred by the inventory system depend greatly on how these opportunities are generated and used in the system.

In Section 7.1, we present the search algorithm employed and some computational remarks for each policy. In Section 7.2 we discuss the advantage of employing the joint replenishment policies under the policy class $\mathcal{P}$ instead of installation stock policies at the retailers. In Section 7.3 and 7.4, we will present the performance of the policies for the systems where the warehouse acts as a cross-dock point and it is allowed to hold stock, respectively. We will also include a discussion on the allocation of cost between the echelons and the difference of echelon costs across the policies. Section 7.5 will discuss the benefit of allowing the warehouse to hold stock instead of employing cross-dock under the joint replenishment policies of interest.

7.1 Computational Issues

Before we proceed with the performance the policies, we will present the search algorithm and search space we use in order to find the optimal policy parameters and some computational remarks for each policy. Even under the simple $(Q, \mathbf{S})$ policy for which closed form expressions of operating characteristics are available, an analytical investigation for unimodality of $AC_{\mathcal{P}}$ provided in Equation (5.20) is not possible since the Lagrangian of fill rate constraints can not be proved to be convex in the policy parameters (See also Agrawal and Seshadri [1]). Therefore, to find the optimal policy parameters we employed either an iterative search algorithm with randomized initial points or an exhaustive search algorithm over a large solution space, whichever is convenient for each of the four policies considered. Notice that the search algorithms provided herein cover the general case where we allow the warehouse to hold stock, *ie.* optimization over (s_0, S_0) is also included. For the problems in which the warehouse employs cross-dock, the steps for the optimization of (s_0, S_0) values are excluded from the algorithm.

Before going on with the details for each policy, we first introduce some notation that will be frequently used: μ_t is the expected number of units demanded from the warehouse in $(0, t]$, ie. $\mu_t = E[D_0(0, t]] = \sum_{k=0}^{\infty} k\varphi(t, k)$. $E[W_0]$ is the expected waiting time of an order placed at the warehouse and is calculated by $E[W_0] = \sum_{q=1}^{\infty} E[W_0(q)]P_{Q_0}(q)$. We let $\lfloor x \rfloor_k$ denote the smallest integer larger than or equal to x which is divisible by k. We also define $Q_r = \left\lceil \sqrt{2K\lambda_0/(\sum_{i=1}^{N} r_i h_i)} \right\rceil$ and $Q_w = \left\lceil \sqrt{2K_0\lambda_0/h_0} \right\rceil$. As also explained in Section 4.1, these values correspond to the optimal order quantities of the retailers and the warehouse under EOQ model with corresponding ordering, holding costs and demand rates and provide a basis to determine the search space for the optimal policy parameters (see Pantumsinchai [58]). We also note that the search space we consider for each of the algorithm represents a very broad range of the policy parameters.

7.1.1 $(Q, \mathbf{S})$ Policy

The $(Q, \mathbf{S})$ policy is the simplest of four policies considered and requires only $N+3$ policy parameters, $(Q, S_1, S_2, \ldots, S_N, s_0, S_0)$ to be optimized. Moreover, as indicated in Section 6.1 and in Cheung and Lee [23], since the size of the order placed by the retailers is constant, s_0 and S_0 values should be integer multiples of Q, which narrows down the search space considerably. This restriction makes the exhaustive search algorithm over a large solution space possible. We also observe that for a given (Q, s_0, S_0) triplet, it is easy to see that the optimization problem to find $\mathbf{S}^*$ can be decomposed into N independent sub-problems in each of which we solve for S_i^* separately where S_i^* is the minimum value of S_i that satisfies the required fill rate constraint as also presented in steps 2.1.2.2.1.2.1 and 2.1.2.2.1.3.1 of the algorithm given below.

The search space consists of $Q \in [Q^{min}, Q^{max}], s_0 \in [s_0^{min}, s_0^{max}], S_0 \in [S_0^{min}, S_0^{max}], S_i \in [S_i^{min}, S_i^{max}]$ $i = 1, 2, \ldots, N$ with increments of $\Delta_Q = 1, \Delta_{s_0} = Q, \Delta_{S_0} = Q, \Delta_{S_i} = 1$. The boundaries of Q are given by

$$Q^{min} = \max(1, Q_r) \quad Q^{max} = \max(2Q_r, Q_r + 50)$$

The limits for (s_0, S_0) will be determined based on the value of the other parameters in the algorithm. The employed search algorithm is provided as below:

Search Algorithm for $(Q, \mathbf{S})$ Policy:

1.1. Set Q_m, Q^{min}, Q^{max}

2.1. For each $Q \in [Q^{min}, Q^{max}]$ by Δ_Q

  2.1.1. Compute $s_0^{min} = \lfloor Q_w \rfloor_Q - 5 \lfloor \mu_{L_0} \rfloor_Q$ and $s_0^{max} = \lfloor Q_w \rfloor_Q + 5 \lfloor \mu_{L_0} \rfloor_Q$.

  2.1.2. For each $s_0 \in [s_0^{min}, s_0^{max}]$ by Δ_{s_0}

    2.1.2.1. Compute $S_0^{min} = s_0 + Q$ and $S_0^{max} = \lfloor Q_w \rfloor_Q + 10 \lfloor \mu_{L_0} \rfloor_Q$.

    2.1.2.2 For each $S_0 \in [S_0^{min}, S_0^{max}]$ by Δ_{S_0}.

      2.1.2.2.1. For each retailer $i \in \{1, 2, \ldots, N\}$

        2.1.2.2.1.1. Set $S_i^{in} = \lceil \lambda_i (L_i + E[W_0]) \rceil$.

        2.1.2.2.1.2. If $\gamma_i < \overline{\gamma_i}$

          2.1.2.2.1.2.1. Set $S_i^* = \min\{S_i : S_i > S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$.

          2.1.2.2.1.2.2. Go to step 2.1.2.2.2.

2.1.2.2.1.3. If $\gamma_i \geq \overline{\gamma_i}$

2.1.2.2.1.3.1. Set $S_i^* = \max\{S_i : S_i \leq S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$.

2.1.2.2.1.4.3. Go to step 2.1.2.2.2.

2.1.2.2.2. Compute and store $AC(Q, \mathbf{S}^*, s_0, S_0)$

3.1. Set $(Q^*, s_0^*, S_0^*) = \text{argmin} AC(Q, \mathbf{S}^*, s_0, S_0)$.

No approximations are used to calculate the cost rate function of $(Q, \mathbf{S})$ policy. Therefore, the best cost rate function $AC(Q^*, \mathbf{S}^*, s_0^*, S_0^*)$ found as a result of the algorithm is exact. We also note that over the 2560 experimental instances where we allow the warehouse to hold stock and whose details will be given in Section 7.3, the search algorithm presented above did not result in best policy parameters which are on the boundaries of the provided search space. Similarly, for 1920 experimental instances where the warehouse employs cross-dock (See the details in Section 7.3), the best Q values were never obtained as Q^{min} or Q^{max}.

7.1.2 $(Q, \mathbf{S}, T)$ Policy

In an N retailer inventory system where the warehouse is allowed to hold stock, the dimensionality of $(Q, \mathbf{S}, T)$ policy is $N + 4$ with the parameters $(Q, S_1, S_2, \ldots, S_N, T, s_0, S_0)$ to be optimized. Although $(Q, \mathbf{S}, T)$ policy requires only one more parameter than that of $(Q, \mathbf{S})$ policy, the optimization is not as easy since none of the operating characteristics such as the steady-state distribution of the warehouse inventory position and waiting time distribution have a closed form expression. They can be calculated only numerically as explained in Section 6.2. From computation time point of view, it becomes almost impossible to use an exhaustive search algorithm to find the optimal policy parameters. Therefore, we used an iterative search algorithm starting with random initial points. A total of 50 initial points $\hat{Q}, \hat{T}, \hat{s}_0$ and $\hat{S}_0$ were selected sequentially over the ranges given below. These ranges are also used in the optimization steps of the iterative algorithm. Below, we also give $\Delta_Q, \Delta_T, \Delta_{s_0}$ and Δ_{S_0} values which represent the increments of each policy parameter within the given ranges.

$$\hat{Q} \in [Q^{min}, Q^{max}], T \in [T^{min}, T^{max}], s_0 \in [s_0^{min}, s_0^{max}], S_0 \in [S_0^{min}, S_0^{max}]$$

$$Q^{min} = \max(1, Q_r), \quad Q^{max} = \max(3Q_r, Q_r + 100), \quad \Delta_Q = 1$$
$$T^{min} = 0.5Q^{min}/\lambda_0, \quad T^{max} = 2Q^{max}/\lambda_0 \quad \Delta_T = (T^{max} - T^{min})/30$$
$$s_0^{min} = Q_w - 5\lfloor \mu_{L_0} \rfloor, \quad s_0^{max} = Q_w + 5\lfloor \mu_{L_0} \rfloor, \quad \Delta_{s_0} = 1$$
$$S_0^{min} = s_0 + 1, S_0^{max} = Q_w + 10\lfloor \mu_{L_0} \rfloor, \quad \Delta_{S_0} = 1$$

The iterative algorithm, as will be presented below, starts with a randomly selected quartet $(\hat{Q}, \hat{T}, \hat{s}_0, \hat{S}_0)$ and ends either when the same policy parameters are obtained in two consecutive iterations or the number of iterations reaches 3000. One iteration of our iterative search algorithm consists of five consecutive optimization problems for one of the policy parameters while keeping the other four parameters constant. We next give the details of the employed iterative search algorithm for one initial random point. In the algorithm, n_{it} corresponds to the iteration number whereas $Q^p, T^P, \mathbf{S}^p, s_0^p, S_0^p$ represent the corresponding parameter values in the previous iteration.

Search Algorithm for $(Q, \mathbf{S}, T)$ Policy:

1.1. Set $n_{it} = 0$.

1.2. Select $\hat{Q}, \hat{T}, \hat{s}_0, \hat{S}_0$.

1.3. Set $\hat{\mathbf{S}} = \mathbf{0}, \mathbf{S}^p = \mathbf{1}, Q^p = \hat{Q} + 1, T^p = \hat{T} + 1, s_0^p = \hat{s}_0 + 1, S_0^p = \hat{S}_0 + 1$.

2.1. If $(n_{it} \leq 3000)$ and $(\hat{Q} \neq Q^p$ or $\hat{T} \neq T^p$ or $\hat{\mathbf{S}} \neq \mathbf{S}^p$ or $\hat{s}_0 \neq s_0^p$ or $\hat{S}_0 \neq S_0^p)$

  2.1.1. Go to step 2.3.

2.2. If $(n_{it} > 3000)$ or $(\hat{Q} = Q^p$ and $\hat{T} = T^p$ and $\hat{\mathbf{S}} = \mathbf{S}^p$ and $\hat{s}_0 = s_0^p$ and $\hat{S}_0 = S_0^p)$

  2.2.1. Go to step 10.1

2.3. Set $Q^p = \hat{Q}, \mathbf{S}^p = \hat{\mathbf{S}}, T^p = \hat{T}, s_0^p = \hat{s}_0, S_0^p = \hat{S}_0$.

3.1. For each retailer $i \in \{1, 2, \ldots, N\}$

  3.1.1. Set $S_i^{in} = \lceil \lambda_i(L_i + E[W_0]) \rceil$.

  3.1.2. If $\gamma_i < \overline{\gamma_i}$

    3.1.2.1. Set $\hat{S}_i = \min\{S_i : S_i > S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$

    3.1.2.2. Go to step 3.2.

  3.1.3. If $\gamma_i \geq \overline{\gamma_i}$

    3.1.3.1. Set $\hat{S}_i = \max\{S_i : S_i \leq S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$

    3.1.3.2. Go to step 3.2.

3.2. Set $\hat{\mathbf{S}} = (\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N)$

4.1. Compute $\hat{S}_0 = \text{argmin}_{\{S_0 \in [\hat{s}_0+1, S_0^{max}] : \gamma_i \geq \overline{\gamma_i} \; i=1,2,\ldots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, S_0)$

5.1. Compute $\hat{s}_0 = \text{argmin}_{\{s_0 \in [s_0^{min}, \hat{S}_0 - 1]: \gamma_i \geq \overline{\gamma_i} \;\; i=1,2,\dots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, \hat{T}, s_0, \hat{S}_0)$.

6.1. Compute $\hat{T} = \text{argmin}_{\{T \in [T^{min}, T^{max}]: \gamma_i \geq \overline{\gamma_i} \;\; i=1,2,\dots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, T, \hat{s}_0, \hat{S}_0)$.

7.1. Compute $\hat{Q} = \text{argmin}_{\{Q \in [Q^{min}, Q^{max}]: \gamma_i \geq \overline{\gamma_i} \;\; i=1,2,\dots,N\}} AC(Q, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, \hat{S}_0)$.

8.1. Set $n_{it} = n_{it} + 1$.

9.1. Go to step 2.3.

10.1. Set $(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*) = (\hat{Q}, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, \hat{S}_0)$.

Observe that, similar to $(Q, \mathbf{S})$ policy, the problem of finding $\hat{\mathbf{S}}$ for given Q, T, s_0 and S_0 values, is decomposed into N independent problems in each of which $\hat{S}_i$ is assigned to the minimum value of S_i that satisfies the target fill rate. In each of the optimization problem for finding $\hat{S}_0, \hat{s}_0, \hat{T}, \hat{Q}$ defined in steps 4.1, 5.1, 6.1. and 7.1., respectively, we check the feasibility of the corresponding parameters, *ie.* whether the fill rate constraints are satisfied or not.

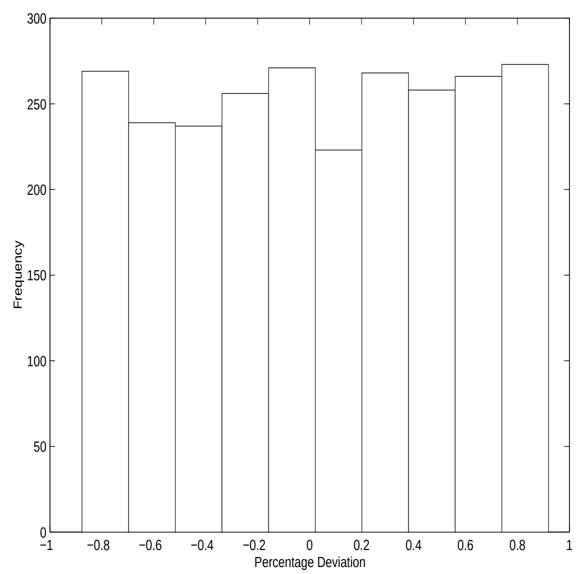

Figure 7.1: Histogram for the Percentage Deviation of the Approximate $AC(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ from simulated $AC(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ for $(Q, \mathbf{S}, T)$ Policy

We also note that at very step of the algorithm where the convolution of the truncated Erlang distribution is necessary, we use the normal approximation explained in Section 6.2. Therefore, the best policy parameters, $(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ are solved using the corresponding approximate cost functions. In order to obtain the true $AC(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ we simulated the inventory system with $(Q, \mathbf{S}, T)$

policy. In our simulations, we used a run length of 20,000 warehouse ordering instances with a warm-up period of 1,000 order placements, and 10 replications. During the remaining part of this chapter, we use the cost figures obtained from the simulations. As a performance of the normal approximation used, we report that, over the 2560 experimental instances where we allow the warehouse to hold stock and whose details will be given in Section 7.4, the average percentage deviation of the approximate cost figures from the simulated ones is 0.06% whereas the maximum and minimum absolute values are 0.91% and 0.002%, respectively. The histogram for the percentage deviations is presented in Figure 7.1. We also note that for the experimental instances where the warehouse acts as a cross-dock facility, the convolutions of truncated Erlang random variables are not required and hence the cost rate functions obtained are exact.

The iterative search algorithm presented above converged to the same policy parameter values for all 50 starting points in 2378 of 2560 experimental instances before hitting the maximum number of iterations. In the remaining 146 of the remaining 182 instances, we hit the maximum number of iterations for at least one initial point. We observed that the solution of the algorithm converged to the same policy parameter values for the other starting points. In 36 experimental instances, the algorithm exceeded the maximum iteration number for all initial points.

In the 1920 experimental instances where the warehouse employs cross-dock, the above algorithm is used excluding steps 4.1. and 5.1. Incidentally, we never hit the maximum number of iterations in these instances and the solution of the algorithm converged to the same policy parameters values for all initial points. Recall that we had a similar numerical observation for $(Q, \mathbf{S}, T)$ policy in the single-location, multi-item context.

7.1.3 $(Q, \mathbf{S}|T)$ Policy

Similar to the $(Q, \mathbf{S}, T)$ policy, the parameters to be optimized under the $(Q, \mathbf{S}|T)$ policy are $(Q, S_1, S_2, \ldots, S_N, T, s_0, S_0)$ with a total of $N + 4$ parameters for an

N retailer inventory system. Although it is easier to calculate the operating characteristics of the $(Q, \mathbf{S}|T)$ policy than that of the $(Q, \mathbf{S}, T)$ policy, an exhaustive search algorithm is also very difficult to implement from computational point of view. Therefore, we also used an iterative search algorithm with random initial points to find the optimal policy parameters for the $(Q, \mathbf{S}|T)$ policy. A total of 50 initial points, $\hat{T}, \hat{Q}, \hat{s}_0$ and $\hat{S}_0$ were selected sequentially over the ranges given below. The increments of each policy parameter within the given ranges, $\Delta_T, \Delta_Q, \Delta_{s_0}$ and Δ_{S_0} are given as follows:

$$T \in [T^{min}, T^{max}], \hat{Q} \in [Q^{min}, Q^{max}], s_0 \in [s_0^{min}, s_0^{max}], S_0 \in [S_0^{min}, S_0^{max}]$$
$$T^{min} = Q_r/\lambda_0, \;\; T^{max} = 20Q_r/\lambda_0 \;\; \Delta_T = (T^{max} - T^{min})/30$$
$$Q^{min} = \left\lceil T^{min}\lambda_0 \right\rceil, Q^{max} = \left\lceil T^{max}\lambda_0 \right\rceil, \Delta_Q = 1$$
$$s_0^{min} = Q_w - 5\left\lfloor \mu_{L_0} \right\rfloor, \;\; s_0^{max} = Q_w + 5\left\lfloor \mu_{L_0} \right\rfloor, \;\; \Delta_{s_0} = 1$$
$$S_0^{min} = s_0 + 1, S_0^{max} = Q_w + 10\left\lfloor \mu_{L_0} \right\rfloor, \;\; \Delta_{S_0} = 1$$

As in the $(Q, \mathbf{S}, T)$ policy, the iterative algorithm starts with randomly selected $\hat{T}, \hat{Q}, \hat{s}_0$ and $\hat{S}_0$ values and ends either when the same policy parameters are obtained in two consecutive iterations or the number of iterations exceeds 3000. One iteration also consists of five consecutive optimization problems for one of the policy parameters while keeping the other four constant. What is different from the $(Q, \mathbf{S}, T)$ policy is the boundary of the ranges as given above and a change in the sequence the optimization problems of a single iteration. The details of the employed iterative search algorithm for a single initial point are presented below. Recall that this search algorithm is repeated for 50 different initial points to obtain the best policy parameters. In the algorithm, n_{it} and $Q^p, T^P, \mathbf{S}^p, s_0^p, S_0^p$ represent the iteration number and the corresponding parameter values in the previous iteration, as before.

Search Algorithm for $(Q, \mathbf{S}|T)$ Policy:

1.1. Set $n_{it} = 0$.

1.2. Select $\hat{T}, \hat{Q}, \hat{s}_0, \hat{S}_0$.

1.3. Set $\hat{\mathbf{S}} = \mathbf{0}, \mathbf{S}^p = \mathbf{1}, T^p = \hat{T} + 1, Q^p = \hat{Q} + 1, s_0^p = \hat{s}_0 + 1, S_0^p = \hat{S}_0 + 1$.

2.1. If $(n_{it} \leq 3000)$ and $(\hat{T} \neq T^p$ or $\hat{Q} \neq Q^p$ or $\hat{\mathbf{S}} \neq \mathbf{S}^p$ or $\hat{s}_0 \neq s_0^p$ or $\hat{S}_0 \neq S_0^p)$

  2.1.1. Go to step 2.3.

2.2. If ($n_{it} > 3000$) or ($\hat{T} = T^p$ and $\hat{Q} = Q^p$ and $\hat{\mathbf{S}} = \mathbf{S}^p$ and $\hat{s}_0 = s_0^p$ and $\hat{S}_0 = S_0^p$)

2.2.1. Go to step 10.1

2.3. Set $T^p = \hat{T}, Q^p = \hat{Q}, \mathbf{S}^p = \hat{\mathbf{S}}, s_0^p = \hat{s}_0, S_0^p = \hat{S}_0$.

3.1. For each retailer $i \in \{1, 2, \ldots, N\}$

3.1.1. Set $S_i^{in} = \lceil \lambda_i(L_i + E[W_0]) \rceil$.

3.1.2. If $\gamma_i < \overline{\gamma_i}$

3.1.2.1. Set $\hat{S}_i = \min\{S_i : S_i > S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$

3.1.2.2. Go to step 3.2.

3.1.3. If $\gamma_i \geq \overline{\gamma_i}$

3.1.3.1. Set $\hat{S}_i = \max\{S_i : S_i \leq S_i^{in}, \gamma_i \geq \overline{\gamma_i}\}$

3.1.3.2. Go to step 3.2.

3.2. Set $\hat{\mathbf{S}} = (\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N)$

4.1. Compute $\hat{S}_0 = \text{argmin}_{\{S_0 \in [\hat{s}_0+1, S_0^{max}]: \gamma_i \geq \overline{\gamma_i} \ \ i=1,2,\ldots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, S_0)$

5.1. Compute $\hat{s}_0 = \text{argmin}_{\{s_0 \in [s_0^{min}, \hat{S}_0 - 1]: \gamma_i \geq \overline{\gamma_i} \ \ i=1,2,\ldots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, \hat{T}, s_0, \hat{S}_0)$.

6.1. Compute $\hat{Q} = \text{argmin}_{\{Q \in [Q^{min}, Q^{max}]: \gamma_i \geq \overline{\gamma_i} \ \ i=1,2,\ldots,N\}} AC(Q, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, \hat{S}_0)$.

7.1. Compute $\hat{T} = \text{argmin}_{\{T \in [T^{min}, T^{max}]: \gamma_i \geq \overline{\gamma_i} \ \ i=1,2,\ldots,N\}} AC(\hat{Q}, \hat{\mathbf{S}}, T, \hat{s}_0, \hat{S}_0)$.

8.1. Set $n_{it} = n_{it} + 1$.

9.1. Go to step 2.3.

10.1. Set $(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*) = (\hat{Q}, \hat{\mathbf{S}}, \hat{T}, \hat{s}_0, \hat{S}_0)$.

For the $(Q, \mathbf{S}|T)$ policy, what makes it easier to calculate the operating characteristics is the discrete structure of both Y and Q_0. For fixed values of Q and T, once $f_{Y^{(n)}, Q_0^{(n)}}(y, q)$ values are calculated for possible values of n, y and q in an iterative way given in Equation (6.7), they can be used in steps 3.1,4.1 and 5.1 without being recalculated. On the other hand, the unbounded structure of both Y and Q_0 is a source of difficulty from computational point of view. Since $f_{Y,Q_0}(y, q)$ can not be calculated for infinitely many values of y and q, we have to truncate the probability mass function at some point while computing it numerically. The convolutions and all other operating characteristics are computed based on the truncated $f_{Y,Q_0}(y, q)$. Hence, the truncation point is important for a correct computation of the cost rate function. While we compute $f_{Y,Q_0}(y, q)$ numerically we stop at the values of $y_0 T$ and q_0 values if $\sum_{m=1}^{y_0} \sum_{k=Q}^{q_0} f_{Y,Q_0}(mT, k) \geq 0.99999$ and $f_{Y,Q_0}(y_0 T, q_0 + 1) \leq 10^{-6}$ and

$f_{Y,Q_0}((y_0+1)T, q_0) \leq 10^{-6}$ hold. The same arguments also hold true for $P_{Q_0}(q)$ and $f_Y(y)$.

The best policy parameters under the $(Q, \mathbf{S}, T)$ policy are found using the truncated $f_{Y,Q_0}(y,q)$, $P_{Q_0}(q)$ and $f_Y(y)$ values and hence using the approximate cost rate function. Therefore, to obtain the true best cost rate function, we simulated the inventory system under the $(Q, \mathbf{S}, T)$ policy with the parameters $(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ found in the iterative search algorithm. In the remaining parts of this chapter, we use the simulated cost figures which are computed by 20,000 warehouse ordering instances after a warm-up period of 1,000 order placements, and 10 replications. Since the truncation issue is also valid for the cases where the warehouse employs cross-dock, we also use the simulated cost figures for these cases.

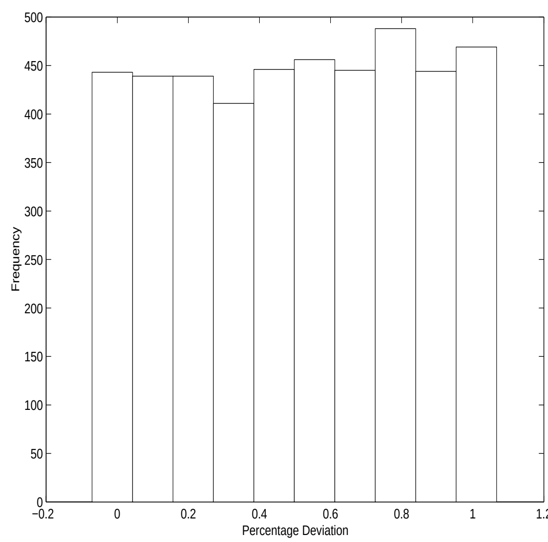

Figure 7.2: Histogram for the Percentage Deviation of the Approximate $AC(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ from simulated $AC(Q^*, \mathbf{S}^*, T^*, s_0^*, S_0^*)$ for $(Q, \mathbf{S}|T)$ Policy

To illustrate the performance of the truncation rule explained above, the average percentage deviation of the approximate cost figures calculated with truncated $f_{Y,Q_0}(y,q)$ from the simulated ones over a total of 4480 (2560 for the instances with the warehouse allowed to hold stock and 1920 cases for cross-dock) cases is 0.52% whereas the maximum and minimum values are 1.01% and 0.01%, respectively. The histogram for the percentage deviations is given in Figure 7.2.

We last present a summary of the performance of the iterative search algorithm presented above. The employed algorithm converged to the same policy parameter values for all 50 starting points in 2159 of 2560 experimental instances where the warehouse is allowed to hold stock. In 297 instances, the algorithm exceeded the maximum number of iterations for at least one initial point. In the remaining 104 experimental instances, the algorithm exceeded the maximum iteration number for all initial points. For the 1920 cross-dock instances in which we omit steps 4.1. and 5.1, the iterative algorithm resulted in the same policy parameter values in 1799 instances. In the remaining 121 instances, maximum number of iterations exceeded for some initial points. However, in these instances, the algorithm converged to the same parameters for the other initial points.

7.1.4 $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ Policy

In an N retailer inventory system, the dimensionality of $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is $2N+2$ with the parameters $(s_1, s_2, \ldots, s_N, S_1, S_2, \ldots, S_N, s_0, S_0)$ to be optimized. If $N > 2$, dimensionality of $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ becomes larger than the dimensionality of $(Q, \mathbf{S}, T)$ or $(Q, \mathbf{S}|T)$ policies. Similarly, if $N > 1$, the dimensionality of $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is larger than that of $(Q, \mathbf{S})$ policy.

As explained in Section 6.4, the structure of the $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is different from the other three policies, the distribution of order quantity is a function of the individual demand rates and individual order quantities. Moreover, in the previous policies, keeping the other parameters fixed, the problem of finding the best $(S_1, S_2, \ldots, S_N)$ can be decomposed into N subproblems. However, for the $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy, this decomposition does not work, since keeping all other parameters constant, when S_i changes for some $i \in [1, 2, \ldots, N]$, $P_{Q_0}(q)$ and hence all the remaining operating characteristics change and need to be recalculated. This complexity makes an exhaustive search algorithm impossible to be implemented. For the $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy , we design a search algorithm which is a combination of the iterative and the exhaustive search procedures.

The search space in the designed algorithm consists of $s_0 \in [s_0^{min}, s_0^{max}], S_0 \in [S_0^{min}, S_0^{max}], s_i \in [s_i^{min}, s_i^{max}], S_i \in [S_i^{min}, S_i^{max}],\ i = 1, 2, \ldots, N$ with increments of $\Delta_{s_0} = 1, \Delta_{S_0} = 1, \Delta_{s_i} = 1, \Delta_{S_i} = 1$. The boundaries of the search space are given as follows:

$$s_0^{min} = Q_w - 5\lfloor \lambda_0 L_0 \rfloor, s_0^{max} = Q_w + 5\lfloor \lambda_0 L_0 \rfloor$$
$$S_0^{min} = s_0 + 1, S_0^{max} = Q_w + 10\lfloor \lambda_0 L_0 \rfloor$$
$$s_i^{min} = \left\lfloor \sqrt{2Kr_i\lambda_i/h_i} \right\rfloor - 5\lfloor \lambda_i(L_0 + L_i) \rfloor, s_i^{max} = \left\lfloor \sqrt{2Kr_i\lambda_i/h_i} \right\rfloor + 5\lfloor \lambda_i(L_0 + L_i) \rfloor$$
$$S_i^{min} = s_i + 1, S_i^{max} = \left\lfloor \sqrt{2Kr_i\lambda_i/h_i} \right\rfloor + 10\lfloor \lambda_i(L_0 + L_i) \rfloor$$

Search Algorithm for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ Policy:

1.1. Set $s_0^{min}, s_0^{max}, S_0^{min}, S_0^{max}$

2.1. For each $s_0 \in [s_0^{min}, s_0^{max}]$ by Δ_{s_0}

2.1.1. For each $S_0 \in [S_0^{min}, S_0^{max}]$ by Δ_{S_0}

2.1.1.1. Set $n_{it} = 0$.

2.1.1.2. Select $\hat{\mathbf{s}}$ and $\hat{\mathbf{S}}$.

2.1.1.3. Set $\mathbf{S}^p = \hat{\mathbf{S}} + 1$ and $\mathbf{s}^p = \hat{\mathbf{s}} + 1$

2.1.1.4. If $(n_{it} \leq 3000)$ and $(\hat{\mathbf{S}} \neq \mathbf{S}^p$ or $\hat{\mathbf{s}} \neq \mathbf{s}^p)$

2.1.1.4.1. Go to step 2.1.1.6.

2.1.1.5. If $(n_{it} > 3000)$ or $(\hat{\mathbf{S}} = \mathbf{S}^p$ and $\hat{\mathbf{s}} = \mathbf{s}^p)$

2.1.1.5.1. Go to step 2.1.1.11.

2.1.1.6. Set $\mathbf{S}^p = \hat{\mathbf{S}}, \mathbf{s}^p = \hat{\mathbf{s}}$

2.1.1.7. For each item $i \in \{1, 2, \ldots, N\}$

2.1.1.7.1. Set $\hat{s}_i = \text{argmin}_{s_i} \left\{ \begin{array}{l} AC(\hat{\mathbf{s}}, \hat{\mathbf{S}}, \hat{s}_0, \hat{S}_0) : s_i \in [s_i^{min}, \hat{S}_i - 1], \\ \gamma_j \geq \overline{\gamma_j} \text{ for } j = 1, 2, \ldots, N \end{array} \right\}$

2.1.1.8. For each item $i \in \{1, 2, \ldots, N\}$

2.1.1.8.1. Set $\hat{S}_i = \text{argmin}_{S_i} \left\{ \begin{array}{l} AC(\hat{\mathbf{s}}, \hat{\mathbf{S}}, \hat{s}_0, \hat{S}_0) : S_i \in [\hat{s}_i + 1, S_i^{max}], \\ \gamma_j \geq \overline{\gamma_j} \text{ for } j = 1, 2, \ldots, N \end{array} \right\}$

2.1.1.9. Set $n_{it} = n_{it} + 1$.

2.1.1.10. Go to step 2.1.1.4.

2.1.1.11. Set $\mathbf{S}^* = \hat{\mathbf{S}}, \mathbf{s}^* = \hat{\mathbf{s}}$.

2.1.1.12. Compute and store $AC(\mathbf{s}^*, \mathbf{S}^*, s_0, S_0)$

3.1. Set $(\mathbf{s}^*, \mathbf{S}^*, s_0^*, S_0^*) = \text{argmin}_{s_0, S_0} AC(\mathbf{s}^*, \mathbf{S}^*, s_0, S_0)$

In the algorithm presented above, notice that we use an iterative procedure to find the best $\mathbf{s}$ and $\mathbf{S}$ values for fixed (s_0, S_0) and use exhaustive search to find the best (s_0, S_0) values. We should also mention that, during the iterative part of the algorithm, in order to find the best $\hat{s}_i$ or $\hat{S}_i$, we also consider the total cost rate in the system in addition to fill rates of all the retailers. Therefore, the computational requirements for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy are much larger than those of the other three policies.

At every step of the algorithm where the average cost rate function and fill rate of the retailers are calculated, if the number of retailers, N, is greater than 2, we use the normal approximation explained in Section 6.4 for the distribution of the order quantity, Q_0, and the retailer order quantities, $R_i(q)$. Therefore, in the 1920 cases where we allow the warehouse to hold stock in an N retailer inventory system with $N > 2$ and 1440 cases where the warehouse acts as a cross-docking facility for $N > 2$, the best policy parameters are computed using the approximate cost rate function. The corresponding true cost rate functions which will be used in the subsequent sections are calculated via a simulation model. The simulation model is run for 20,000 warehouse orders after a warm-up period of 1,000 order placements, and 10 replications.

The histogram of the percentage deviations of the approximate cost function from the simulated ones is presented in Figure 7.3. The average percentage deviation is 0.58 over a total of 3360 instances and the minimum and maximum absolute values are 0.03 and 1.29, respectively.

Over the 2560 instances where we allow the warehouse to hold stock, the best s_0 and S_0 values were never found to be on the boundary of the ranges presented above and we never hit the maximum iteration number in the iterative part of the search algorithm. Similarly, for the cross-dock cases in which the optimization steps over the (s_0, S_0) values are not carried out, the maximum number of iterations is never exceeded.

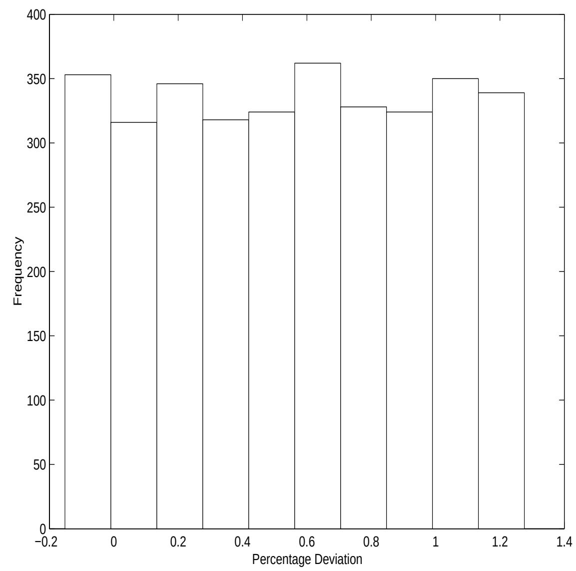

Figure 7.3: Histogram for the Percentage Deviation of the Approximate $AC(\mathbf{s}^*, \mathbf{S}^*, s_0^*, S_0^*)$ from simulated $AC(\mathbf{s}^*, \mathbf{S}^*, s_0^*, S_0^*)$ for $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ Policy

7.2 Advantage of Joint Replenishment Policy in a Two-echelon Inventory System

In this section, we will present some sample results to illustrate the advantage of using various policies under the policy class $\mathcal{P}$ instead of installation stock policies at the retailers.

We first introduce some notation. We let AC_P^* denote the optimal cost rate of a policy P within the policy class $\mathcal{P}$. Also, let $AC(s,S)^*$ be the optimal cost rate of the model where the retailers employ independent (s,S) policies. As a measure of the performance of the policy P over the installation stock (s,S) policy, we define $\Delta\%_P^{(s,S)}$ as follows:

$$\Delta\%_P^{(s,S)} = \frac{AC(s,S)^* - AC_P^*}{AC(s,S)^*} \times 100$$

By definition, a positive $\Delta\%_P^{(s,S)}$ value indicates that the joint replenishment policy P within the class $\mathcal{P}$ performs better than the independent (s,S) policy of the retailers.

We should also mention that an exact analytical model is not available for the installation stock policy with fixed ordering costs at the warehouse and fill rate constraints at the retailers. Therefore, in order to obtain the corresponding best cost rate function for installation stock policy, we have constructed a simulation model and simulated the inventory system operating under independent (s, S) policies at the retailers over a moderate range of policy parameters. For our simulations, we used a run length of 10000 warehouse ordering instances after a warm-up period of 2000 order placements and 10 replications to obtain the average cost rate function of a selected policy parameter. $AC(s, S)^*$ is then found by selecting the minimum of simulated cost rate functions.

In Figures 7.4(a) and (b), we present $\Delta\%_P^{(s,S)}$ values with varying number of retailers for $L_0 = 1$ and $L_0 = 5$, respectively, where the other system parameters are taken as $K_0 = 0, K = 100, h_0 = 1, h_i = 1.2, \lambda_i = 5, L_i = 1, \gamma_i = 0.95, i = 1, 2, \ldots, N$. We should mention that in all of the cases presented in Figures 7.4(a) and (b), the warehouse acts as a cross-dock facility. Therefore, the effective lead time of an order for retailer i is $L_0 + L_i$ (See the details of a cross-dock facility in Section 7.3). The $\Delta\%_P^{(s,S)}$ values presented in Figure 7.4 illustrate that if the retailers use independent replenishments instead of employing joint replenishment decisions, the system wide costs may increase quite significantly, *ie.* the improvements achieved by the joint replenishment decisions may go up to 33.01% and 28.19% for $L_0 = 1$ and $L_0 = 5$, respectively. In addition, we observe an increasing behavior of $\Delta\%_P^{(s,S)}$ values with increasing N for all policies within the class $\mathcal{P}$. This is to be expected because as we increase N, keeping the demand rate of each retailer fixed, the average number of orders placed per time increase and hence the savings of ordering costs obtained from the joint replenishment decisions increase. Also notice that the advantage of using joint replenishment policies decreases as L_0 increases. This may possibly result from a disadvantage of joint replenishment policies in that the retailers lose their flexibility in ordering decisions due to joint replenishment and hence the response to an increase in lead time and/or lead time uncertainty is more effective with independent replenishments. These results are consistent with the results presented in Atkins and Iyogun [4] and Viswanathan [79] for the joint replenishment problem in a single-location, multi-item setting.

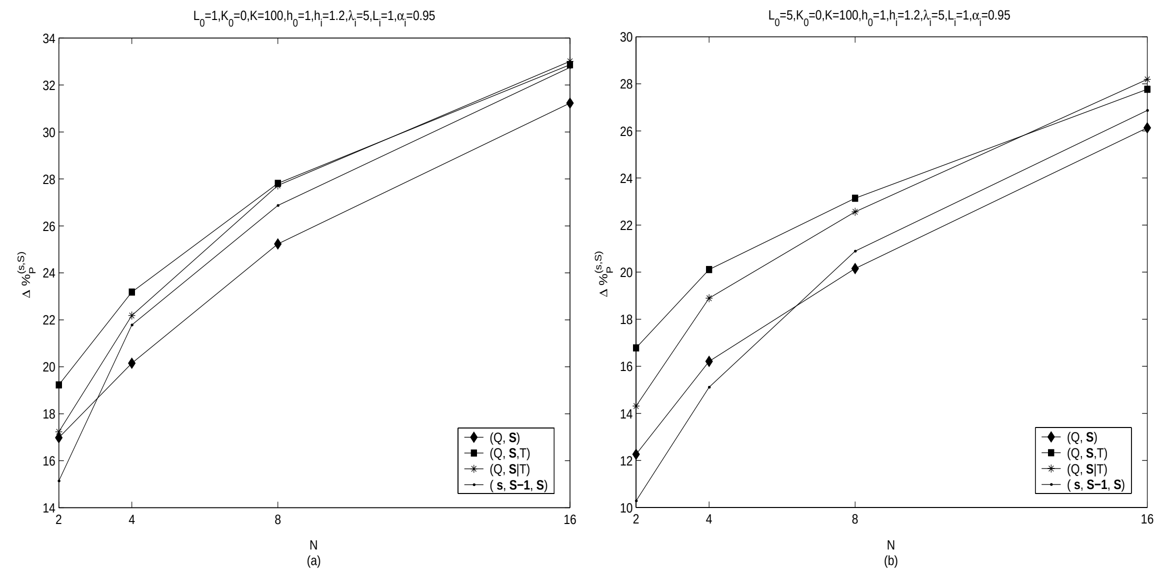

Figure 7.4: Comparison of joint replenishment policies with installation stock policies - $K_0 = 0$

Figures 7.5(a) and (b) illustrate the advantage of using a policy from class $\mathcal{P}$ for the cases $\overline{\gamma}_i = 0.95$ and $\overline{\gamma}_i = 0.99$, respectively. The other parameters in the system are taken as $L_0 = 1, K_0 = 200, K = 100, h_0 = 1, h_i = 1.2, \lambda_i = 5, L_i = 1, i = 1, 2, \ldots, N$. The cases of $K_0 = 200$ and $K_0 = 0$ presented in Figures 7.5(a) and 7.4(a), respectively indicate that the advantage of using joint replenishment generally increases by including an ordering cost at the warehouse. This may be due to the fact that the ordering cost pushes stock and a batch ordering policy at the warehouse. The stock at the warehouse decreases the effective lead time of an order (in case $s_0 \geq 0$) and hence the advantage of using joint replenishments at the retailers increase due to the explanation given above. As we increase the required fill rate of the retailers, $\overline{\gamma}_i$, we observe that $\Delta\%_P^{(s,S)}$ values decrease. This may also be explained by the reduced flexibility of the retailers that employ joint replenishment. On the other hand, the decrease in $\Delta\%_P^{(s,S)}$ values is smaller for $(Q, \mathbf{S}, T)$ and $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policies. $(Q, \mathbf{S}, T)$ policy is a proactive policy which is demonstrated to perform better for larger target fill rates in Table 4.10 in a single-location and multi-item setting. The $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy bases the ordering decisions on the individual inventory positions and hence it can attain the higher fill rate values easier than $(Q, \mathbf{S})$ and $(Q, \mathbf{S}|T)$ policies.

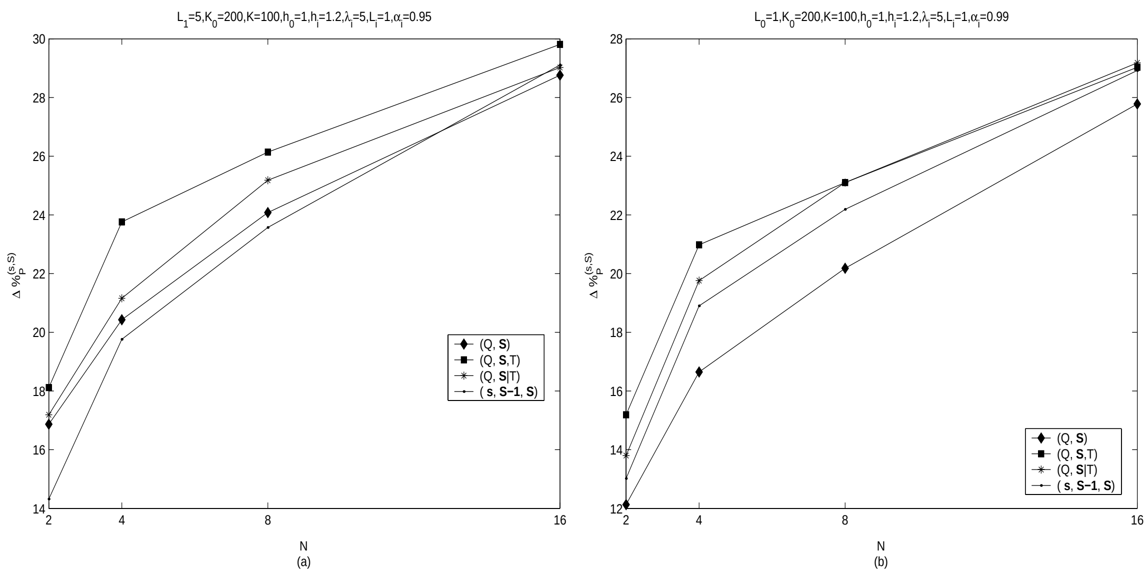

Figure 7.5: Comparison of joint replenishment policies with installation stock policies - $K_0 = 200$

Although joint replenishment policies provide significant cost savings over independent replenishments, Figures 7.4-7.5 illustrate that there is not a single policy within the class $\mathcal{P}$ that dominates the other policies for all system and cost parameters. The performance of a policy P differs depending on the system and cost parameters. In the next two sections, we investigate the performance of policies with respect to each other in detail. In particular, we attempt to identify the operational environments in which it is beneficial to implement a specific policy within the class $\mathcal{P}$.

7.3 Comparison of Policies within Class $\mathcal{P}$ with Warehouse Employing Cross-Dock

In this section, we examine the performance of the policies within the class $\mathcal{P}$ for the special case where the warehouse acts as a cross-dock point. In cross-docking systems, the warehouse functions as a transit point which is in charge of ordering, receiving, unloading, allocating and dispatching shipments. Inventory spends very little time at the warehouse so we assume that the items are shipped to the retailers as soon as they arrive at the warehouse with no reallocation, *i.e.* the warehouse does not hold any inventory and inventory is only held at the retailers.

In the multi-location, single-item model constructed in Chapter 5, if the warehouse employs cross-dock, it operates with $(s_0, S_0) = (-1, 0)$ and the warehouse places an order at the outside supplier for every retailer order and hence the effective lead time of an order for retailer i is given by $L_0 + L_i$. In other words, as also conjectured in Chapter 5, the single-item, two-echelon inventory system where the warehouse acts as a cross-dock point with no reallocation can be represented by the single-location, multi-item model (See also Figure 7.6) in which the lead time for retailer i is given by $L_0 + L_i$ and the fixed ordering cost in the system is $K_0 + K$. The numerical comparison provided in this section is similar to the one presented in Chapter 4 except that the average cost rates of the policies are minimized subject to modified fill rate constraints instead of explicit backorder costs.

For policy comparisons, we let AC_P^* denote the optimal cost rate of a policy P within class $\mathcal{P}$. As a measure of the performance of the policy P, we use the percentage deviation $\Delta_P\%$ defined as follows:

$$\Delta_P\% \quad = \quad \frac{AC_P^* - AC^*}{AC^*} \times 100 \tag{7.1}$$

where AC^* is the cost rate of the best policy within the policy class $\mathcal{P}$, ie. $AC^* = \min_{P \in \mathcal{P}} AC_P^*$. A zero entry for $\Delta_P\%$, by definition, means that the policy P gives the minimum cost rate among the four policies considered. Similarly, a positive,

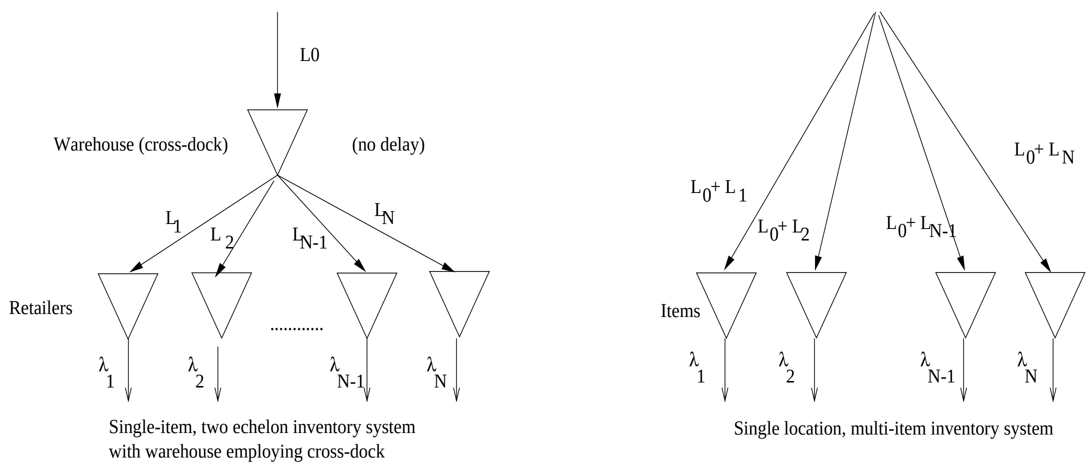

Figure 7.6: Illustration of the analogy of the single-location, multi-item inventory system and the single-item, two-echelon inventory system with cross-dock

lower value of $\Delta_P\%$ indicates that the policy P achieves a lower cost difference from the best policy within the class $\mathcal{P}$.

Before we proceed with individual comparisons, we first present a summary of our findings over all 1920 experiment instances in which the warehouse is assumed to employ cross-dock. We observed that the $(Q, \mathbf{S}, T)$ policy is the best policy of four in 1178 out of 1920 instances with an average and maximum improvement of 0.45% and 1.97% over the next best policy in these instances. In the remaining 742 cases, $(Q, \mathbf{S}|T)$ policy is the best policy in 586 and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is the best in 156 instances. Obviously, $(Q, \mathbf{S})$ policy is never the best one. In the 586 cases where $(Q, \mathbf{S}|T)$ policy is the best, the average deviation of the next best policy from $(Q, \mathbf{S}|T)$ is 0.34% with a maximum deviation of 1.53%. The corresponding figures for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy are 0.23% and 1.21% over the 156 instances where $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ is the best. We also report the average $\Delta_{\mathcal{P}}\%$ values over 1920 experimental instances as 0.23%, 0.98 %, 2.01% and 3.48% for $(Q, \mathbf{S}, T)$, $(Q, \mathbf{S}|T)$, $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policies, respectively.

Next, we discuss our findings in detail. The test bed we are considering consists of $N \in \{2, 4, 8, 16\}$ identical retailers with identical cost and system parameters. We let $K \in \{25, 50, 100, 250\}$, $K_0 \in \{0, K, 2K, 4K\}$, $L_0 \in \{1, 5\}$,

$L_i = L = 1$, $\overline{\gamma}_i = \overline{\gamma} \in \{0.95, 0.99\}$, $\lambda_i = \lambda \in \{0.1, 1, 5, 10, 20\}$, $h_i = h \in \{1, 1.2\}$. The system and cost parameters presented here provide a total of 2560 instances. Notice that these experimental points represent a wide range of parameters from very low ordering and holding costs as well as demand rates to very high values.

Recall that the two-echelon inventory system with the warehouse employing cross-dock acts like a single-location, multi-item inventory system with fixed ordering cost $K_0 + K$. With K_0 values assumed above, we have $K_0 + K \in \{K, 2K, 3K, 5K\}$. If the same $K_0 + K$ value is generated more than once by different K values, we consider it as one instance and that's why we have a total of 1920 experimental instances for cross-dock case.

In Tables 7.1-7.2, we tabulate a total of 64 instances from the 1920 instances which are representative to illustrate the general behavior of the policies with varying cost and demand parameters where we let $L_0 = 5, K_0 = K$. The $(Q, \mathbf{S}, T)$ policy performs better than the other three policies in 50 out of 64 instances. For the remaining 14 experiment instances, it is dominated in 8 cases by $(Q, \mathbf{S}|T)$, and 6 times by $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy. Over the 50 instances where $(Q, \mathbf{S}, T)$ is the best policy among four, the average improvement that it attains over the next best policy is 0.61% with a maximum improvement of 1.13%. Over the 8 instances where $(Q, \mathbf{S}|T)$ is the best policy, the average improvement over the next best policy is 0.16% with a maximum improvement of 0.24%. The corresponding average and maximum improvements for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy are 0.12% and 0.19%, respectively.

The figures presented in Tables 7.1-7.2 and other untabulated results illustrate that time-based policies, namely the $(Q, \mathbf{S}, T)$ and the $(Q, \mathbf{S}|T)$ perform better with smaller demand rates when compared with the quantity based $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policies. Due to the integer values of the parameters with quantity based policies, the optimal parameters of $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ may be quite insensitive to the system parameters. Since the lead time demand is low, a unit change in (s, S) and/or Q values may result in significant cost changes. On the other hand, the continuous time dimension of $(Q, \mathbf{S}, T)$ and $(Q, \mathbf{S}|T)$ policies usually captures the cost/system parameters and this increases the performance

Problem Parameters					$\Delta_P\%$			
λ	K	h	N	AC^*	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S}-\mathbf{1},\mathbf{S})$
1	50	1	2	27.19	0.00	0.72	2.55	2.76
			4	49.23	0.00	0.63	2.02	2.14
			8	95.01	0.00	0.38	1.75	1.58
			16	187.87	0.14	0.00	1.23	1.45
1	50	1.2	2	33.14	0.00	0.99	2.81	2.88
			4	60.68	0.00	0.81	2.23	2.29
			8	126.61	0.00	0.55	1.82	1.76
			16	225.01	0.07	0.00	1.43	1.65
1	100	1	2	45.38	0.00	0.54	2.21	2.12
			4	88.76	0.00	0.38	1.82	1.75
			8	172.34	0.00	0.21	1.56	1.23
			16	334.21	0.22	0.00	1.02	0.93
1	100	1.2	2	59.19	0.00	0.68	2.43	2.21
			4	115.81	0.00	0.59	2.01	1.84
			8	245.93	0.00	0.52	1.75	1.45
			16	454.19	0.13	0.00	1.22	1.14
10	50	1	2	247.43	0.00	0.68	1.85	2.52
			4	457.84	0.00	0.57	1.47	1.95
			8	845.59	0.00	0.32	1.31	1.47
			16	1702.10	0.18	0.00	0.92	1.05
10	50	1.2	2	300.91	0.00	0.91	2.01	2.68
			4	552.79	0.00	0.72	1.64	2.09
			8	1140.76	0.00	0.48	1.28	1.59
			16	2013.84	0.12	0.00	0.99	1.48
10	100	1	2	399.80	0.00	0.51	1.52	1.93
			4	780.20	0.00	0.35	1.29	1.61
			8	1521.76	0.00	0.19	1.03	1.09
			16	2927.68	0.24	0.00	0.71	0.85
10	100	1.2	2	524.42	0.00	0.62	1.72	2.02
			4	1008.71	0.00	0.54	1.43	1.63
			8	2208.45	0.00	0.37	1.31	1.29
			16	4019.58	0.17	0.00	0.85	0.98

Table 7.1: Comparison of policies with the warehouse employing cross-dock - $\overline{\gamma} = 0.95$

of these policies with respect to the others for low demand rates.

The $(Q, \mathbf{S}, T)$ policy best performs with low ordering cost, lower demand rates and higher holding costs. The proactive behaviour of the $(Q, \mathbf{S}, T)$ policy increases its performance for higher fill rate constraints. These observations for the $(Q, \mathbf{S}, T)$ policy are consistent with those presented in Chapter 4, as expected. The performance of the $(Q, \mathbf{S}|T)$ policy is generally better with higher ordering cost and larger values of N. In other words, $\Delta_{(Q,\mathbf{S}|T)}\%$ values decreases with increasing K and/or N. The periodic structure increases the performance of the $(Q, \mathbf{S}|T)$ policy for lower target fill rates. We also observe that the instances where $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy dominates the other policies usually correspond to the

Problem Parameters					$\Delta_P\%$			
λ	K	h	N	AC^*	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
1	50	1	2	34.80	0.00	0.98	2.73	1.48
			4	64.49	0.00	0.76	2.19	1.04
			8	125.41	0.00	0.45	1.88	0.67
			16	251.75	0.09	0.11	1.46	0.00
1	50	1.2	2	40.76	0.00	1.13	2.89	1.56
			4	79.49	0.00	0.92	2.43	1.24
			8	158.26	0.00	0.65	2.04	1.03
			16	274.51	0.00	0.08	1.55	0.45
1	100	1	2	58.09	0.00	0.71	2.35	1.19
			4	118.05	0.00	0.58	1.98	0.97
			8	225.77	0.00	0.35	1.77	0.54
			16	451.18	0.14	0.19	1.32	0.00
1	100	1.2	2	75.17	0.00	0.83	2.54	1.27
			4	152.87	0.00	0.70	2.19	1.11
			8	334.46	0.00	0.59	1.87	0.77
			16	631.32	0.00	0.37	1.65	0.33
10	50	1	2	304.34	0.00	0.86	2.07	1.32
			4	572.30	0.00	0.67	1.67	0.99
			8	1082.35	0.00	0.40	1.49	0.63
			16	2246.77	0.17	0.10	1.09	0.00
10	50	1.2	2	382.16	0.00	0.99	2.19	1.41
			4	729.69	0.00	0.81	1.72	1.14
			8	1551.43	0.00	0.57	1.38	0.95
			16	2799.24	0.12	0.42	1.11	0.00
10	100	1	2	507.74	0.00	0.62	1.72	1.09
			4	1029.86	0.00	0.51	1.45	0.89
			8	2069.60	0.00	0.31	1.19	0.49
			16	4069.47	0.24	0.07	0.83	0.00
10	100	1.2	2	666.02	0.00	0.73	1.91	1.19
			4	1331.49	0.00	0.62	1.55	1.03
			8	2716.40	0.00	0.52	1.43	0.75
			16	5145.06	0.16	0.03	0.92	0.00

Table 7.2: Comparison of policies with the warehouse employing cross-dock - $\overline{\gamma} = 0.99$

cases with larger N, higher fill rate constraints and higher ordering cost. The individual reorder levels for the retailers under $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy enable to cope with the higher fill rates better especially for larger demand rates, *ie.* the integer values of the policy parameters do not eliminate the advantage of reorder levels.

Also notice that the performance of the policies become alike as the number of retailers, N increases. We have a similar observation with increasing demand rates, *e.g.* $\Delta_P\%$ values decrease with increasing N and λ.

In order to give a broader view of the performance of the policies, we next present two tables illustrating the summary for the comparison of the policies. In Table 7.3, we provide a pairwise comparison across the instances where one policy dominates the other.

Policy	Dimensionality	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	N+2	-	1.52 (1246)	2.88 (1920)	3.27 (1481)
$(Q, \mathbf{S}\|T)$	N+2	0.61 (674)	-	1.85 (1920)	2.01 (1075)
$(Q, \mathbf{S})$	N+1	(0)	(0)	-	1.19 (1415)
$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$	2N	0.96 (439)	1.25 (845)	1.03 (505)	-

Table 7.3: The summary comparison of policies across pairwise dominated instances - cross-dock case

Each element (the entry corresponding to the ith row and jth column) of the table reports two entities: the average improvement in the expected cost rate achieved by policy $\mathcal{P}_i$ over policy $\mathcal{P}_j$ in the experimental instances where $\mathcal{P}_i$ dominates $\mathcal{P}_j$; and, the number of such instances in parentheses. The first row of the table gives the performance of the proposed policy in comparison with the other policies. For example, we see that $(Q, \mathbf{S}, T)$ dominates $(Q, \mathbf{S}|T)$ policy in 1246 out of 1920 considered instances; and, the average improvement in such instances achieved over $(Q, \mathbf{S}|T)$ policy is 1.52%. Similarly, $(Q, \mathbf{S}, T)$ policy is better than $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy with an average improvement of 3.27% in 1481 out of 1920 considered instances.

In Table 7.4, we provide an overall comparison of the *average* performance of the policies. In the same format as in Table 7.3, we present the *average* percentage change in the expected total cost rate under policy P_i versus P_j as the entry corresponding to ith row jth column of the table. Differently from Table 7.3, we consider all of the 1920 experiment instances, where P_i may or may not dominate P_j. A positive entry indicates that policy P_i provides that much average percentage improvement in the cost rate over P_j. A negative entry indicates that the performance of P_i is worse by that much, on average, in comparison with the performance of policy P_j.

What is significant about the performance of the policies is that $(Q, \mathbf{S}, T)$ policy is the only one providing cost improvements over the other policies over a very wide range of parameter set. $(Q, \mathbf{S}|T)$ policy is the second if we rank the policies with respect to the average performance presented in Table 7.4. On the average, $(Q, \mathbf{S}|T)$ policy dominates $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policies and is inferior

Policy	Dimensionality	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	N+2	-	1.38	2.88	3.11
$(Q, \mathbf{S}\|T)$	N+2	-1.17	-	1.85	1.79
$(Q, \mathbf{S})$	N+1	-2.54	-1.57	-	1.03
$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$	2N	-2.69	-1.55	-0.92	-

Table 7.4: The overall average performance of policies across all 1920 instances - cross-dock case

only when compared with $(Q, \mathbf{S}, T)$ policy. Although there are instances where $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ dominates the other policies, *ie.* when the ordering cost and/or demand rate and/or target fill rates are high as presented in Table 7.2, the average performance is the worst among the considered policies.

In this study, we assume that the system is centralized and hence our objective is to minimize the total cost rate in the system, rather than dealing with the individual cost of the warehouse and retailers, separately. Tables 7.3 and 7.4 indicate that the performance of $(Q, \mathbf{S}, T)$ policy is better than the other policies when the objective is to minimize the total cost rate in the system. On the other hand, if we considered a decentralized system, the allocation of the costs among the echelons and the difference of echelon costs would be two main issues to be addressed. Moreover, it is important to distinguish the savings/losses of the echelons when the retailers change the policy that they use. In order to gain insight on these issues, we next give summaries of the allocation of the costs among the echelons.

Recall that when the warehouse employs cross-dock, the cost components at the warehouse are the ordering cost incurred for every warehouse order (or for every retailer order) and the holding cost of the items during their transportation from the outside supplier to the warehouse. Therefore, the cost of the warehouse, C_W is given by $(K_0 + h_0 E[Q_0] L_0)/E[Y] = K_0/E[Y] + h_0 \lambda_0 L_0$. Over all 1920 instances where the warehouse employs cross-dock, the average of the proportion of the cost incurred by the warehouse, $C_W^*/AC_{\mathcal{P}}^*$ is the smallest for $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy with 30% and the highest for $(Q, \mathbf{S}, T)$ policy with 39%. The corresponding average values are 33% and 36% for $(Q, \mathbf{S})$ and $(Q, \mathbf{S}|T)$ policies, respectively.

Since the warehouse incurs the maximum proportion of costs under $(Q, \mathbf{S}, T)$

policy, it seems that when the warehouse acts as a cross-dock point, $(Q, \mathbf{S}, T)$ is a policy which is designed for the benefit of the retailers as well as the inventory system itself. Similarly, it is the warehouse which benefits most from implementing $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy under cross-dock. Recall that $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ performs, on the average, the worst when compared with the other policies.

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	-0.72 (492)	-1.01 (197)	-1.58 (239)
$(Q, \mathbf{S}\|T)$	0.83 (1428)	-	-0.49 (724)	-0.12 (819)
$(Q, \mathbf{S})$	1.13 (1723)	0.53 (1196)	-	-0.09 (902)
$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$	1.64 (1681)	0.16 (1101)	0.11 (1018)	-

Table 7.5: The overall average performance of the warehouse - cross-dock case

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	3.19 (1793)	5.26 (1920)	5.81 (1764)
$(Q, \mathbf{S}\|T)$	-3.08 (127)	-	3.62 (1578)	-2.44 (1232)
$(Q, \mathbf{S})$	-5.17 (0)	-3.49 (342)	-	1.68 (1053)
$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$	-5.60 (156)	2.61 (688)	-1.55 (867)	-

Table 7.6: The overall average performance of the retailer - cross-dock case

Table 7.5 (7.6) presents the comparison of the warehouse (retailer) costs across the policies over all 1920 experimental instances. The entries in the ith row and jth column of the table present the *average* percentage change in the expected warehouse (retailer) costs under policy P_i versus P_j and the number of instances in which policy P_i dominates policy P_j in terms of warehouse (retailer) costs. Although the average proportion of costs incurred by the warehouse is achieved by $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy, the average performance of the warehouse is quite close to each other for $(Q, \mathbf{S}|T)$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policies.

We should mention that the performance of the policies is more distinguishable in terms of the retailer costs. When we include the warehouse costs, the percentage deviation of the cost rates of the policies decreases significantly and hence the performance of the policies becomes alike. When the warehouse employs cross-dock, the $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ are two policies in which the warehouse benefits much, *ie.* the entries corresponding to these two policies are in favor of them.

In the next section, we will investigate the performance of the policies for the cases where the warehouse is allowed to hold stock and the differentiation of the echelon costs will become more clear.

7.4 Comparison of Policies within Class $\mathcal{P}$ with Warehouse Allowed to Hold Stock

In the previous section, we examine the performance of the joint replenishment policies studied within the policy class $\mathcal{P}$ for the special case of cross-dock. In this section, we extend the numerical study to the general case where the warehouse is allowed to hold stock. Unlike the special case of cross-dock, the optimization over (s_0, S_0) values is carried out and the optimal policy parameters for each of the four policies within the class $\mathcal{P}$ are computed via the search algorithms presented in Section 7.1. As in the case of cross-dock, we use $\Delta_P\%$ as a performance measure of the policy P, which is given in Equation (7.1).

In this part of the numerical study, our test bed is mainly composed of the experimental instances with $N \in \{2, 4, 8, 16\}$ retailers with identical cost, demand and lead time parameters. In our experimental set, we vary $K \in \{25, 50, 100, 250\}$, $K_0 \in \{0, K, 2K, 4K\}$, $L_0 \in \{1, 5\}$, $L_i = L = 1$, $\overline{\gamma}_i = \overline{\gamma} \in \{0.95, 0.99\}$, $\lambda_i = \lambda \in \{0.1, 1, 5, 10, 20\}$, $h_i = h \in \{1, 1.2\}$ and consider a total of 2560 experimental instances. Notice that the experimental points represent a wide range of retailer/warehouse cost and demand parameters.

Recall that, among the considered policies, $(Q, \mathbf{S})$ policy in Cheung and Lee [23] is the only one which has been previously studied in a two-echelon inventory system. We should also mention that the numerical study provided there is quite restrictive, *e.g.* the size of the orders are assumed to be fixed and hence the ordering cost of the retailers is not considered explicitly in their model and only very low demand rates are considered. Therefore, the numerical study provided herein also presents a detailed performance analysis of the $(Q, \mathbf{S})$ policy over a wide range of parameter set for the first time in the literature.

Before we go on with the details of our numerical study, we first give a summary of our numerical findings over all 2560 experimental instances in which the warehouse is allowed to hold stock. We observe that the $(Q, \mathbf{S}, T)$ policy is the best policy among four in 1928 out of 2560 instances. Over these 1928 instances, the average improvement of $(Q, \mathbf{S}, T)$ policy over the next best policy is 0.51%. The maximum and minimum improvements over the next best policy in these instances is 1.74% and 0.003%, respectively.

$(Q, \mathbf{S}|T)$ policy is the best one in all of the remaining 632 experimental instances where $(Q, \mathbf{S}, T)$ is not the best. The average improvement of $(Q, \mathbf{S}|T)$ over the next best policy is 0.18% with maximum and minimum values of 1.32% and 0.001%, respectively over these 636 instances. Incidentally, among the considered instances $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policy is never the best one.

The average $\Delta_{\mathcal{P}}\%$ values over 2560 experimental instances are found as 0.11%, 0.35%, 1.29% and 1.59% for $(Q, \mathbf{S}, T)$, $(Q, \mathbf{S}|T)$, $(Q, \mathbf{S})$ and $(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$ policies, respectively. The corresponding maximum $\Delta_{\mathcal{P}}\%$ values are found as 0.28%, 1.88%, 2.17% and 2.92%. The summary of our findings indicates that the $\Delta_{\mathcal{P}}\%$ values for the case where the warehouse is allowed to hold stock are generally much smaller than the corresponding $\Delta_{\mathcal{P}}\%$ values for the cross-dock cases.

This observation is also illustrated in Tables 7.7- 7.8 in which we present the results of a subset of 64 representative instances. In order to compare the performance of the joint replenishment policies where the warehouse is allowed to hold stock with that of cross-dock case, in Tables 7.7- 7.8, we use the same cost and system parameters as in Tables 7.1 - 7.2, respectively.

In Tables 7.7-7.8, we present the results of 64 representative instances with $\overline{\gamma} = 0.95$ and 0.99, respectively, for the cases with $\lambda \in \{1, 10\}, N \in \{2, 4, 8, 16\}, h \in \{1, 1.2\}, K \in \{50, 100\}, K_0 = K, L_0 = 5$. The $(Q, \mathbf{S}, T)$ policy is the best among four in 54 of these instances with an average and maximum improvement of 0.48% and 0.94% over the next best policy. $(Q, \mathbf{S}|T)$ policy is the best in the remaining 10 instances. Observe that these 10 instances all correspond to a target fill rate of $\overline{\gamma} = 0.95$. Over these instances, the average and maximum improvement over the next best policy is 0.14% and 0.22%, respectively.

Problem Parameters					$\Delta_P\%$			
λ	K	h	N	AC^*	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S}-\mathbf{1},\mathbf{S})$
1	50	1	2	26.43	0.00	0.60	1.73	2.37
			4	47.41	0.00	0.52	1.37	1.84
			8	90.46	0.00	0.32	1.19	1.36
			16	176.32	0.12	0.00	0.83	1.24
1	50	1.2	2	32.08	0.00	0.82	1.91	2.48
			4	58.19	0.00	0.67	1.52	1.97
			8	120.06	0.00	0.46	1.23	1.50
			16	209.69	0.07	0.00	0.97	1.42
1	100	1	2	44.55	0.00	0.45	1.50	1.82
			4	88.04	0.00	0.32	1.24	1.51
			8	169.08	0.00	0.18	1.05	1.06
			16	323.73	0.18	0.00	0.68	0.81
1	100	1.2	2	58.13	0.00	0.56	1.65	1.90
			4	112.69	0.00	0.49	1.37	1.58
			8	236.59	0.00	0.43	1.19	1.25
			16	431.53	0.12	0.00	0.83	0.98
10	50	1	2	237.86	0.00	0.56	1.26	2.17
			4	436.05	0.00	0.47	1.00	1.68
			8	796.29	0.00	0.26	0.88	1.26
			16	1582.75	0.18	0.00	0.62	0.90
10	50	1.2	2	288.08	0.00	0.76	1.37	2.30
			4	524.30	0.00	0.60	1.12	1.80
			8	1070.33	0.00	0.40	0.87	1.37
			16	1856.05	0.11	0.00	0.67	1.27
10	100	1	2	395.86	0.00	0.42	1.03	1.66
			4	765.36	0.00	0.29	0.88	1.38
			8	1475.92	0.18	0.00	0.70	0.94
			16	2804.29	0.22	0.00	0.48	0.73
10	100	1.2	2	510.86	0.00	0.51	1.17	1.74
			4	973.51	0.00	0.28	0.97	1.40
			8	2108.80	0.10	0.00	0.89	1.11
			16	3785.91	0.17	0.00	0.59	0.86

Table 7.7: Comparison of policies with the warehouse allowed to hold stock - $K_0 = K, \overline{\gamma} = 0.95$

The figures presented in Tables 7.7-7.8 also indicate that the $(Q, \mathbf{S}, T)$ policy generally performs better for higher target fill rates of the retailers when the number of retailers and/or the ordering cost of the retailers are smaller. As we can also observe from the untabulated results, the experimental instances at which the $(Q, \mathbf{S}, T)$ policy is not the best policy usually correspond to the cases where N is large. In these instances, $\Delta_{(Q,\mathbf{S},T)}$ values are usually smaller when K and h values are smaller.

As also tabulated in Tables 7.7-7.8 and expressed above, the performance of the $(Q, \mathbf{S}|T)$ policy is better for lower target fill rates, which is also valid in the untabulated results. Over the 1280 instances of $\overline{\gamma} = 0.99$, the $(Q, \mathbf{S}|T)$ policy

Problem Parameters					$\Delta_P\%$			
λ	K	h	N	AC^*	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S-1},\mathbf{S})$
1	50	1	2	33.19	0.00	0.87	1.99	1.27
			4	60.93	0.00	0.68	1.60	0.89
			8	117.14	0.00	0.40	1.38	0.58
			16	232.22	0.00	0.10	1.08	0.29
1	50	1.2	2	38.71	0.00	0.94	2.11	1.34
			4	74.78	0.00	0.82	1.77	1.07
			8	147.29	0.00	0.58	1.49	0.89
			16	250.96	0.00	0.43	1.13	0.39
1	100	1	2	55.94	0.00	0.63	1.72	1.02
			4	114.87	0.00	0.52	1.45	0.83
			8	217.19	0.00	0.31	1.29	0.46
			16	428.67	0.00	0.17	0.96	0.35
1	100	1.2	2	72.43	0.00	0.74	1.85	1.09
			4	145.92	0.00	0.62	1.60	0.95
			8	315.65	0.00	0.53	1.37	0.66
			16	588.43	0.00	0.33	1.20	0.28
10	50	1	2	287.00	0.00	0.77	1.51	1.13
			4	534.71	0.00	0.60	1.22	0.85
			8	999.81	0.00	0.35	1.09	0.54
			16	2049.70	0.00	0.09	0.80	0.41
10	50	1.2	2	358.91	0.00	0.89	1.60	1.21
			4	678.92	0.00	0.72	1.26	0.98
			8	1427.99	0.00	0.51	1.01	0.82
			16	2530.90	0.00	0.38	0.81	0.73
10	100	1	2	493.19	0.00	0.56	1.26	0.94
			4	991.08	0.00	0.45	1.06	0.77
			8	1969.11	0.00	0.27	0.87	0.42
			16	3823.91	0.00	0.17	0.61	0.19
10	100	1.2	2	636.46	0.00	0.65	1.39	1.02
			4	1260.62	0.00	0.55	1.13	0.89
			8	2542.70	0.00	0.46	1.04	0.65
			16	4756.38	0.00	0.03	0.67	0.45

Table 7.8: Comparison of policies with the warehouse allowed to hold stock - $K_0 = K, \overline{\gamma} = 0.99$

is never the best and the average $\Delta_{(Q,\mathbf{S}|T)}\%$ values over these 1280 instances is 0.74 %. We also observe that $\Delta_{(Q,\mathbf{S},T)}$ values are generally decreasing in N and λ whereas they are generally increasing in K and h. We observe a similar behaviour for $\Delta_{(\mathbf{s},\mathbf{S-1},\mathbf{S})}\%$ values, except that they are decreasing in $\overline{\gamma}$.

In order to illustrate the behaviour of the policies with respect to larger values of K_0, we also present 32 illustrative intances in Table 7.9 where $K_0 = 2K$ and $\overline{\gamma} = 0.95$ $(Q,\mathbf{S},T)$ policy performs the best in 24 of 32 instances with an average improvement of 0.50% over the next best policy. Over the remaining 8 instances, the improvement that $(Q,\mathbf{S}|T)$ policy attains over the second best policy is 0.12 %. Therefore, the performance of $(Q,\mathbf{S}|T)$ policy decreases whereas $\Delta_{(Q,\mathbf{S},T)}\%$

values usually decrease with increasing K_0.

Problem Parameters					$\Delta_P\%$			
λ	K	h	N	AC^*	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S-1},\mathbf{S})$
1	50	1	2	31.19	0.00	0.68	1.55	2.53
			4	55.94	0.00	0.59	1.22	1.94
			8	106.74	0.00	0.36	1.09	1.44
			16	208.45	0.11	0.00	0.74	1.35
1	50	1.2	2	37.85	0.00	0.93	1.72	2.61
			4	68.67	0.00	0.76	1.31	2.08
			8	141.74	0.00	0.52	1.09	1.59
			16	247.43	0.04	0.00	0.86	1.48
1	100	1	2	52.57	0.00	0.51	1.36	1.90
			4	103.89	0.00	0.36	1.12	1.58
			8	199.43	0.00	0.20	0.96	1.11
			16	381.95	0.16	0.00	0.61	0.84
1	100	1.2	2	68.60	0.00	0.64	1.49	1.97
			4	132.97	0.00	0.55	1.22	1.64
			8	279.18	0.00	0.49	1.07	1.28
			16	509.20	0.11	0.00	0.73	1.02
10	50	1	2	280.67	0.00	0.64	1.12	2.28
			4	514.54	0.00	0.53	0.90	1.77
			8	939.55	0.00	0.30	0.80	1.33
			16	1867.80	0.14	0.00	0.55	0.95
10	50	1.2	2	339.93	0.00	0.85	1.21	2.42
			4	618.67	0.00	0.68	0.99	1.90
			8	1262.99	0.00	0.45	0.77	1.42
			16	2190.14	0.09	0.00	0.59	1.31
10	100	1	2	467.11	0.00	0.48	0.92	1.68
			4	903.13	0.00	0.33	0.78	1.43
			8	1741.59	0.00	0.17	0.62	0.97
			16	3309.07	0.20	0.00	0.43	0.74
10	100	1.2	2	602.81	0.00	0.58	1.05	1.82
			4	1148.74	0.00	0.32	0.87	1.43
			8	2486.58	0.00	0.19	0.80	1.17
			16	4469.71	0.13	0.00	0.52	0.87

Table 7.9: Comparison of policies with the warehouse allowed to hold stock - $K_0 = 2K, \overline{\gamma} = 0.95$

An important observation is that the decrease in $\Delta_{(Q,\mathbf{S})}\%$ values are more significant than the decrease in $\Delta_{(Q,\mathbf{S},T)}\%$ values and hence $(Q,\mathbf{S},T)$ and $(Q,\mathbf{S})$ policies become alike when K_0 increases. The decreasing behaviour of $\Delta_{(Q,\mathbf{S})}\%$ values is also valid when the warehouse is allowed to hold stock instead of employing cross-dock.

When we compare the $\Delta_P\%$ values presented in Tables 7.7- 7.8 with those in Tables 7.1-7.2 and other unreported results, we observe that the general behaviour of the policies in the instances where the warehouse is allowed to hold stock are quite similar to those for the cross-dock case. On the other hand, as indicated above $\Delta_P\%$ values are usually much smaller than the corresponding figures of cross-dock case as the summary performance of the policies presented in Tables 7.10-7.11.

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	0.51 (1928)	1.24 (2560)	1.54 (2451)
$(Q, \mathbf{S}\|T)$	0.12 (632)	-	1.02 (2560)	1.27 (1974)
$(Q, \mathbf{S})$	(0)	(0)	-	0.75 (1693)
$(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$	0.06 (109)	0.22 (586)	0.52 (867)	-

Table 7.10: The summary comparison of policies across pairwise dominated instances - warehouse allowed to hold stock

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	0.35	1.24	1.46
$(Q, \mathbf{S}\|T)$	-0.27	-	1.02	0.97
$(Q, \mathbf{S})$	-1.17	-0.95	-	0.36
$(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$	-1.31	-0.81	-0.31	-

Table 7.11: The overall average performance of policies across all 2560 instances - warehouse allowed to hold stock

The summary figures presented in Tables 7.10-7.11 illustrate that it becomes very difficult to differentiate the policies with the objective of minimizing the total cost rate function although the ranking of the policies is the same with that of the case where the warehouse employs cross-dock. In order to distinguish the policies, we next present a summary of the allocation of the costs among the echelons, which is more distinctive to compare the policies.

Over the 2560 instances where the warehouse is allowed to hold stock, the average proportion of the warehouse cost expressed by $C_W^*/AC_{\mathcal{P}}^*$ is given by 45%, 48%, 51% and 53% for $(Q, \mathbf{S})$, $(Q, \mathbf{S}, T)$, $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ and $(Q, \mathbf{S}|T)$ policies, respectively. Observe that, when the warehouse is allowed to hold stock, the proportion of the costs that the warehouse incurs is larger than that of the cross-dock case for each of the four policies considered. Although this is an expected behaviour, the order of the policies in terms of the proportion of warehouse costs changes completely, which may be explained by the different behaviour of μ_{L_0} and c_{L_0} which correspond to the expected value and the coefficient of variation of warehouse lead time demand, respectively. Both of these quantities can be considered as a measure of the safety stock of the warehouse.

Figure 7.7 illustrates μ_{L_0} vs c_{L_0} values for the particular instance with $L_0 = 1, \lambda = 5, K = 50, K_0 = K, h = 1.2$. We should mention that this instance represents the general behaviour of μ_{L_0} and c_{L_0} across the policies. Although $(Q, \mathbf{S})$ policy imposes a constant reorder size, it is interesting to see that $(Q, \mathbf{S}, T)$ policy has the smallest c_{L_0} value which is followed by c_{L_0} value of $(Q, \mathbf{S})$ policy. This advantage of $(Q, \mathbf{S}, T)$ policy possibly results from the bounded structure of both Y and Q_0. When we compare μ_{L_0} values across the policies, we see that the $(Q, \mathbf{S})$ policy usually attains the smallest value which is followed by $(Q, \mathbf{S}, T)$ policy. Since the smallest values of both μ_{L_0} and c_{L_0} values are achieved by $(Q, \mathbf{S})$ and $(Q, \mathbf{S}, T)$ policies, the warehouse level incurs the least proportion of the total costs by these policies.

On the other hand, the highest c_{L_0} and μ_{L_0} values are accomplished with either $(Q, \mathbf{S}, T)$ or $(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$ policies. The warehouse, on the average, incurs more cost with these policies since more safety stock must exist at the warehouse. On the other hand, the proportion of the warehouse costs at each of the experimental instance obviously depends on the trade-off between the on-hand inventory and the ordering frequency at the warehouse.

We next present the summary of the comparison of echelon costs across the policies in Tables 7.12-7.13. When the warehouse is allowed to hold stock, the warehouse costs under $(Q, \mathbf{S})$ policy are smaller than those of the other three policies. In terms of the retailer costs, the $(Q, \mathbf{S}|T)$ policy is the one which benefits much by allowing the warehouse to hold stock.

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	1.03 (1871)	-0.95 (438)	0.42 (2560)
$(Q, \mathbf{S}\|T)$	-0.94 (689)	-	-0.33 (111)	-0.05 (1819)
$(Q, \mathbf{S})$	1.02 (2122)	1.73 (2449)	-	1.31 (2560)
$(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$	-0.39 (0)	0.06 (741)	-1.25 (0)	-

Table 7.12: The overall average performance of the warehouse - warehouse allowed to hold stock

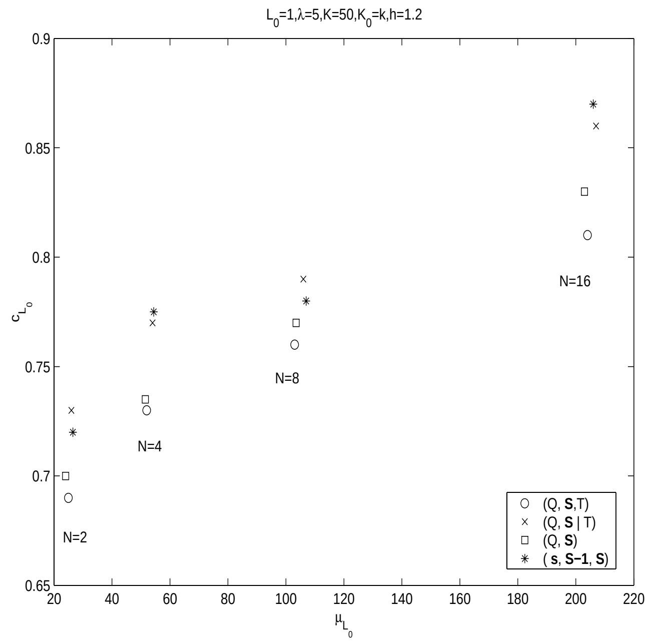

Figure 7.7: Illustration of μ_{L_0} vs c_{L_0} across policies

Policy	$(Q, \mathbf{S}, T)$	$(Q, \mathbf{S}\|T)$	$(Q, \mathbf{S})$	$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$
$(Q, \mathbf{S}, T)$	-	-1.15 (1625)	3.13 (2560)	2.12 (2125)
$(Q, \mathbf{S}\|T)$	1.18 (935)	-	1.97 (2560)	1.84 (2411)
$(Q, \mathbf{S})$	-3.02 (0)	-1.89 (0)	-	-0.75 (1076)
$(\mathbf{s}, \mathbf{S}-\mathbf{1}, \mathbf{S})$	-2.04(435)	-1.79 (149)	0.81 (1484)	-

Table 7.13: The overall average performance of the retailers - warehouse allowed to hold stock

7.5 Advantage of Allowing the Warehouse to Hold Stock

Although implementing cross-dock mainly aims to reduce the average inventory level in a supply chain but it can result in more inventory required to achieve the same fill rates at the retailers and the costs in the system may increase significantly. On the other hand, cross-dock systems are often implemented due to the ease of optimization and implementation. In this part of the numerical study, we aim to quantify the adavantage of allowing the warehouse to hold stock.

As a measure of the advantage of allowing the warehouse to hold stock, we define $\Delta^*_{W/C}\%$ as follows:

$$\Delta^*_{W/C}\% = \frac{AC^*_c - AC^*_w}{AC^*_w}$$

Here, AC^*_c and AC^*_w correspond to the best cost rate under the policy class $\mathcal{P}$ where the warehouse employs cross-dock and is allowed to hold stock, respectively. Obviously, $AC^*_c \geq AC^*_w$.

Problem Parameters					$\Delta^P_{W/C}\%$			
λ	K	h	N	$\Delta^*_{W/C}\%$	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S}-\mathbf{1},\mathbf{S})$
1	50	1	2	2.88	2.88	3.01	3.71	3.27
			4	3.84	3.84	3.95	4.50	4.15
			8	5.03	5.03	5.10	5.61	5.26
			16	6.35	6.36	6.35	6.76	6.56
1	50	1.2	2	3.31	3.31	3.48	4.22	3.71
			4	4.28	4.28	4.42	5.01	4.60
			8	5.41	5.41	5.51	6.01	5.66
			16	7.31	7.31	7.31	7.79	7.55
1	100	1	2	1.86	1.86	1.96	2.57	2.16
			4	0.82	0.82	0.88	1.40	1.06
			8	1.97	1.97	2.01	2.48	2.15
			16	3.25	3.29	3.25	3.59	3.38
1	100	1.2	2	1.82	1.82	1.94	2.60	2.13
			4	2.77	2.77	2.87	3.42	3.03
			8	3.95	3.95	4.04	4.52	4.16
			16	5.25	5.26	5.25	5.66	5.42
10	50	1	2	4.02	4.02	4.14	4.63	4.38
			4	5.00	5.00	5.10	5.49	5.28
			8	6.20	6.20	6.26	6.64	6.42
			16	7.53	7.55	7.53	7.85	7.69
10	50	1.2	2	4.45	4.45	4.62	5.12	4.84
			4	5.44	5.44	5.56	5.98	5.74
			8	6.58	6.58	6.67	7.01	6.81
			16	8.50	8.51	8.50	8.84	8.72
10	100	1	2	1.00	1.00	1.08	1.48	1.26
			4	1.94	1.94	2.00	2.36	2.17
			8	3.11	2.92	3.30	3.44	3.26
			16	4.40	4.42	4.40	4.64	4.52
10	100	1.2	2	2.66	2.66	2.76	3.21	2.94
			4	3.62	3.62	3.88	4.08	3.85
			8	4.80	4.71	5.31	5.24	4.99
			16	6.12	6.13	6.12	6.40	6.26

Table 7.14: Advantage of Allowing the Warehouse to Hold Stock - $K_0 = K, \overline{\gamma} = 0.95$

Similarly, we define $\Delta^P_{W/C}\%$ as a measure of the performance of allowing the warehouse to hold stock at the warehouse under policy P within the class $\mathcal{P}$.

$$\Delta^P_{W/C}\% = \frac{AC^P_c - AC^P_w}{AC^P_w}$$

where AC_c^P and AC_w^P correspond to the optimal cost rate of the policy P for the cases where the warehouse employs cross-dock and is allowed to hold stock, respectively.

We present 64 representative results to illustrate the general behaviour of $\Delta^*_{W/C}\%$ and $\Delta^P_{W/C}\%$ in Tables 7.14 and 7.15 where we compare the values presented in Tables 7.1-7.7 and 7.2-7.8, respectively.

Problem Parameters					$\Delta^P_{W/C}\%$			
λ	K	h	N	$\Delta^*_{W/C}\%$	$(Q,\mathbf{S},T)$	$(Q,\mathbf{S}\|T)$	$(Q,\mathbf{S})$	$(\mathbf{s},\mathbf{S-1},\mathbf{S})$
1	50	1	2	4.87	4.87	4.99	5.63	5.09
			4	5.85	5.85	5.94	6.47	6.01
			8	7.07	7.07	7.12	7.60	7.17
			16	8.42	8.52	8.43	8.82	8.10
1	50	1.2	2	5.31	5.31	5.44	6.11	5.53
			4	6.29	6.29	6.40	6.98	6.48
			8	7.46	7.46	7.53	8.03	7.60
			16	9.39	9.39	9.44	9.84	9.45
1	100	1	2	3.83	3.83	3.92	4.48	4.01
			4	2.77	2.77	2.84	3.31	2.91
			8	3.95	3.95	3.99	4.44	4.03
			16	5.25	5.40	5.27	5.62	4.88
1	100	1.2	2	3.79	3.79	3.88	4.49	3.97
			4	4.76	4.76	4.84	5.37	4.92
			8	5.96	5.96	6.03	6.49	6.07
			16	7.29	7.29	7.33	7.76	7.34
10	50	1	2	6.04	6.04	6.14	6.62	6.23
			4	7.03	7.03	7.11	7.51	7.18
			8	8.26	8.26	8.30	8.69	8.35
			16	9.61	9.80	9.63	9.93	9.17
10	50	1.2	2	6.48	6.48	6.59	7.10	6.69
			4	7.48	7.48	7.57	7.97	7.65
			8	8.64	8.64	8.71	9.04	8.79
			16	10.60	10.74	10.65	10.93	9.80
10	100	1	2	2.95	2.95	3.02	3.42	3.11
			4	3.91	3.91	3.97	4.32	4.04
			8	5.10	5.10	5.14	5.44	5.17
			16	6.42	6.68	6.44	6.66	6.22
10	100	1.2	2	4.64	4.64	4.73	5.18	4.82
			4	5.62	5.62	5.69	6.06	5.77
			8	6.83	6.83	6.89	7.24	6.94
			16	8.17	8.34	8.18	8.44	7.69

Table 7.15: Advantage of allowing the warehouse to hold stock - $K_0 = K, \overline{\gamma} = 0.99$

As presented in Tables 7.14-7.15, we observe that $\Delta^*_{W/C}\%$ values increase with increasing K, N, h and λ values. Increasing K decreases the ordering frequency of the retailers and the warehouse tries to hold stock in order to decreaese the waiting time of an infrequent order. When h changes from 1 to 1.2, inventory is pushed to the warehouse level to cope with the increase in the unit holding cost of the retailers. Increasing λ and N increases the ordering rate of the retailers and holding stock at the warehouse decreases the ordering rate of the warehouse which will be more frequent with cross-dock case. Similarly, when $\overline{\gamma}$ increases, the advantage of holding stock at the warehouse also increases due to the reduced waiting time of an order at the warehouse. Similar observations are also valid for $\Delta^P_{W/C}\%$ values and the maximum $\Delta^P_{W/C}\%$ is generally obtained with $(Q, \mathbf{S})$ policy. The minimum $\Delta^P_{W/C}\%$ values are usually achieved by $(Q, \mathbf{S}, T)$ policy when N is smaller or $(Q, \mathbf{S}|T)$ policy when N is large. We should also mention that there are some instances where the cases of warehouse acting as cross-dock and allowed to hold stock result in different best policies. These instances usually correspond to higher number of retailers ($N = 8$ *or* 16) in which the best policy switches from $(Q, \mathbf{S}|T)$ to $(Q, \mathbf{S}, T)$ policy and higher fill rate constraints in which the best policy switches from $(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$ to $(Q, \mathbf{S}, T)$ policy. For these cases, $\Delta^*_{W/C}\%$ values are different from $\Delta^P_{W/C}\%$ values. Otherwise, $\Delta^*_{W/C}\%$ values are equal to $\Delta^P_{W/C}\%$ value of the best policy for both cases.

We should also mention that, among the considered instances, only for $K_0 = 0$, the optimal policy parameters indicate that the warehouse should act as a cross-docking point. In all 640 experiments with $K_0 = 0$, the $(\mathbf{s}, \mathbf{S} - \mathbf{1}, \mathbf{S})$ and $(Q, \mathbf{S})$ policies employ cross-docking at the optimal parameters. The number of such instances is 526 and 562 for $(Q, \mathbf{S}, T)$ and $(Q, \mathbf{S}|T)$ policies, respectively. In the remaining 1920 instances, the optimal policy parameters indicate that the warehouse should hold stock. We also observe that as the ratio K_0/K increases, $\Delta^*_{W/C}\%$ and $\Delta^P_{W/C}\%$ values increase significantly, as expected. Over all the 2560 instances, the average of $\Delta^*_{W/C}\%$ values is found to be 5.12 % which means that employing cross-dock at the warehouse may lead to significantly higher cost values especially if the warehouse ordering cost is high.

Chapter 8

Conclusion

In this study, we consider the stochastic joint replenishment problem in a single-location/multi-item and single-item/multi-location inventory settings. The stochastic joint replenishment problem is originally defined in a multi-item inventory setting and it aims to determine the optimal replenishment and stocking decisions to minimize the expected total ordering, holding and backorder costs in the system in the presence of random demands and joint ordering cost structure. Because of the applicability to multi-location inventory systems, stochastic joint replenishment problem is a challenging research area. This study is among the recent few studies considering the stochastic joint replenishment problem in two-echelon supply chain.

In this chapter, the contributions of this study will be explained and some future research directions will be provided.

8.1 Contributions

The problem of replenishment coordination strategies has been one of the most important issues faced especially by practitioners for years. This issue has become even more critical in recent years with the recent advances in information technology as information sharing between the parties involved in the supply chain. Therefore, the stochastic joint replenishment problem is a a real problem faced by retailers and is an integral part of supply chain management in general. Despite its practical importance, solution of the stochastic joint replenishment problem is extremely difficult. The existing policies in the literature do not dominate each other uniformly over the entire parameter space.

In the first part of the study, we have proposed a new parsimonious ordering policy for stochastic joint replenishment problem in a single location, N-item setting. The replenishment decisions are based on both group reorder point-group order quantity and the time since the last decision epoch. We derive the expressions for the key operating characteristics of the inventory system for both unit and compound Poisson demands and constructed the expected cost rate function explicitly.

An extensive numerical study has been conducted to study the sensitivity of the policy to various system parameters and to assess the performance of the proposed policy over the existing policies in the literature. The numerical experiments indicate that there is no clear demarcation of operating environments for the dominance of proposed policies in the literature and that the dominance of the proposed policy is not monotone over the experiments. However, similarity of items in their cost structure appears to be most critical factor in the dominance of the proposed policy. The diversity of the individual demand rates is also an important factor. We have found that the proposed policy provides significant savings over the existing policies for items similar in their cost structures and individual demand rates. This finding may have important implications for supply chain design.

The proposed policy attains such performance levels with parsimony. This parsimony reduces the computational effort in optimization enormously and eases implementation in practice greatly. Viewing the comparison in this broader perspective, we believe that the proposed policy and the model developed herein provide significant improvements over the existing models in terms of cost savings, optimization effort and ease of implementation. Although we motivate our model in a single-location, multi-item setting, it can also be used in a two-echelon, single-item, multi-retailer setting with cross docking at the upper echelon.

In the second part of the study, we extend our model to incorporate a two-echelon inventory system where the upper echelon also holds inventory. We studied a policy class under the joint replenishment problem in a two-echelon divergent inventory system. The policy class bases the ordering decisions on the ordering opportunities that arrive according to some prespecified rule. At each ordering opportunity, the retailers are all replenished to their maximum inventory positions to take full advantage of savings of the ordering cost. In order to analyze the two-echelon inventory system under a policy class, we have developed a new generic framework which is only based on the ordering process of the retailers. The proposed methodology is not specific to a particular policy but is applicable to any policy that satisfies the characteristics of the considered class.

Our modeling methodology provides us an analytical tool to investigate various joint replenishment policies under the considered policy class. We have provided the expressions and approximations for the operating characteristics of four different policies and provided insights for the behaviour of these policies. Among these policies, only $(Q, \mathbf{S})$ policy was previously studied in a two-echelon inventory system [23].

An extensive numerical study was conducted to investigate the performance of these policies. The numerical experiments indicate that the policy, which has been proposed proposed for the multi-item/single-location model, also provides cost savings over the other policies within the considered policy class especially when the number of retailers in the system is small. When the warehouse is allowed to hold stock, the cost savings achieved by a policy over another are not as significant as in the case of cross-dock (or single-location/multi-item) case. However, the allocation of the costs among the echelons and the comparison of these costs across the policies provide better distinction between these policies. We also note that the holding cost associated with the pipeline inventory, which is the same for all policies, is omitted the magnitude of the percentage deviations across the policies would increase.

Moreover, the methodology proposed for the warehouse deals with one of the research questions provided in Hill [44], *ie.* modeling the lead time with an additional delay in case the partial shipments are not allowed. The distributions presented for different policies in Chapter 6 also provide insights for the relationship between the order size, inter-order time and waiting time and implications for the inventory model. The analysis at the warehouse is also applicable to a single-item/single-location model facing renewal batch demands where the inter-order time and the inter-order quantity have a bivariate distribution.

Finally, we believe that this part of the study serves as a starting point for the analysis of more complex joint replenishment policies in multi-echelon inventory systems.

8.2 Future Research Directions

The basic objective of this study was to present the joint replenishment concept in a 2-echelon inventory system and to develop a basic analytical model for a class of joint replenishment policies. In this section, we provide possible research extensions.

A common practice in typical retail supply chains is to employ cross-dock at the warehouse although it may lead to significant cost rate increases as presented in Section 7.5. On the other hand, if cross-docks have real time information about the inventory status of the retailers and are also in charge of the allocation of goods to the retailers, it is possible to reallocate the units in an order before they are shipped to the retailers according to a prespecified allocation rule. This effectively reduces the lead time variability of the retailers since the number of units to be shipped to each of the retailer is decided according to the inventory status L_0 time units after the order is given. A quite preliminary study conducted indicated that, for each of the joint replenishment policies within policy class $\mathcal{P}$, it is possible to reduce the costs in the system even by using an allocation rule which ensures equal stock-out probabilities for the retailers. The numerical study demonstrated that when an allocation is performed for $(Q, \mathbf{S})$ policy, it may outperform the $(Q, \mathbf{S}, T)$ policy without an allocation especially if the demand diversity among the items is significant. The allocation problem in a cross-dock setting is an important issue that should be studied in detail.

Recall that if the warehouse holds stock the dispatching policy does not allow a partial shipment and is based on a first-come-first-serve rule. Within the joint replenishment context, this assumption may be a fair and adequate one since we are basically trying to reduce the ordering costs in the system. However, it is quite restrictive and may lead to higher average cost rates since an order has to wait until sufficient inventory accumulates at the warehouse, which increases the effective lead time of an order. An alternative way to relax this assumption may be to allow partial shipments, which may lead to more complex problems *e.g.* how to allocate the existing stock to the retailers. There is an additional option of subcontracting for immediate additional shipments from the outside supplier and ship the whole order immediately to the retailers. In case the stock at the warehouse is not enough to satisfy an order of the retailers, the unsatisfied part will be shipped to the warehouse from the outside supplier immediately at an additional cost so that the order will be dispatched without waiting at the warehouse. In that case, the effective lead time of an order will always be L_0 time units. However, there will be two different supply modes for the warehouse, which may also increase the average inventory level. A more intelligent subcontracting option may consider the number of unsatisfied units and/or the time that remains until all units in the order are satisfied. This option will obviously capture the trade off between the increased stock at the warehouse and better service levels at the retailers.

In this study, we basically considered the joint replenishment problem under single-location/multi-item and single-item/two-echelon settings. An obvious extension is to make a complete analysis of the multi-item/two-echelon inventory system which is extremely difficult. Therefore, a 2-echelon serial system with a single warehouse and a single retailer may be a good starting point to analyze multi-item and multi-location inventory systems. In addition, this setting is important since it represents a typical supermarket chain where the warehouse represents the central distribution center and the retailer is the market itself. In order to dispatch a joint order without a partial shipment, there must be enough units from each item in the order at the warehouse and hence the order will wait at the warehouse until sufficient units from each item exist. Then, the waiting time of an order at the warehouse will be the maximum time that elapses until sufficient inventory of the items in the order accumulates at the warehouse.On the other hand, once analytical model for this setting can be developed, it will be easy to extend it to a more general multi-location, multi-item setting.

In this study, we considered inventory systems in which the items or the retailers are jointly ordered according to a prespecified joint replenishment rule, *ie.* all items/retailers use the same joint replenishment policy. However, as demonstrated in Section 4.3.2, the dominance of each policy is strongly dependent on how the individual demand rates are distributed among the items/retailers. This finding has important implications for supply chain design and management. In a multi-item setting, the problem of clustering the items and determining a joint replenishment policy for each cluster is important to eliminate the disadvantage of reduced flexibility in joint replenishment policies. It would also be interesting to investigate the joint location-allocation-replenishment problem in a supply chain. The extension of correlated demands across the items/retailers would also be an interesting issue to focus on.

Chapter 9

Appendix

Notation Table

$\mathcal{A}(n, y, q)$	:	Set of Y_i's and Q_i's such that $\sum_{i=1}^{n} Y_i = y$ and $\sum_{i=1}^{n} Q_i = q$
$AC^*_{\mathcal{P}}$	:	Optimal cost rate under policy $\mathcal{P}$
AC^*_c	:	Optimal cost rate with warehouse employing cross-dock
AC^*_w	:	Optimal cost rate with warehouse allowed to hold stock
AC^P_c	:	Optimal cost rate with warehouse employing cross-dock under policy P
AC^P_w	:	Optimal cost rate with warehouse allowed to hold stock under policy P
$AC(p_1, p_2, ...p_n)$	:	Expected cost rate function of a policy with parameter $p_1, p_2, \ldots, p_n$
$AR_i(m_i, q)$	:	Age of m_i units allocated to retailer i in an order of size q
α	:	Average number of units demanded per time
$B_i(m_i, q)$	:	Number of items from m_i units in an order of size q which are used to satisfy backordered demands at retailer i
$BO_0(t)$	:	Backorder level of the warehouse at time t
$BO_i(t)$	:	Backorder level of item i at time t
BO_i	:	Steady-state backorder level of item i
$\mathbf{c}$	:	$(c_1, c_2, \ldots, c_N)$
C_0	:	Normalizing constant for $g(\cdot, \cdot, \cdot)$ under unit Poisson demand

C_1	:	Normalizing constant for $g(\cdot,\cdot,\cdot)$ under compound Poisson demand
c_i	:	Can-order level of retailer i
C_i	:	Normalizing constant for $g(i,\cdot)$
C_R	:	Cost rate of the retailers
C_W	:	Cost rate of the warehouse
C_R^*	:	Optimal cost rate of the retailers
C_W^*	:	Optimal cost rate of the warehouse
c.d.f	:	Cumulative distribution function
c_{L_0}	:	Coefficient of variation of warehouse lead time demand
$D_0(t_1,t_2]$	:	Number of units demanded from the warehouse in $(0,t]$
$D_0[t_1,t_2)$	:	Number of units demanded from the warehouse in $[0,t)$
$D_0(t_1,t_2)$	:	Number of units demanded from the warehouse in $(0,t)$
$D_i(t_1,t_2]$	:	Number of demands arriving for item i during $(t_1,t_2]$
Δ_0	:	$S_0 - s_0$
Δ_i	:	$S_i - s_i$
Δ_T	:	Increments for T in the search space
Δ_Q	:	Increments for Q in the search space
Δ_{S_i}	:	Increments for S_i in the search space
Δ_{s_0}	:	Increments for s_0 in the search space
Δ_{S_0}	:	Increments for S_0 in the search space
$\Delta_{\mathcal{P}}\%$	:	Percentage improvement of $(Q,\mathbf{S},T)$ policy over policy $\mathcal{P}$ in the multi-item model or percentage deviation of policy $\mathcal{P}$ from the best one in the two-echelon model
$\Delta\%_{\mathcal{P}}^{(s,S)}$	:	Percentage improvement of a policy $\mathcal{P}$ over (s,S) policy
$\Delta_{W/C}^*\%$	:	Percentage improvement achieved by holding stock over cross-dock
$\Delta_{W/C}^P\%$	:	Percentage improvement achieved by holding stock over cross-dock under policy P
$E[BT_i]$	:	Expected number of backorders given by retailer i per time
$f_Y(\cdot)$	:	P.d.f. of Y
$f_Z(\cdot)$	:	Steady-state p.d.f. of $Z(t)$
$f_{Y,Q_0}(\cdot,\cdot)$	:	Joint p.d.f. of Y and Q_0
$f_{Y,Q_0}(\cdot,\cdot)$	:	Joint p.d.f. of Y_d and Q_0

$f_{Y^{(n)},Q_0^{(n)}}(\cdot,\cdot)$	:	Joint density of $Y^{(n)}$ and $Q_0^{(n)}$
$f_{Y_d^{(n)},Q_0^{(n)}}(\cdot,\cdot)$	:	Joint density of $Y_d^{(n)}$ and $Q_0^{(n)}$
$f_{Te}(x,t,q,\lambda)$	:	P.d.f. of a Truncated Erlang random variable with parameters q and λ at t
$f(x,k,\lambda)$	:	P.d.f. of Erlang random variable with parameters k and λ
$F(x,k,\lambda)$	:	C.d.f. of Erlang random variable with parameters k and λ
$F_{Q_0}(\cdot)$	:	C.d.f. of Q_0
$F_{Te}(x,t,q,\lambda)$	:	C.d.f. of a Truncated Erlang random variable with parameters q and λ at t
$F_{Te}^{(n)}(x,t,q,\lambda)$	:	C.d.f. of n^{th} convolution of a Truncated Erlang random variable with parameters q and λ at t
$F_{W_0(q)}(\cdot)$	:	Steady-state c.d.f. of $W_0(q)$
$F_Y(\cdot)$	:	C.d.f. of Y
$\overline{F}$	:	$1-F$ for any distribution function F
$F_{Y^{(n)},Q_0^{(n)}}(\cdot,\cdot)$	:	Sub-distribution function of $Y^{(n)}$ and $Q_0^{(n)}$
$F_{Y_d^{(n)},Q_0^{(n)}}(\cdot,\cdot)$	:	Sub-distribution function of $Y_d^{(n)}$ and $Q_0^{(n)}$
$\overline{F}$	:	$1-F$ for any distribution function F
ϕ_0	:	$p_0(0,\lambda_0 T)$, probability that no demands arrive in $(0,T]$
$\phi(t,\mu,\sigma^2)$	:	C.d.f. of a Normal random variable with mean μ and variance σ^2
$g(t,\cdot,\cdot,\cdot)$	:	P.d.f. of $\xi_i(t)$
$g(\cdot,\cdot,\cdot)$	:	Steady-state p.d.f. of $\xi_i(t)$
$g(\cdot,\cdot)$	:	Steady-state p.d.f. of $\xi(t)$
γ_i	:	Modified fill rate for retailer i
$\overline{\gamma}_i$	:	Target modified fill rate for retailer i
h_0	:	Unit inventory holding cost per time at the warehouse
h_i	:	Unit inventory holding cost per time of item i or retailer i
$\mathcal{H}_T$	:	Probability that an order is given by time trigger in $(Q,\mathbf{S},T)$ policy
$\eta(q)$	:	Minimum number of retailer orders for which the order size exceeds q units
$I(\cdot)$	:	Indicator function of its argument

$IL_0(t)$	:	Inventory level of the warehouse at time t
$IP(t)$	:	$\sum_{i=1}^{N} IP_i(t)$, total inventory position of the system at time t
$IP_0(t)$	:	Inventory position of the warehouse at time t
$IP_i(t)$	:	Inventory position of item i or retailer i at time t
K	:	Common ordering costs of items or retailers
K_0	:	Warehouse fixed ordering cost
k_i	:	Item i or retailer i specific ordering cost
$\kappa(q)$	:	Minimum total order size such that it exceeds q units
λ_0	:	$\sum_{i=1}^{N} \lambda_i$, total demand rate of items or retailers
λ_i	:	Demand rate for item i or retailer i
λ_Θ	:	$\sum_{i\in\Theta} \lambda_i$, demand rate for the items in set Θ
L_0	:	Lead time of the warehouse
L_i	:	Lead time of either item i or retailer i
μ_t	:	$E[D_0(0,t)]$
μ_{Te}	:	Mean of a truncated Erlang random variable
$\mu_{i,j}$	:	Mean of the order quantity of retailer j in an order triggered by retailer i
N	:	Number of items or retailers
$\mathcal{N}$	:	Set of all the items in the inventory system
n_{it}	:	Iteration number
$N_{0,i}(t)$	:	Number of demands that have arrived for items other than item i since last decision epoch
$N_i(t)$	:	Number of demands that have arrived for item i since
	:	last decision epoch
$N(t)$	:	Counting process of system demands in $(0,t]$ item i since last decision epoch
$NI_i(t)$	:	Net inventory level of item i at time t
$OH_0(t)$	:	On-hand inventory of the warehouse at time t
$OH_0(t)$	:	Steady-state on-hand inventory of the warehouse
$OH_i(t)$	:	On-hand inventory of item i at time t
OH_i	:	Steady-state on-hand inventory of item i
$p_0(x,\lambda)$	:	P.m.f. of a Poisson random variable with rate λ
$P_0(x,\lambda)$	:	C.d.f. of a Poisson random variable with rate λ

$\tilde{p}_0(q, \lambda_\Theta z, \Theta)$	:	Probability that a total of q units are demanded for items in set Θ in z time units
$P_{Q_0}(\cdot)$	:	P.m.f. of Q_0
$P_{Q_0}^{(n)}(\cdot)$	:	P.m.f. of n^{th} convolution of Q_0
$P'_{Q_0}(\cdot)$	:	Function corresponding to p.m.f. of Q_0 for $Q_0 < Q$
$P'_{Q_0}(\cdot)$	:	n^{th} convolution of $P'_{Q_0}(\cdot)$
$P_{R_i(q)}(\cdot)$	:	P.m.f. of $R_i(q)$
p.d.f.	:	Probability density function
p.m.f.	:	Probability mass function
π_i	:	Unit shortage cost of item i
θ_i	:	Probability that item i is included in the order
Θ	:	A subset of $\mathcal{N}$
Q	:	Quantity trigger under $(Q, \mathbf{S})$, $(Q, \mathbf{S}, T)$, $(Q, \mathbf{S}\|T)$ policies
Q_0	:	Order quantity
$\underline{Q}_0$	:	Minimum order quantity under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policy
$\overline{Q}_0$	:	Maxium order quantity under $(\mathbf{s}, \mathbf{S-1}, \mathbf{S})$ policy
Q_i	:	i'th replicant of Q_0
Q_r	:	A temporary variable to construct the search region for Q
Q_w	:	Warehouse order size
Q^{min}	:	Minimum value of the search range of Q
Q^{max}	:	Maximum value of the search range of Q
Q^*	:	Optimal value of Q
$\hat{Q}$	:	Solution for Q at an iteration of an iterative search procedure
r_i	:	λ_i/λ_0, probability that a demand arrives at item i or retailer i
$R_i(q)$	:	Order quantity of retailer i in an order of size q
$r'_{j,i}$	:	$\lambda_j/(\lambda_0 - \lambda_i)$ for $j \neq i$
ρ_i	:	Time weighted shortage cost of item i
s_0	:	Reorder level of the warehouse
s_0^{min}	:	Minimum value of the search range of s_0
s_0^{max}	:	Maximum value of the search range of s_0
s_0^*	:	Optimal value of s_0
$\hat{s}_0$	:	Solution for s_0 at an iteration of an iterative search procedure

S_0	:	Order-up-to level of the warehouse
S_0^{min}	:	Minimum value of the search range of S_0
S_0^{max}	:	Maximum value of the search range of S_0
S_0^*	:	Optimal value of S_0
$\hat{S}_0$	:	Solution for S_0 at an iteration of an iterative search procedure
s_i	:	Reorder level of item i or retailer i
$\mathbf{s}$	:	$(s_1, s_2, \ldots, s_N)$
s_i^{min}	:	Minimum value of the search range of s_i
s_i^{max}	:	Maximum value of the search range of s_i
$\hat{\mathbf{s}}$	:	Solution for $\mathbf{s}$ at an iteration of an iterative search procedure
$\mathbf{s}^*$	:	Optimal value of $\mathbf{s}$
$\hat{s}_i$	:	Solution for s_i at an iteration of an iterative search procedure
S_i	:	Order-up-to level of item i or retailer i
$\mathbf{S}$	:	$(S_1, S_2, \ldots, S_N)$
S_T	:	$\sum_{i=1}^{N} S_i$, maximum inventory position of the retailer or the items
S_i^{min}	:	Minimum value of the search range of S_i
S_i^{max}	:	Maximum value of the search range of S_i
$\hat{\mathbf{S}}$	:	Solution for $\mathbf{S}$ at an iteration of an iterative search procedure
$\mathbf{S}^*$	:	Optimal value of $\mathbf{S}$
σ_{Te}^2	:	Variance of a truncated Erlang random variable
$\sigma_{i,j}^2$	:	Variance of the order quantity of retailer j in an order triggered by retailer i
T	:	Time trigger under $(Q, \mathbf{S}, T)$, $(Q, \mathbf{S}\|T)$ policies
T^{min}	:	Minimum value of the search range of T
T^{max}	:	Maximum value of the search range of T
T^*	:	Optimal value of T
$\hat{T}$	:	Solution for T at an iteration of an iterative search procedure
$T_i(q)$	:	Effective lead time of an order of size q for retailer i
$v_i(\cdot)$	:	P.m.f. of demand size for item i
$V_i(\cdot)$	:	C.d.f. of demand size for item i
$v_i^{(k)}(\cdot)$	:	P.m.f. of k^{th} convolution of demand size for item i
$\varphi_i(\cdot)$	:	Steady-state p.m.f. of $IP_i(t)$
$\varphi(t, k)$	:	P.m.f. of $D_0(0, t]$

ϑ_i	:	Steady-state p.m.f. of $OH_0(t)$
$w_\Theta(q,k)$	:	Probability that k customers demand a total of q units
	:	for the items in the set Θ
W_0	:	Steady-state waiting time of an order
$W_0(t,q)$	:	Waiting time of an order of size q arriving at time t
$W_0(q)$	:	Steady-state waiting time of an order of size q
$\hat{X}$	:	Approximation for quantity X
X^p	:	Value of a policy parameter X in the previous iteration
X_j^i	:	Arrival time of j'th demand a retailer i after an order
X_n	:	Arrival time of n'th system demand after the last decision epoch
$X^{(n)}$	:	n'th convolution of a random variable X
$\xi(t)$	:	$\{IP_0(t), Z(t)\}$
$\xi_i(t)$	:	$\{N_i(t), N_{0,i}(t), Z(t)\}$
Y	:	Inter-order time
Y_0	:	Time between consecutive warehouse orders
Y_d	:	Time since last decision epoch until an order is given
Y_i	:	i'th replicant of Y
ψ_i	:	Steady-state no-stockout probability of item i
$\overline{\psi}_i$	:	Target no-stockout probability of item i
$Z(t)$	:	Time elapsed since last decision epoch for the item or last
		order arrival at the warehouse
ζ_q	:	q/λ_0
$[x]^+$	:	$\max(x,0)$
$\lceil x \rceil$	:	Smallest integer larger than or equal to x
$\lfloor x \rfloor$	:	Largest integer smaller than or equal to x
$\lfloor x \rfloor_k$	:	Smallest integer larger than or equal to x which is divisible by k

Proof of Lemma 3.2.1:

Let $N(t)$ be the counting process of system demands in $(0,t]$ where $t=0$ is taken as the beginning of a replenishment cycle.

First, suppose $y = mT$ for $m \geq 1$ and let $0 < q < Q$. This case corresponds to a replenishment cycle depicted in Figure 3.1(a).

$$\begin{aligned} P(Y = mT, Q_0 = q) &= P(N((m-1)T) = 0, N(mT) - N((m-1)T) = q) \\ &= P(N(T) = 0)^{m-1} P(N(y-(m-1)T)) = q) = \phi_0^{m-1} p_0(q, \lambda_0(y-(m-1)T)) \\ & \qquad m \geq 1, 0 < q < Q \end{aligned}$$

Now, suppose $(m-1)T < y < mT$ for $m \geq 1$ and $q = Q$. This case corresponds to a replenishment cycle depicted in Figure 3.1(b). An order of size $Q_0 = Q$ is triggered in time interval $(y, y+\delta y]$ if the following events occur: $N((m-1)T) = 0, N(y) - N(y-(m-1)T) = Q-1$, and $N(y+\delta y) - N(y) = 1$. Then, we have

$$\begin{aligned} & P(Y \in (y, y+\delta y], Q_0 = Q) = \\ & \quad = P(N((m-1)T) = 0, N(y) - N((m-1)T) = Q-1, N(y+\delta y) - N(y) = 1) \\ & \quad = \phi_0^{m-1} p_0(Q-1, \lambda_0(y-(m-1)T))\lambda_0[\delta y + o(\delta y)] \end{aligned}$$

Result follows by dividing both sides by δy and taking the limit as $\delta y \to 0$.

Proof of Corollary 3.2.1:

a) First, let $Q_0 = q \in [1, 2, \ldots, Q-1]$. Then,

$$P_{Q_0}(q) = \sum_{m=1}^{\infty} \phi_0^{m-1} p_0(q, \lambda_0 T) = p_0(q, \lambda_0 T)/(1-\phi_0)$$

Now, let $Q_0 = Q$. Then, we have

$$\begin{aligned} P_{Q_0}(Q) &= \sum_{m=1}^{\infty} \phi_0^{m-1} \int_{y=(m-1)T}^{mT} f(y-(m-1)T, Q, \lambda_0) dy \\ &= \sum_{m=1}^{\infty} \phi_0^{m-1} \int_{y=0}^{T} f(y, Q, \lambda_0) dy = F(T, Q, \lambda_0)/(1-\phi_0) \\ &= \overline{P}_0(Q-1, \lambda_0 T)/(1-\phi_0) \end{aligned}$$

b) For $y = mT, m \geq 1$, we can write:

$$\begin{aligned} f_Y(y) &= \sum_{q=1}^{Q-1} \phi_0^{m-1} p_0(q, \lambda_0 T) = \phi_0^{m-1} \left[P_0(Q-1, \lambda_0 T) - P_0(0, \lambda_0 T) \right] \\ &= \phi_0^{m-1} [P_0(Q-1, \lambda_0 T) - \phi_0] \end{aligned}$$

Now, let $y \in ((m-1)T, mT)$ for $m \geq 1$. It is obvious that

$$f_Y(y) = f_{Y,Q_0}(y, Q)$$

Proof of Proposition 3.2.1:

The proof is based on the development of the partial differential equations describing the dynamics of the stochastic process, $\xi_i(t)$ via supplementary variables. We refer the reader to Cox [27] and Schmidt and Nahmias [64] for further details of the technique.

Let $g_i(t, n_i, n_0, z)$ denote the probability density function of $\xi_i(t)$ being in state $\{n_i, n_0, z\}$ at time t. We first derive the partial differential equations that $g_i(t, n_i, n_0, z)$ satisfies and use them to obtain the partial differential equations for the steady-state distribution, $g_i(n_i, n_0, z)$. We derive the equations for four different cases and then we verify that the proposed solution (3.1) satisfies these equations.

Case 1: $n_0 = 0, n_i = 0, 0 < z < T$.

$$g_i(t + \delta_t, 0, 0, z + \delta t) = g_i(t, 0, 0, z)(1 - \lambda_0 \delta t) + o(\delta t)$$

where $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$. This follows because the state of item i will be $(0, 0, z + \delta t)$ at time $t + \delta t$ if it is in state $(0, 0, z)$ at time t and no demands arrive for any of the items during the interval $(t, t + \delta t]$ which has probability $1 - \lambda_0 \delta t + o(\delta t)$. For sufficiently small δt, $z + \delta t < T$ should also hold so that a review is not carried out.

Subtracting the term $g_i(t, 0, 0, z + \delta t)$ from both sides and dividing both sides by δt and letting $\delta t \to 0$ gives

$$\frac{\partial g_i(t, 0, 0, z)}{\partial t} = -\frac{\partial g_i(t, 0, 0, z)}{\partial z} + \lambda_0 g_i(t, 0, 0, z)$$

Taking the limit as $t \to \infty$ results in

$$\frac{\partial g_i(0,0,z)}{\partial z} = -\lambda_0 g_i(0,0,z) \quad 0 < z < T \tag{9.1}$$

$$\frac{\partial g_i(0,0,z)}{\partial z} = -C_0\lambda_0 e^{-\lambda_0 z} = -\lambda_0 g_i(0,0,z)$$

Case 2: $n_0 = 0, 0 < n_i < Q$ and $0 < z < T$

$$g_i(t+\delta_t, n_i, 0, z+\delta t) = g_i(t, n_i, 0, z)(1-\lambda_0\delta t) + g_i(t, n_i - 1, 0, z)\lambda_i \delta t + o(\delta t)$$

The state of item i will be $(n_i, 0, z+\delta t)$ at time $t+\delta t$ if at time t the state is $(n_i, 0, z)$ and no demands have arrived in $[t, t+\delta t)$; or the state at time t is $(n_i - 1, 0, z)$ and a demand has arrived for item i during the interval $[t, t+\delta t)$ with probability $\lambda_i \delta t + o(\delta t)$. Subtracting the term $g_i(t, n_i, 0, z+\delta t)$ from both sides and dividing both sides by δt and letting $\delta t \to 0$ results in

$$\frac{\partial g_i(t, n_i, 0, z)}{\partial t} = -\frac{\partial g_i(t, n_i, 0, z)}{\partial z} - \lambda_0 g_i(t, n_i, 0, z) + \lambda_i g_i(t, n_i - 1, 0, z)$$

Then, letting $t \to \infty$,

$$\frac{\partial g_i(n_i, 0, z)}{\partial z} = -\lambda_0 g_i(n_i, 0, z) + \lambda_i g_i(n_i - 1, 0, z) \quad 0 < n_i < Q, z < T \tag{9.2}$$

$$\begin{aligned} g_i(n_i, 0, z) &= C_0 e^{-\lambda_0 z}\frac{(\lambda_i z)^{n_i}}{n_i!} \\ \frac{\partial g_i(n_i, 0, z)}{\partial z} &= -\lambda_0 C_0 e^{-\lambda_0 z}\frac{(\lambda_i z)^{n_i}}{n_i!} + C_0 e^{-\lambda_0 z}\lambda_i \frac{(\lambda_i z)^{n_i - 1}}{(n_i - 1)!} \\ &= -\lambda_0 g_i(n_i, 0, z) + \lambda_i g_i(n_i - 1, 0, z) \end{aligned}$$

Case 3: $0 < n_0 < Q, n_i = 0$ and $0 < z < T$

$$g_i(t+\delta_t, 0, n_0, z+\delta t) = g_i(t, 0, n_0, z)(1-\lambda_0\delta t) + g_i(t, 0, n_0 - 1, z)(\lambda_0 - \lambda_i)\delta t + o(\delta t)$$

This case is very similar to Case 2 and no further details will be given except the following partial differential equation for the steady-state probability density function:

$$\frac{\partial g_i(0,n_0,z)}{\partial z} = -\lambda_0 g_i(0,n_0,z) + (\lambda_0-\lambda_i)g_i(0,n_0-1,z) \quad 0<n_0<Q, z<T \tag{9.3}$$

Case 3: $0<n_0<Q, n_i=0, 0<z<T$

$$\begin{aligned} g_i(0,n_0,z) &= C_0 e^{-\lambda_0 z}\frac{((\lambda_0-\lambda_i)z)^{n_0}}{n_0!} \\ \frac{\partial g_i(0,n_0,z)}{\partial z} &= -\lambda_0 C_0 e^{-\lambda_0 z}\frac{((\lambda_0-\lambda_i)z)^{n_0}}{n_0!} + C_0 e^{-\lambda_0 z}(\lambda_0-\lambda_i)\frac{((\lambda_0-\lambda_i)z)^{n_0-1}}{(n_0-1)!} \\ &= -\lambda_0 g_i(0,n_0,z) + (\lambda_0-\lambda_i)g_i(0,n_0-1,z) \end{aligned}$$

Case 4: $0<n_0<Q, 0<n_i<Q, 0<n_i+n_0<Q$ and $0<z<T$

$$\begin{aligned} g_i(t+\delta_t,n_i,n_0,z+\delta t) &= g_i(t,n_i,n_0,z)(1-\lambda_0\delta t) + g_i(t,n_i-1,n_0,z)\lambda_i\delta t \\ &+ g_i(t,n_i-1,n_0,z)(\lambda_0-\lambda_i)\delta t + o(\delta t) \end{aligned}$$

This follows because the state at time $t+\delta t$ will be $(n_i,n_0,z+\delta t)$ only if one of the following three events occur: the state at time t is (n_i,n_0,z) and no demands have arrived at the system in $[t,t+\delta t)$; the state at time t is (n_i-1,n_0,z) and a demand has arrived for item i in $[t,t+\delta t)$; the state is (n_i,n_0-1,z) at time t and a demand has arrived for an item other than i in $[t,t+\delta t)$.

Subtracting the term $g_i(t,n_i,n_0,z+\delta t)$ from both sides, dividing by δt and letting $\delta t\to 0$ results in

$$\begin{aligned} \frac{\partial g_i(t,n_i,n_0,z)}{\partial t} &= -\frac{\partial g_i(t,n_i,n_0,z)}{\partial z} - \lambda_0 g_i(t,n_i,n_0,z) \\ &+ \lambda_i g_i(t,n_i-1,n_0,z) + (\lambda_0-\lambda_i)g_i(t,n_i,n_0-1,z) \end{aligned}$$

Finally, letting $t\to\infty$, we have

$$\begin{aligned} \frac{\partial g_i(n_i,n_0,z)}{\partial z} &= -\lambda_0 g_i(n_i,n_0,z) + \lambda_i g_i(n_i-1,n_0,z) + (\lambda_0-\lambda_i)g_i(n_i,n_0-1,z) \\ & \qquad 0<n_i<Q, 0<n_0<Q, 0<n_i+n_0<Q, 0<z<T \end{aligned} \tag{9.4}$$

$$\begin{aligned}
g_i(n_i, n_0, z) &= C_0 e^{-\lambda_0 z} \frac{(\lambda_i z)^{n_i}}{n_i!} \frac{((\lambda_0-\lambda_i)z)^{n_0}}{n_0!} \\
\frac{\partial g_i(n_i, n_0, z)}{\partial z} &= -\lambda_0 C_0 e^{-\lambda_0 z} \frac{(\lambda_i z)^{n_i}}{n_i!} \frac{((\lambda_0-\lambda_i)z)^{n_0}}{n_0!} \\
&+ \lambda_i C_0 e^{-\lambda_0 z} \frac{(\lambda_i z)^{n_i-1}}{(n_i-1)!} + (\lambda_0-\lambda_i) C_0 e^{-\lambda_0 z} \frac{((\lambda_0-\lambda_i)z)^{n_0-1}}{(n_0-1)!} \\
&= -\lambda_0 g_i(n_i, n_0, z) + \lambda_i g_i(n_i-1, n_0, z) + (\lambda_0-\lambda_i) g_i(n_i, n_0-1, z)
\end{aligned}$$

Boundary Condition: The state of the system at time $t+\delta t$ will be $(0,0,0)$ if one of the following two events occur: the state at time t is $(n_i, Q-1-n_i, z)$ and a demand has arrived at the system in $[t, t+\delta t)$ which means that Q demands accumulate at the system triggering an order; and the state at time t is $(n_i, n_0, T-\delta t)$ and no demands have arrived at the system in $[t, t+\delta t)$ which indicates an order is placed by the time trigger. Then, we can write:

$$\begin{aligned}
g(t+\delta t, 0,0,0) &= \sum_{n_i=0}^{Q-1} \int_{z=0}^{T} \lambda_0 g_i(t, n_i, Q-1-n_i, z) dz \\
&+ \sum_{n_i=0}^{Q-1} \sum_{n_0=0}^{Q-1-n_i} g_i(t, n_i, n_0, T-\delta t)(1-\lambda_0 \delta t)
\end{aligned}$$

Let $t \to 0$ and $\delta t \to 0$. Then, the boundary condition to the system of partial differential equations described above is as follows:

$$g_i(0,0,0) = \sum_{n_i=0}^{Q-1} \int_{z=0}^{T} \lambda_0 g_i(n_i, Q-1-n_i, z) dz + \sum_{n_i=0}^{Q-1} \sum_{n_0=0}^{Q-1-n_i} g_i(n_i, n_0, T) \tag{9.5}$$

We now verify that the proposed solution (3.1) satisfies the boundary condition given in Equation (9.5).

$$\begin{aligned}
g_i(0,0,0) &= \sum_{n_i=0}^{Q-1} \int_{z=0}^{T} C_0 \lambda_0 \frac{e^{-\lambda_i z}(\lambda_i z)^{n_i}}{n_i!} \frac{e^{-(\lambda_0-\lambda_i)z}((\lambda_0-\lambda_i)z)^{Q-1-n_i}}{(Q-1-n_i)!} dz \\
&+ \sum_{n_i=0}^{Q-1} \sum_{n_0=0}^{Q-1-n_i} C_0 p_0(n_i, \lambda_i T) p_0(n_0, (\lambda_0-\lambda_i)T) \\
&= C_0 \sum_{n_i=0}^{Q-1} \frac{(Q-1)!}{n_i!(Q-1-n_i)!} \left(\frac{\lambda_i}{\lambda_0}\right)^{n_i} \left(\frac{\lambda_0-\lambda_i}{\lambda_0}\right)^{Q-1-n_i} \int_{z=0}^{T} \lambda_0 \frac{e^{-\lambda_0 z}(\lambda_0 z)^{Q-1}}{(Q-1)!} dz
\end{aligned}$$

$$\begin{aligned}
&+ \quad C_0 \sum_{n_i=0}^{Q-1} p_0(n_i, \lambda_i T) P_0(Q-1-n_i, (\lambda_0 - \lambda_i)T) \\
&= \quad C_0 \sum_{n_i=0}^{Q-1} \binom{Q-1}{n_i} r_i^{n_i}(1-r_i)^{Q-1-n_i} F(T, Q, \lambda_0) + C_0 \lim_{z \to T} P_0(Q-1, \lambda_0 z) \\
&= \quad C_0 F(T, Q, \lambda_0) + C_0 P_0(Q-1, \lambda_0 T) = C_0
\end{aligned}$$

Thus, the steady-state probability density function has the structure given in Equation in (3.1). Moreover, for $g_i(n_i, n_0, z)$ to be a probability density function,

$$\int_{z=0}^{T} \left[\sum_{n_i=0}^{Q-1} \sum_{n_0=0}^{Q-1-n_i} g_i(n_i, n_0, z) \right] dz = 1$$

Therefore,

$$\int_{z=0}^{T} \left[\sum_{n_i=0}^{Q-1} C_0 p_0(n_i, \lambda_i z) P_0(Q-1-n_i, (\lambda_0 - \lambda_i) z) \right] dz = \int_{z=0}^{T} C_0 P_0(Q-1, \lambda_0 z) dz = 1$$

Hence,

$$C_0 = \left[\int_{z=0}^{T} P_0(Q-1, \lambda_0 z) dz \right]^{-1}$$

Derivation of Equation 3.2:

Using Corollary 3.2.1 we can write:

$$\begin{aligned}
E[Y] &= \sum_{m=1}^{\infty} mT\phi_0^{m-1} \left[P_0(Q-1, \lambda_0 T) - \phi_0 \right] \\
&+ \sum_{m=1}^{\infty} \int_{y=(m-1)T}^{mT} \phi_0^{m-1} y f(y-(m-1)T, Q, \lambda_0) dy \\
&= \frac{T\left[P_0(Q-1, \lambda_0 T) - \phi_0 \right]}{(1-\phi_0)^2} + \sum_{m=1}^{\infty} \phi_0^{m-1} \int_{t=0}^{T} (t+(m-1)T) f(t, Q, \lambda_0) dt \\
&= \frac{T[P_0(Q-1, \lambda_0 T) - \phi_0]}{(1-\phi_0)^2} + \frac{1}{1-\phi_0} \int_{t=0}^{T} t f(t, Q, \lambda_0) dt \\
&+ \sum_{m=1}^{\infty} (m-1)\phi_0^{m-1} T F(T, Q, \lambda_0) \\
&= \frac{T P_0(Q-1, \lambda_0 T) - T\phi_0}{(1-\phi_0)^2} + \frac{Q}{\lambda_0(1-\phi_0)} \int_{t=0}^{T} f(t, Q+1, \lambda_0) dt
\end{aligned}$$

$$+\sum_{m=1}^{\infty} m\phi_0^{m-1} TF(T,Q,\lambda_0) - \sum_{m=1}^{\infty} \phi_0^{m-1} TF(T,Q,\lambda_0)$$
$$= \frac{TP_0(Q-1,\lambda_0 T) - T\phi_0}{(1-\phi_0)^2} + \frac{Q}{\lambda_0(1-\phi_0)} F(T,Q+1,\lambda_0)$$
$$+\frac{T}{(1-\phi_0)^2} F(T,Q,\lambda_0) - \frac{T}{1-\phi_0} F(T,Q,\lambda_0)$$
$$= \frac{T}{1-\phi_0} - \frac{T\overline{P}_0(Q-1,\lambda_0 T)}{1-\phi_0} + \frac{Q}{\lambda_0(1-\phi_0)}\overline{P}_0(Q,\lambda_0 T)$$
$$= \frac{TP_0(Q-1,\lambda_0 T)}{1-\phi_0} + \frac{Q\overline{P}_0(Q,\lambda_0 T)}{\lambda_0(1-\phi_0)}$$

Proof of Lemma 3.4.1:

Recall that $N(t)$ is the counting process of system demands in $(0,t]$. Then, we have, $P(N(T)=0) = p_0(0,\lambda_0 T)$ as in unit Poisson demands and $P(N(t)=k) = \tilde{p}_0(k,\lambda_{\mathcal{N}} z,\mathcal{N})$ for $k \geq 1$.

Case 1: $y = mT$ for $m \geq 1$ and $0 < q < Q$.

$$P(Y = mT, Q_0 = q) = P(N((m-1)T) = 0, N(mT) - N((m-1)T) = q)$$
$$= P(N(T)=0)^{m-1} P(N(T)=q) = \phi_0^{m-1}\tilde{p}_0(q,\lambda_{\mathcal{N}} T,\mathcal{N}) \quad m \geq 1, 0 < q < Q$$

Case 2: $y \in ((m-1)T, mT)$ for $m \geq 1$ and $q \geq Q$.

$$P(Y \in [y, y+\delta y), Q_0 = q)$$
$$= \sum_{j=0}^{Q-1} P(N((m-1)T) = 0, N(y) - N((m-1)T) = j, N(y+\delta y) - N(y) = q-j)$$
$$= \phi_0^{m-1} \sum_{j=0}^{Q-1} \tilde{p}_0(j,\lambda_{\mathcal{N}} T,\mathcal{N}) \left[\sum_{i=1}^{N} \lambda_i(\delta y + o(\delta y)) v_i(q-j)\right]$$
$$= \phi_0^{m-1} \sum_{j=0}^{Q-1} \sum_{k=0}^{j} p_0(k,\lambda_0(y-(m-1)T)) w_{\mathcal{N}}(j,k) \left[\sum_{i=1}^{N} \lambda_i(\delta y + o(\delta y)) v_i(q-j)\right]$$
$$= \phi_0^{m-1} \sum_{j=0}^{Q-1} \sum_{k=0}^{j} \lambda_0 p_0(k,\lambda_0 y-(m-1)T)) w_{\mathcal{N}}(j,k) \left[\sum_{i=1}^{N} r_i(\delta y + o(\delta y)) v_i(q-j)\right]$$
$$= \phi_0^{m-1} \sum_{k=0}^{Q-1} \sum_{j=k}^{Q-1} f(y-(m-1)T), k+1,\lambda_0) w_{\mathcal{N}}(j,k) \left[\sum_{i=1}^{N} r_i(\delta y + o(\delta y)) v_i(q-j)\right]$$

Result follows by dividing both sides by δy and taking the limit as $\delta y \to 0$.

Proof of Lemma 3.4.1:

Similar to unit Poisson demands, we will prove that the proposed solution satisfies the differential equations for the steady-state distribution, $g_i(n_i, n_0, z)$.

Since the derivations of these differential equations are quite similar to those in unit Poisson demand, we just present the differential equations for $g_i(n_i, n_0, z)$ without giving the derivations. Then, for each case we show that the proposed solution satisfies the differential equations.

Case 1: $n_0 = 0, n_i = 0, 0 < z < T$.

$$\frac{\partial g_i(0,0,z)}{\partial z} = -\lambda_0 g_i(0,0,z) \tag{9.6}$$

$$\begin{aligned} g_i(0,0,z) &= C_1 p_0(0, \lambda_i z) p_0(0, (\lambda_0 - \lambda_i)z) = C_1 e^{-\lambda_0 z} \\ \frac{\partial g_i(0,0,z)}{\partial z} &= -\lambda_0 C_1 e^{-\lambda_0 z} = -\lambda_0 g_i(0,0,z) \end{aligned} \tag{9.7}$$

Case 2: $n_0 = 0, 0 < n_i < Q$ and $0 < z < T$

$$\frac{\partial g_i(n_i,0,z)}{\partial z} = -\lambda_0 g_i(n_i,0,z) + \lambda_i \sum_{k=1}^{n_i} g_i(n_i - k, 0, z) v_i(k) \qquad 0 < n_i < Q, z < T \tag{9.8}$$

$$\begin{aligned} g_i(n_i,0,z) &= C_1 \tilde{p}_0(n_i, \lambda_{\{i\}} z, \lambda_{\{i\}}) p_0(0, (\lambda_0 - \lambda_i)z) \\ &= C_1 \sum_{k=1}^{n_i} p_0(k, \lambda_i z) v_i^{(k)}(n_i) p_0(0, (\lambda_0 - \lambda_i)z) = C_1 \sum_{k=1}^{n_i} \frac{e^{-\lambda_0 z}(\lambda_i z)^k}{k!} v_i^{(k)}(n_i) \end{aligned} \tag{9.9}$$

$$\begin{aligned} \frac{\partial g_i(n_i,0,z)}{\partial z} &= -\lambda_0 C_1 \sum_{k=1}^{n_i} \frac{e^{-\lambda_0 z}(\lambda_i z)^k}{k!} v_i^{(k)}(n_i) + C_1 \sum_{k=1}^{n_i} \frac{\lambda_i e^{-\lambda_0 z}(\lambda_i z)^{k-1}}{(k-1)!} v_i^{(k)}(n_i) \\ &= -\lambda_0 C_1 g_i(n_i,0,z) + C_1 \sum_{k=0}^{n_i-1} \frac{\lambda_i e^{-\lambda_0 z}(\lambda_i z)^k}{k!} v_i^{(k+1)}(n_i) \end{aligned}$$

Next, we should show that

$$\lambda_i \sum_{k=1}^{n_i} g_i(n_i - k, 0, z) v_i(k) = \sum_{k=0}^{n_i - 1} \frac{\lambda_i e^{-\lambda_0 z} (\lambda_i z)^k}{k!} v_i^{(k+1)}(n_i) \tag{9.10}$$

Using the proposed solution given in Equations (9.7) and (9.9), we can rewrite the right hand side of Equation (9.10) as follows:

$$\begin{aligned}
&\lambda_i \sum_{k=1}^{n_i} g_i(n_i - k, 0, z) v_i(k) \\
&= \lambda_i C_1 \sum_{k=1}^{n_i - 1} \sum_{j=1}^{n_i - k} \frac{\lambda_i e^{-\lambda_0 z} (\lambda_i z)^j}{j!} v_i^{(j)}(n_i - k) v_i(k) + \lambda_i C_1 e^{-\lambda_0 z} v_i(n_i) \\
&= \lambda_i C_1 \sum_{j=1}^{n_i - 1} \sum_{k=1}^{n_i - j} \frac{\lambda_i e^{-\lambda_0 z} (\lambda_i z)^j}{j!} v_i^{(j)}(n_i - k) v_i(k) + \lambda_i C_1 e^{-\lambda_0 z} v_i(n_i) \\
&= \lambda_i C_1 \sum_{j=0}^{n_i - 1} \frac{\lambda_i e^{-\lambda_0 z} (\lambda_i z)^j}{j!} \left[\sum_{k=1}^{n_i - j} v_i^{(j)}(n_i - k) v_i(k) \right] \\
&= \lambda_i C_1 \sum_{j=0}^{n_i - 1} \frac{\lambda_i e^{-\lambda_0 z} (\lambda_i z)^j}{j!} v_i^{(j+1)}(n_i)
\end{aligned}$$

Case 3: $0 < n_0 < Q, n_i = 0$ and $0 < z < T$

$$\frac{\partial g_i(0, n_0, z)}{\partial z} = -\lambda_0 g_i(0, n_0, z) + \sum_{k=1}^{n_0} g_i(0, n_0 - k, z) \left[\sum_{j \neq i} \lambda_j v_j(k) \right] \qquad 0 < n_0 < Q, z < T \tag{9.11}$$

$$\begin{aligned}
g_i(0, n_0, z) &= C_1 p_0(0, \lambda_i z) \tilde{p}_0(n_0, \lambda_{\{\mathcal{N} \setminus \{i\}\}} z, \{\mathcal{N} \setminus \{i\}\}) \\
&= C_1 \sum_{k=1}^{n_0} \frac{e^{-\lambda_0 z} ((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N} \setminus \{i\}}(n_0, k)
\end{aligned} \tag{9.12}$$

$$\begin{aligned}
\frac{\partial g_i(0, n_0, z)}{\partial z} &= -\lambda_0 C_1 \sum_{k=1}^{n_0} \frac{e^{-\lambda_0 z} ((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N} \setminus \{i\}}(n_0, k) \\
&+ (\lambda_0 - \lambda_i) C_1 \sum_{k=1}^{n_0} \frac{e^{-\lambda_0 z} ((\lambda_0 - \lambda_i) z)^{k-1}}{(k-1)!} w_{\mathcal{N} \setminus \{i\}}(n_0, k) \\
&= -\lambda_0 g_i(0, n_0, z) \\
&+ (\lambda_0 - \lambda_i) C_1 \sum_{k=0}^{n_0 - 1} \frac{e^{-\lambda_0 z} ((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N} \setminus \{i\}}(n_0, k+1)
\end{aligned}$$

The next step will be to prove that

$$\sum_{k=1}^{n_0} g_i(0, n_0-k, z)\left[\sum_{j\neq i}\lambda_j v_j(k)\right] = (\lambda_0-\lambda_i)C_1\sum_{k=0}^{n_0-1}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^{k-1}}{(k-1)!}w_{\mathcal{N}\setminus\{i\}}(n_0, k+1) \tag{9.13}$$

The proposed solutions in Equations (9.7) and (9.12) give

$$\begin{aligned}
&\sum_{k=1}^{n_0} g_i(0, n_0-k, z)\left[\sum_{j\neq i}\lambda_j v_j(k)\right]\\
&= C_1\sum_{k=1}^{n_0-1}\sum_{m=1}^{n_0-k}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^m}{m!}w_{\mathcal{N}\setminus\{i\}}(n_0-k, m)\left[\sum_{j\neq i}\lambda_j v_j(k)\right]\\
&+C_1 e^{-\lambda_0 z}\left[\sum_{j\neq i}\lambda_j v_j(n_0)\right]\\
&= C_1\sum_{m=1}^{n_0-1}\sum_{k=1}^{n_0-m}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^m}{m!}w_{\mathcal{N}\setminus\{i\}}(n_0-k, m)\left[\sum_{j\neq i}\lambda_j v_j(k)\right]\\
&+C_1 e^{-\lambda_0 z}\left[\sum_{j\neq i}\lambda_j v_j(n_0)\right]\\
&= C_1\sum_{m=0}^{n_0-1}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^m}{m!}\left\{\sum_{k=1}^{n_0-m}w_{\mathcal{N}\setminus\{i\}}(n_0-k, m)\left[\sum_{j\neq i}\lambda_j v_j(k)\right]\right\} \qquad (9.14)
\end{aligned}$$

Multiply and divide the right hand side of Equation (9.14) by $(\lambda_0-\lambda_i)/\lambda_0$. Then,

$$\begin{aligned}
&\sum_{k=1}^{n_0} g_i(0, n_0-k, z)\left[\sum_{j\neq i}\lambda_j v_j(k)\right]\\
&= (\lambda_0-\lambda_i)C_1\sum_{m=0}^{n_0-1}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^m}{m!}\left\{\frac{\lambda_0}{\lambda_0-\lambda_i}\sum_{k=1}^{n_0-m}w_{\mathcal{N}\setminus\{i\}}(n_0-k, m)\left[\sum_{j\neq i}r_j v_j(k)\right]\right\}\\
&= (\lambda_0-\lambda_i)C_1\sum_{m=0}^{n_0-1}\frac{e^{-\lambda_0 z}((\lambda_0-\lambda_i)z)^m}{m!}w_{\mathcal{N}\setminus\{i\}}(n_0, m+1) \qquad (9.15)
\end{aligned}$$

Case 4: $0 < n_0 < Q, 0 < n_i < Q, 0 < n_i + n_0 < Q$ and $0 < z < T$

$$\frac{\partial g_i(n_i, n_0, z)}{\partial z} = -\lambda_0 g_i(n_i, n_0, z) + \sum_{k=1}^{n_i} g_i(n_i-k, n_0, z)\lambda_i v_i(k)$$

$$+ \sum_{k=1}^{n_0} g_i(n_i, n_0 - k, z) \left[\sum_{j \neq i} \lambda_j v_j(k) \right]$$
$$0 < n_i < Q, 0 < n_0 < Q, 0 < n_i + n_0 < Q, 0 < z < T \quad (9.16)$$

$$\begin{aligned} g_i(n_i, n_0, z) &= C_1 \left[\sum_{j=1}^{n_i} p_0(j, \lambda_i z) v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} p_0(k, (\lambda_0 - \lambda_i) z) w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right] \\ &= C_1 e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^j}{j!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right] \end{aligned}$$

$$\frac{\partial g_i(n_i, n_0, z)}{\partial z} = -\lambda_0 C_1 e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^j}{j!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right]$$
$$= C_1 \lambda_i e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^{j-1}}{(j-1)!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right]$$
$$+ C_1 (\lambda_0 - \lambda_i) e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^j}{j!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^{k-1}}{(k-1)!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right]$$
$$= -\lambda_0 g_i(n_i, n_0, z) + C_1 \lambda_i e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^{j-1}}{(j-1)!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^k}{k!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right]$$
$$+ C_1 (\lambda_0 - \lambda_i) e^{-\lambda_0 z} \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^j}{j!} v_i^{(j)}(n_i) \right] \left[\sum_{k=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^{k-1}}{(k-1)!} w_{\mathcal{N}\setminus\{i\}}(n_0, k) \right]$$

Observe that Equation (9.16) can be rewritten as:

$$\begin{aligned} \frac{\partial g_i(n_i, n_0, z)}{\partial z} &= -\lambda_0 g_i(n_i, n_0, z) \\ &+ \lambda_i \sum_{k=1}^{n_i} g_i(n_i - k, 0, z) v_i(k) \left[\sum_{m=1}^{n_0} \frac{((\lambda_0 - \lambda_i) z)^m}{m!} w_{\mathcal{N}\setminus\{i\}}(n_0, m) \right] \\ &+ \sum_{k=1}^{n_0} g_i(0, n_0 - k, z) \left[\sum_{j \neq i} \lambda_j v_j(k) \right] \left[\sum_{j=1}^{n_i} \frac{(\lambda_i z)^j}{j!} v_i^{(j)}(n_i) \right] \\ &0 < n_i < Q, 0 < n_0 < Q, 0 < n_i + n_0 < Q, 0 < z < T \end{aligned}$$

Equations (9.10)-(9.13) provide the result.

Thus, the steady-state probability density function of $\xi_i(t)$ has the structure given in Equation 3.15. Moreover, for $g_i(n_i, n_0, z)$ to represent a probability density function, we should have:

$$\int_{z=0}^{T}\left[\sum_{n_0=0}^{Q-1}\sum_{n_i=0}^{Q-1-n_0} g_i(n_i, n_0, z)\right] dz = 1$$

Therefore,

$$C_1\int_{z=0}^{T}\left\{\sum_{n_0=0}^{Q-1}\sum_{n_i=0}^{Q-1-n_0}\int_{z=0}^{T}\tilde{p}_0(n_i, \lambda_{\{i\}}z, \{i\})\tilde{p}_0(n_0, \lambda_{\{\mathcal{N}\backslash\{i\}\}}z, \{\mathcal{N}\backslash\{i\}\})dz\right] = 1$$

should hold.

Proof of Lemma 5.2.1: We first derive the partial differential equations that $g(t,i,z)$ satisfies and use them to obtain the partial differential equations for the steady-state distribution, $g(i,z)$.

Observe that $\xi(t+\delta t)$ is in state $(i, z+\delta t)$ if and only if $\xi(t)$ is in state (i,z) and no order is placed at the warehouse during $[t, t+\delta t)$ (or in $[z, z+\delta t)$). Therefore, we can write:

$$\begin{aligned}
g(t+\delta t, i, z+\delta t) &= P(IP_0(t) = i, Z(t) = z, Y > z+\delta t) \\
&= P(IP_0(t) = i, Z(t) = z)P(Y > z+\delta t \,|IP_0(t) = i, Z(t) = z) \\
&= g(t,i,z)P(Y > z+\delta t \,|Y > z) \qquad (9.17)
\end{aligned}$$

where Equation (9.17) follows the independence of the inventory position of the warehouse and the interarrival time of the orders placed at the warehouse. Hence

$$g(t+\delta t, i, z+\delta t) = g(t,i,z)\frac{\overline{F}_Y(z+\delta t)}{\overline{F}_Y(z)} \qquad (9.18)$$

and

$$\overline{F}_Y(z)g(t+\delta t, i, z+\delta t) = g(t,i,z) - g(t,i,z)F_Y(z+\delta t)$$

Rearrangement of the terms after adding $F_Y(z)g(t,i,z)$ and $-\overline{F}_Y(z)g(t,i,z+\delta t)$ to both sides gives:

$$\begin{aligned}
&\overline{F}_Y(z)[g(t+\delta t, i, z+\delta t) - g(t,i,z+\delta t)] = \\
&\quad -\overline{F}_Y(z)[g(t,i,z+\delta t) - g(t,i,z)] - g(t,i,z)[F_Y(z+\delta t) - F_Y(z)]
\end{aligned}$$

Dividing both sides by δt and letting $\delta t \to 0$, we have

$$\overline{F}_Y(z)\frac{\partial g(t,i,z)}{\partial t} = -\overline{F}_Y(z)\frac{\partial g(t,i,z)}{\partial z} - g(t,i,z)f_Y(z)$$

Finally, taking the limit as $t \to \infty$ results in

$$\frac{\partial g(i,z)}{\partial z} = -g(i,z)\frac{f_Y(z)}{\overline{F}_Y(z)} \quad \text{for} \quad (i,z) \in [s_0+1, s_0+2, \ldots, S_0] \times [0,\infty) \tag{9.19}$$

The boundary conditions to the system of partial differential equations described above are given in the following two cases:

$g(S_0, 0)$: An order that has just been placed at the warehouse triggers a warehouse order itself.

$$g(S_0,0) = \sum_{j=s_0+1}^{S_0} \sum_{q=j-s_0}^{\infty} \int_{z=0}^{\infty} g(j,z)\frac{f_{Y,Q_0}(z,q)}{\overline{F}_Y(z)}dz \tag{9.20}$$

$g(i, 0)$ for $i \in [s_0+1, s_0+2, \ldots, S_0-1]$: An order that has just been placed at the warehouse does not trigger a warehouse order itself.

$$g(i,0) = \sum_{j=i+1}^{S_0} \int_{z=0}^{\infty} g(j,z)\frac{f_{Y,Q_0}(z,j-i)}{\overline{F}_Y(z)}dz \tag{9.21}$$

We now verify that the proposed solution satisfies Equation (9.19).

$$\begin{aligned} g(i,z) &= C_i\overline{F}_Y(z) \\ \frac{\partial g(i,z)}{\partial z} &= -C_i f_Y(z) = -g(i,z)\frac{f_Y(z)}{\overline{F}_Y(z)} = -C_i\overline{F}_Y(z)\frac{f_Y(z)}{\overline{F}_Y(z)} \end{aligned}$$

The boundary conditions given in (9.20) and (9.21) results in

$$\begin{aligned} g(S_0,0) &= C_{S_0} = \sum_{j=s_0+1}^{S_0} \sum_{q=j-s_0}^{\infty} \int_{z=0}^{\infty} C_j\overline{F}_Y(z)\frac{f_{Y,Q_0}(z,q)}{\overline{F}_Y(z)}dz \\ &= \sum_{j=s_0+1}^{S_0} \sum_{q=j-s_0}^{\infty} \int_{z=0}^{\infty} C_j f_{Y,Q_0}(z,q)dz = \sum_{j=s_0+1}^{S_0} \sum_{q=j-s_0}^{\infty} C_j P_{Q_0}(q) \end{aligned}$$

and for $i \in [s_0+1, s_0+2, \ldots, S_0-1]$, we have

$$\begin{aligned} g(i,0) &= C_i = \sum_{j=i+1}^{S_0} \int_{z=0}^{\infty} C_j\overline{F}_Y(z)\frac{f_{Y,Q_0}(z,j-i)}{\overline{F}_Y(z)}dz \\ &= \sum_{j=i+1}^{\infty} \int_{z=0}^{\infty} C_j f_{Y,Q_0}(z,j-i)dz = \sum_{j=i+1}^{S_0} C_j P_{Q_0}(j-i) \end{aligned}$$

For $g(i,z)$ to represent a probability density function,

$$\sum_{i=s_0+1}^{S_0} \int_{z=0}^{\infty} g(i,z)dz = 1$$

should hold. Therefore,

$$\sum_{i=s_0+1}^{S_0} \int_{z=0}^{\infty} C_i \overline{F}_Y(z)dz = \sum_{i=s_0+1}^{S_0} C_i E[Y] = 1$$

Hence,

$$\sum_{i=s_0+1}^{S_0} C_i = 1/E[Y]$$

The steady-state probability mass function of IP_0 is obtained by integrating $g(i,z)$ over z.

$$\pi_i \;=\; \int_{z=0}^{\infty} g(i,z)dz = \int_{z=0}^{\infty} C_i \overline{F}_Y(z)dz = C_i E[Y] \quad \text{for} \quad i \in [s_0+1, s_0+2, \ldots, S_0]$$

The steady-state p.d.f. of Z is obtained by summing $g(i,z)$ over all possible values of i.

$$f_Z(z) \;=\; \sum_{i=s_0+1}^{S_0} g(i,z) = \sum_{i=s_0+1}^{S_0} C_i \overline{F}_Y(z)$$

The result follows from Equation (5.4).

Proof of Lemma 5.2.2:

To prove the independence of $IP_0(t)$ and $D_0(t, t+\tau]$, we should prove that

$$\lim_{t\to\infty} P(IP_0(t) = i, D_0(t, t+\tau] = k) = \pi_i \lim_{t\to\infty} P(D_0(t, t+\tau] = k)$$
$$\forall i \in [s_0+1, s_0+2, \ldots, S_0], k \in [0, 1, 2, \ldots)$$

$$\begin{aligned}
\lim_{t\to\infty} P(IP_0(t) = i, D_0(t, t+\tau] = k) &= \lim_{t\to\infty} \int_{z=0}^{t} P(IP_0(t) = i, Z(t) = z, D_0(t, t+\tau] = k)dz \\
&= \lim_{t\to\infty} \int_{z=0}^{t} P(D_0(t, t+\tau] = k | IP_0(t) = i, Z(t) = z) g(t, i, z)dz \\
&= \int_{z=0}^{\infty} \lim_{t\to\infty} P\left(D_0(t, t+\tau] = 0 | IP_0(t) = i, Z(t) = z\right) g(i,z)dz \\
&= \int_{z=0}^{\infty} \frac{P(D_0(0, z+\tau] = k)}{\overline{F}_Y(z)} C_i \overline{F}_Y(z)dz = C_i \int_{z=0}^{\infty} \varphi(z+\tau, k)dz \qquad (9.22)
\end{aligned}$$

Equation (9.22) follows since, given that the last order is placed at time $t - z$ *ie.* *Z(t)=z*, the distribution of $D_0(t, t+\tau]$ is the same as $D_0(0, z+\tau]$.
The next step is to find $\lim_{t\to\infty} P(D_0(t, t+\tau] = k)$.

$$\begin{aligned}
\lim_{t\to\infty} P(D_0(t,t+\tau] = k) &= \lim_{t\to\infty} \int_{z=0}^{t} P(D_0(t,t+\tau] = k | Z(t) = z) f_{Z(t)}(z) dz \\
&= \int_{z=0}^{\infty} \lim_{t\to\infty} P(D_0(t,t+\tau] = k | Z(t) = z) f_Z(z) dz = \int_{z=0}^{\infty} \frac{\varphi(z+\tau,k)}{\overline{F}_Y(z)} \frac{\overline{F}_Y(z)}{E[Y]} dz \\
&= \frac{1}{E[Y]} \int_{z=0}^{\infty} \varphi(z+\tau,k) dz
\end{aligned}$$

Then,

$$\begin{aligned}
\pi_i \lim_{t\to\infty} P(D_0(t,t+\tau] = k) = C_i E[Y] \frac{1}{E[Y]} \int_{z=0}^{\infty} \varphi(z+\tau,k) dz = C_i \int_{z=0}^{\infty} \varphi(z+\tau,k) dz \\
\lim_{t\to\infty} P(IP_0(t) = i, D_0(t,t+\tau] = k)
\end{aligned}$$

Proof of Lemma 5.2.3:

a) $s_0 \geq 0$

- $\tau < L_0$. Using Equation (5.5), we can write

$$\begin{aligned}
P(W_0(t,q) \leq \tau) &= P(IP_0(t+\tau-L_0) - D_0[t+\tau-L_0, t) \geq q) \\
&= \sum_{i=\max(s_0+1,q)}^{S_0} P(IP_0(t+\tau-L_0) = i, D_0[t+\tau-L_0,t) \geq i-q)
\end{aligned}$$

Due to Lemma 5.2.2, we can write

$$\begin{aligned}
F_{W_0(q)}(\tau) &= \lim_{t\to\infty} P(W_0(t,q) \leq \tau) \\
&= \lim_{t\to\infty} \sum_{i=\max(s_0+1,q)}^{S_0} P(IP_0(t+\tau-L_0) = i) P(D_0[t+\tau-L_0,t) \leq i-q)
\end{aligned}$$

Note that since t is an order point, looking backward from t, $D_0[t+\tau-L_0, t)$ has the same distribution as $D_0(0, L_0-\tau]$. Then, it follows that

$$\begin{aligned}
F_{W_0(q)}(\tau) &= \sum_{i=\max(s_0+1,q)}^{S_0} P(IP_0 = i) P\left(D_0(0, L_0-\tau] \leq i-q\right) \\
&= \sum_{i=\max(s_0+1,q)}^{S_0} \sum_{k=0}^{i-q} \pi_i \varphi(L_0-\tau, k)
\end{aligned}$$

where $\varphi(L_0 - \tau, k)$ and π_i are given in Lemmas (5.1.1) and (5.2.1), respectively.

- $\tau \geq L_0$ Since we know that $W_0(t,q) \leq L_0$ for $s_0 \geq 0$, we have

$$\lim_{t\to\infty} P(W_0(t,q) \leq \tau) \;=\; F_{W_0(q)}(\tau) = 1 \quad \text{for} \quad \tau \geq L_0$$

b) $s_0 < 0$

- $\tau < L_0$. $F_{W_0(q)}(\tau)$ for this case is quite similar to $s_0 \geq 0$. So, we omit here.

- $\tau = L_0$.

$$F_{W_0(q)}(L_0) \;=\; \lim_{\tau\to L_0^-} F_{W_0(q)}(\tau) + P(W_0(q) = L_0)$$

Using Equation (5.8), we have

$$\begin{aligned}\lim_{t\to\infty} P(W_0(t,q) = L_0) \;&=\; P(W_0(q) = L_0) \\ &= \lim_{t\to\infty} \sum_{i=s_0+1}^{s_0+q} P(IP_0(t^-) = i) = \sum_{i=s_0+1}^{s_0+q} \pi_i\end{aligned}$$

and

$$\begin{aligned}\lim_{\tau\to L_0^-} F_{W_0(q)}(\tau) \;&=\; \lim_{\tau\to L_0^-} \sum_{i=\max(s_0+1,q)}^{S_0} \sum_{k=0}^{i-q} \pi_i \varphi(L_0-\tau,k) \\ &= \sum_{i=q}^{S_0} \sum_{k=0}^{i-q} \pi_i \lim_{t\to 0^+} \varphi(t,k) = \sum_{i=q}^{S_0} \pi_i\end{aligned}$$

since $\lim_{t\to 0^+} \varphi(t,k)$ takes the value of 1 if $k=0$ and 0 otherwise. Therefore,

$$F_{W_0(q)}(L_0) \;=\; \sum_{i=q}^{S_0} \pi_i + \sum_{i=s_0+1}^{s_0+q} \pi_i = 1 - \sum_{i=s_0+q+1}^{q-1} \pi_i$$

- $\tau > L_0$. Using Equation (5.9), we have

$$P(L_0 < W_0(t,q) \leq \tau) \;=\; P(IP_0(t^-) - D_0[t, t+\tau-L_0] \leq s_0)$$

Here, we can write $D_0[t, t+\tau-L_0] = q + D_0(t, t+\tau-L_0]$. Therefore,

$$P(L_0 < W_0(t,q) \le \tau) = P(IP_0(t^-) - D_0(t, t+\tau-L_0] \le s_0 + q)$$
$$= \sum_{i=s_0+q+1}^{q-1} P(IP_0(t^-) = i, D_0(t, t+\tau-L_0] \ge i - q - s_0)$$

Since t is a demand point, $D_0(t, t+\tau-L_0]$ and $D_0(0, \tau-L_0]$ has the same distribution. Due to Lemma 5.2.2, $IP_0(t)$ and $D_0(t, t+\tau-L_0]$ are independent as $t \to \infty$. Therefore,

$$\lim_{t\to\infty} P(L_0 < W_0(t,q) \le \tau)$$
$$= \lim_{t\to\infty} \sum_{i=s_0+q+1}^{q-1} P(IP_0(t^-) = i) P(D_0(t, t+\tau-L_0] \ge i - q - s_0)$$

$$P(L_0 < W_0(q) \le \tau) = \sum_{i=s_0+q+1}^{q-1} \pi_i P(D_0(0, \tau-L_0] \ge i - q - s_0)$$
$$= \sum_{i=s_0+q+1}^{q-1} \pi_i (1 - P(D_0(0, \tau-L_0] \le i - q - s_0 - 1))$$
$$= \sum_{i=s_0+q+1}^{q-1} \pi_i - \sum_{i=s_0+q+1}^{q-1} \sum_{k=0}^{i-q-s_0-1} \pi_i \varphi(\tau - L_0, k)$$

Then, for $\tau > L_0$,

$$\begin{aligned} F_{W_0(q)}(\tau) &= F_{W_0(q)}(L_0) + P(L_0 < W_0(q) \le \tau) \\ &= 1 - \sum_{i=s_0+q+1}^{q-1} \pi_i + \sum_{i=s_0+q+1}^{q-1} \pi_i - \sum_{i=s_0+q+1}^{q-1} \sum_{k=0}^{i-q-s_0-1} \pi_i \varphi(\tau - L_0, k) \\ &= 1 - \sum_{i=s_0+q+1}^{q-1} \sum_{k=0}^{i-q-s_0-1} \pi_i \varphi(\tau - L_0, k) \end{aligned}$$

Proof of Lemma 5.2.4:

Referring to 5.10, we can write $P(OH_0(t+L_0) = i)$ conditioning on $(IP_0(t), Z(t))$ as follows:

$$P(OH_0(t + L_0) = i) =$$

$$= \int_{z=0}^{\infty} \sum_{k=0}^{S_0-i} P(OH_0(t+L_0) = i | IP_0(t) = i+k, Z(t) = z) g(t, i+k, z) dz$$
$$= \int_{z=0}^{t} P(OH_0(t+L_0) = i | IP_0(t) = i, Z(t) = z) g(t, i, z) dz$$
$$+ \int_{z=0}^{t} \sum_{k=1}^{S_0-i} P(OH_0(t+L_0) = i | IP_0(t) = i+k, Z(t) = z) g(t, i+k, z) dz$$

Let $t \to \infty$. Then, we can write:

$$\vartheta_i = \lim_{t\to\infty} P(OH_0(t+L_0) = i) = P(OH_0 = i)$$
$$= \int_{z=0}^{\infty} \left[P(Y_1 > L_0 + z | Y_1 > z) + P(Y_1 \le L_0 + z, Q_1 > i | Y_1 > z)\right] C_i \overline{F}_Y(z) dz$$
$$+ \sum_{k=1}^{S_0-i} \sum_{n=1}^{\infty} \int_{z=0}^{\infty} P(\sum_{i=1}^{n} Y_i \le L_0 + z, \sum_{i=1}^{n} Q_i = k, \sum_{i=1}^{n+1} Q_i > k+i | Y_1 > z) C_{i+k} \overline{F}_Y(z) dz$$
$$= \int_{z=0}^{\infty} C_i \left[\overline{F}_Y(L_0+z) + \sum_{j=i+1}^{\infty} [F_{Y,Q_0}(L_0+z, j) - F_{Y,Q_0}(z, j)]\right] dz$$
$$+ \sum_{k=1}^{S_0-i} \sum_{n=1}^{\infty} \int_{z=0}^{\infty} C_{i+k} P(\sum_{i=1}^{n} Y_i \le L_0 + z, \sum_{i=1}^{n} Q_i = k, \sum_{i=1}^{n+1} Q_i > k+i, Y_1 > z) dz$$
$$= \int_{z=0}^{\infty} C_i \left[\overline{F}_Y(L_0+z) + \sum_{j=i+1}^{\infty} [F_{Y,Q_0}(L_0+z, j) - F_{Y,Q_0}(z, j)]\right] dz$$
$$+ \sum_{k=1}^{S_0-i} \sum_{n=1}^{\infty} \sum_{j=1}^{k} C_{i+k} \int_{z=0}^{\infty} \int_{t=z}^{L_0+z} P(\sum_{i=2}^{n} Y_i \le L_0 + z - t, \sum_{i=2}^{n} Q_i = k-j, Q_{n+1} > i) dF_{Y,Q_0}(t, j) dz$$
$$= C_i \int_{z=0}^{\infty} \left[\overline{F}_Y(L_0+z) + \sum_{j=i+1}^{\infty} [F_{Y,Q_0}(L_0+z, j) - F_{Y,Q_0}(z, j)]\right] dz$$
$$+ \overline{F}_{Q_0}(i) \sum_{k=1}^{S_0-i} \sum_{n=1}^{\infty} \sum_{j=1}^{k} C_{i+k} \int_{z=0}^{\infty} \int_{t=z}^{L_0+z} F_{Y(n-1), Q_0^{(n-1)}}(L_0 + z - t, k-j) dF_{Y,Q_0}(t, j) dz$$

Proof of Equation 5.12:

$$E[\eta(S_0 - s_0)] = \sum_{n=1}^{S_0 - s_0} n P(\eta(S_0 - s_0) = n)$$

Using the definition of $\eta(S_0 - s_0) = \min\{n : \sum_{i=1}^{n} Q_i \ge S_0 - s_0\}$, we can write:

$$E[\eta(S_0 - s_0)] = \sum_{n=1}^{S_0 - s_0} n P(\sum_{i=1}^{n-1} Q_i < S_0 - s_0, \sum_{i=1}^{n-1} Q_i \ge S_0 - s_0)$$

$$
\begin{aligned}
&= \sum_{n=1}^{S_0-s_0} n \left[\sum_{k=0}^{S_0-s_0-1} P(\sum_{i=1}^{n-1} Q_i = k, Q_n \geq S_0 - s_0 - k) \right] \\
&= \sum_{n=1}^{S_0-s_0} n \left[\sum_{k=0}^{S_0-s_0-1} P(Q_0^{(n-1)} = k)[1 - F_{Q_0}(S_0 - s_0 - k - 1)] \right] \\
&= \sum_{n=1}^{S_0-s_0} n \left[F_{Q_0}^{(n-1)}(S_0 - s_0 - 1) - F_{Q_0}^{(n)}(S_0 - s_0 - 1) \right] \\
&= \sum_{n=0}^{S_0-s_0-1} F_{Q_0}^{(n)}(S_0 - s_0 - 1)
\end{aligned}
$$

Proof of Equation 5.13:

$$
\begin{aligned}
E[Q_w] &= \sum_{n=S_0-s_0}^{\infty} nP(\kappa(S_0 - s_0) = n) \\
&= \sum_{n=S_0-s_0}^{\infty} n \left[\sum_{k=1}^{n} P(\sum_{i=1}^{k-1} Q_i \leq S_0 - s_0 - 1, \sum_{i=1}^{k} Q_i = n) \right] \\
&= \sum_{n=S_0-s_0}^{\infty} n \left[\sum_{k=1}^{n} \sum_{j=k-1}^{S_0-s_0-1} P(\sum_{i=1}^{k-1} Q_i = j, Q_k = n - j) \right] \\
&= \sum_{n=S_0-s_0}^{\infty} n \left[\sum_{k=1}^{n} \sum_{j=k-1}^{S_0-s_0-1} P_{Q_0}^{(k-1)})(j) P_{Q_0}(n - j) \right]
\end{aligned}
$$

Proof of Lemma 5.2.5:

First, let $m_i \leq S_i$. Using 5.14, $E[AR_i(m_i, q)]$ can be written as:

$$
\begin{aligned}
E[AR_i(m_i, q)] &= \sum_{j=1}^{m_i} E[(X^i_{S_i-m_i+j} - T_i(q))I(X^i_{S_i-m_i+j} > T_i(q))] \\
&= \sum_{j=1}^{m_i} E[X^i_{S_i-m_i+j} I(X^i_{S_i-m_i+j} > L_i + W_0(q))] \\
&- \sum_{j=1}^{m_i} E[(L_i + W_0(q))I(X^i_{S_i-m_i+j} > L_i + W_0(q))] \\
&= \sum_{j=1}^{m_i} \int_{w=0}^{\infty} [\int_{t=L_i+w}^{\infty} tf(t, S_i - m_i + j, \lambda_i)dt] dF_{W_0(q)}(w) \\
&- \sum_{j=1}^{m_i} \int_{w=0}^{\infty} (L_i + w)P(X^i_{S_i-m_i+j} > L_i + w) dF_{W_0(q)}(w)
\end{aligned}
$$

where $f(x,k,\lambda)$ be the p.d.f. of an Erlang random variable with shape and scale parameters, k and λ, respectively. Substituting Erlang probabilities for $X^i_{S_i-m_i+j}$, we have

$$E[AR_i(m_i,q)] = \sum_{j=1}^{m_i}\int_{w=0}^{\infty}\left[\int_{t=L_i+w}^{\infty} t\lambda_i \frac{e^{-\lambda_i t}(\lambda_i t)^{S_i-m_i+j-1}}{\Gamma(S_i-m_i+j)}dt\right]dF_{W_0(q)}(w)$$
$$-\sum_{j=1}^{m_i}\int_{w=0}^{\infty}(L_i+w)\overline{F}(L_i+w,S_i-m_i+j,\lambda_i)dF_{W_0(q)}(w)$$
$$=\sum_{j=1}^{m_i}\int_{w=0}^{\infty}\left[\frac{S_i-m_i+j}{\lambda_i}\int_{t=L_i+w}^{\infty}\lambda_i\frac{e^{-\lambda_i t}(\lambda_i t)^{S_i-m_i+j}}{\Gamma(S_i-m_i+j+1)}dt\right]dF_{W_0(q)}(w)$$
$$-\sum_{j=1}^{m_i}\int_{w=0}^{\infty}(L_i+w)\overline{F}(L_i+w,S_i-m_i+j,\lambda_i)dF_{W_0(q)}(w)$$
$$=\sum_{j=1}^{m_i}\int_{w=0}^{\infty}\left[\frac{S_i-m_i+j}{\lambda_i}\int_{t=L_i+w}^{\infty}f(t,S_i-m_i+j+1,\lambda_i)dt\right]dF_{W_0(q)}(w)$$
$$-\sum_{j=1}^{m_i}\int_{w=0}^{\infty}(L_i+w)\overline{F}(L_i+w,S_i-m_i+j,\lambda_i)dF_{W_0(q)}(w)$$
$$=\sum_{j=1}^{m_i}\int_{w=0}^{\infty}\left[\frac{S_i-m_i+j}{\lambda_i}\overline{F}(L_i+w,S_i-m_i+j+1,\lambda_i)\right]dF_{W_0(q)}(w)$$
$$-\sum_{j=1}^{m_i}\int_{w=0}^{\infty}(L_i+w)\overline{F}(L_i+w,S_i-m_i+j,\lambda_i)dF_{W_0(q)}(w)$$

Similarly, $E[B_i(m_i,q)]$ will be written as

$$E[B_i(m_i,q)] = \sum_{k=1}^{m_i-1}(m_i-k)P(X^i_{S_i-k} < T_i(q) < X^i_{S_i-k+1}) + m_i P(X^i_{S_i} < T_i(q))$$

After standard algebraic operations, we get:

$$E[B_i(m_i,q)] = m_i - \sum_{j=1}^{m_i}P(T_i(q)\le X^i_{S_i-m_i+j}) = \sum_{j=1}^{m_i}P(X^i_{S_i-m_i+j} < L_i + W_0(q))$$
$$= \sum_{j=1}^{m_i}\int_{w=0}^{\infty}F(L_i+w,S_i-m_i+j,\lambda_i)dF_{W_0(q)}(w)$$

Now, let $m_i > S_i$. Using (5.15), we can write $E[AR_i(m_i,q)]$ as:

$$E[AR_i(m_i,q) = \sum_{j=m_i-S_i+1}^{m_i}E[X^i_{S_i-m_i+j}I(X^i_{S_i-m_i+j} > L_i + W_0(q))]$$

$$- \sum_{j=m_i-S_i+1}^{m_i} E[L_i + W_0(q) I(X^i_{S_i-m_i+j} > L_i + W_0(q))]$$

Using the same steps as in case of $S_i \geq m_i$, we obtain

$$\begin{aligned} E[AR_i(m_i,q) &= \sum_{j=m_i-S_i+1}^{m_i} \int_{w=0}^{L_0} \left[\frac{S_i - m_i + j}{\lambda_i} \overline{F}(L_i + w, S_i - m_i + j + 1, \lambda_i) \right] dF_{W_0(q)}(w) \\ &- \sum_{j=m_i-S_i+1}^{m_i} \int_{w=0}^{L_0} \left[(L_i + w) \overline{F}(L_i + w, S_i - m_i + j, \lambda_i) \right] dF_{W_0(q)}(w) \end{aligned}$$

$E[B_i(m_i,q)]$ can be calculated as

$$E[B_i(m_i,q)] = (m_i - S_i) + \sum_{k=1}^{S_i-1} kP(X^i_k < T_i(q) \leq X^i_{k+1}) + S_i P(X^i_{S_i} < T_i(q))$$

Using the same approach, we write:

$$\begin{aligned} E[B_i(m_i,q)] &= (m_i - S_i) + \sum_{j=m_i-S_i+1}^{m_i} P(X^i_{S_i-m_i+j} < T_i(q)) \\ &= (m_i - S_i) + \sum_{j=m_i-S_i+1}^{m_i} P(X^i_{S_i-m_i+j} < L_i + W_0(q)) \\ &= \sum_{j=m_i-S_i+1}^{m_i} \int_{w=0}^{\infty} F(L_i + w, S_i - m_i + j, \lambda_i) dF_{W_0(q)}(w) \end{aligned}$$

Proof of Corollary 6.1.2:

First suppose $s_0 \geq 0$. Due to restrictions on (s_0, S_0) values and using Lemma 5.2.3, we can write for $\tau < L_0$:

$$\begin{aligned} F_{W_0(q)}(\tau) &= \sum_{i \in \{s_0+Q, s_0+2Q, \ldots, S_0\}} \sum_{k \in \{0, Q, \ldots, i-Q\}} \frac{1}{\zeta_{\Delta_0}} \varphi(L_0 - \tau, k) \\ &= \sum_{i \in \{s_0+Q, s_0+2Q, \ldots, S_0\}} \frac{1}{\zeta_{\Delta_0}} [1 - F(L_0 - \tau, i, \lambda_0)] \\ &= 1 - \sum_{i=1}^{\zeta_{\Delta_0}} \frac{1}{\zeta_{\Delta_0}} F(L_0 - \tau, s_0 + iQ, \lambda_0) \end{aligned}$$

Now, suppose $s_0 < 0$ and let $\tau < L_0$. Then,

$$\begin{aligned}
F_{W_0(q)}(\tau) &= \sum_{i\in\{Q,2Q,\dots,S_0\}} \sum_{k\in\{0,Q,\dots,i-Q\}} \frac{1}{\zeta_{\Delta_0}} \varphi(L_0-\tau,k) \\
&= \sum_{i\in\{Q,2Q,\dots,S_0\}} \frac{1}{\zeta_{\Delta_0}} \left[1 - F(L_0-\tau,i,\lambda_0)\right] \\
&= \frac{S_0}{Q}\frac{1}{\zeta_{\Delta_0}} - \sum_{i\in\{Q,2Q,\dots,S_0\}} \frac{1}{\zeta_{\Delta_0}} F(L_0-\tau,i,\lambda_0) \\
&= \frac{\zeta_{S_0}}{\zeta_{\Delta_0}} - \frac{1}{\zeta_{\Delta_0}} \sum_{i=1}^{\zeta_{S_0}} F(L_0-\tau,iQ,\lambda_0)
\end{aligned}$$

Now, let $\tau = L_0$. Then,

$$F_{W_0(q)}(\tau) = 1 - \sum_{i\in\{s_0+2Q,s_0+3Q,\dots,0\}} \frac{1}{\zeta_{\Delta_0}} = 1 - \frac{-s_0-Q}{Q}\frac{1}{\zeta_{\Delta_0}} = 1 - \frac{\zeta_{-s_0-Q}}{\zeta_{\Delta_0}}$$

For $\tau > L_0$, we have:

$$\begin{aligned}
F_{W_0(q)}(\tau) &= 1 - \sum_{i\in\{s_0+2Q,s_0+3Q,\dots,0\}} \sum_{k\in\{0,Q,\dots,i-s_0-2Q\}} \frac{1}{\zeta_{\Delta_0}} \varphi(\tau-L_0,k) \\
&= 1 - \sum_{i\in\{s_0+2Q,s_0+3Q,\dots,0\}} \frac{1}{\zeta_{\Delta_0}} \left[1 - F(\tau-L_0,i-s_0-Q,\lambda_0)\right] \\
&= 1 - \frac{\zeta_{-s_0-Q}}{\zeta_{\Delta_0}} + \frac{1}{\zeta_{\Delta_0}} \sum_{k=1}^{\zeta_{-s_0-Q}} F(\tau-L_0,kQ,\lambda_0)
\end{aligned}$$

Proof of Corollary 6.2.1:

Observe that $Y_d = Y - MT$ where $M \geq 0$ corresponds to the number of review instances which do not trigger an order between two ordering instances.

Now, suppose that $0 < q < Q$ and $y = T$. Then, we have

$$\begin{aligned}
f_{Y_d,Q_0}(y,q) &= \sum_{m=0}^{\infty} f_{Y,Q_0}((m+1)T,q) = \sum_{m=0}^{\infty} \phi_0^m p_0(q,\lambda_0 T) \\
&= p_0(q,\lambda_0 T)/(1-\phi_0)
\end{aligned}$$

$$f_{Y_d|Q_0}(T|q) = f_{Y_d,Q_0}(T,q)/P_{Q_0}(q) = \frac{p_0(q,\lambda_0 T)/(1-\phi_0)}{p_0(q,\lambda_0 T)/(1-\phi_0)} = 1$$

Similarly, for $q = Q$ and $0 < y < T$, we have

$$\begin{aligned} f_{Y_d,Q_0}(y,Q) &= \sum_{m=0}^{\infty} f_{Y,Q_0}(y+mT,Q) = \sum_{m=0}^{\infty} \phi_0^m f(y,Q,\lambda_0) \\ &= f(y,Q,\lambda_0)/(1-\phi_0) \end{aligned}$$

Then, we have

$$f_{Y_d|Q_0}(y|Q) = \frac{f(y,Q,\lambda_0)/(1-\phi_0)}{\overline{P}_0(Q-1,\lambda_0 T)/(1-\phi_0)} = \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} = f_{T_e}(y,T,Q,\lambda_0)$$

Proof of Corollary 6.2.2:

$a)$ Using conditional distribution, $f_{Y_d^{(n)},Q_0^{(n)}}(y,q)$ can be written as:

$$f_{Y_d^{(n)},Q_0^{(n)}}(y,q) = f_{Y_d^{(n)}|Q_0^{(n)}}(y|q) P_{Q_0}^{(n)}(q) \tag{9.23}$$

First, suppose that $q = nQ$. It is obvious that all n orders are triggered by the the accumulation of Q demands in the system. In view of Corollary 6.2.2, Y_d has a truncated Erlang distribution given an order of size Q. Therefore,

$$f_{Y_d^{(n)},Q_0^{(n)}}(y,nQ) = f_{T_e}^{(n)}(y,T,Q,\lambda_0) P_{Q_0}^{(n)}(Q) = f_{T_e}^{(n)}(y,T,Q,\lambda_0) P_{Q_0}(Q)^n$$

Now, suppose that $n \le q < nQ$ and let $m \ge 1$ be the number of orders with size less than Q units. Then, these m orders will demand a total of $q-(n-m)Q$ units since the remaining $(n-m)$ orders will have a size of Q units. Then,

$$P_{Q_0}^{(n)}(q) = \sum_{m=1}^{n} C(n,m) P_{Q_0}^{(n-m)}((n-m)Q) P_{Q_0}^{\prime(m)}(q-(n-m)Q)$$

$P_{Q_0}^{\prime(m)}(q)$ is the probability that a total of q units are demanded in m orders each of which have a size less than Q. Using Equation (9.23), we can write:

$$\begin{aligned} & f_{Y_d^{(n)},Q_0^{(n)}}(y,q) = \\ & = \sum_{m=1}^{n} \left[\begin{array}{l} C(n,m) P_{Q_0}^{(n-m)}((n-m)Q) P_{Q_0}^{\prime m}(q-(n-m)Q) \\ f_{Y_d^{(n-m)}|Q_0^{(n-m)}}(y-mT|(n-m)Q) f_{Y_d^{(m)}|Q_0^{(m)}}(mT|q-(n-m)Q) \end{array} \right] \end{aligned} \tag{9.24}$$

Equation (9.24) follows because, for $q < Q$, $f_{Y_d|Q_0}(y|q)$ has a probability mass of one if $y = T$. Then, we have:

$$f_{Y_d^{(n)},Q_0^{(n)}}(y,q) =$$
$$= \sum_{m=1}^{n} C(n,m) P_{Q_0}((n-m)Q)^{n-m} P'^{m}_{Q_0}(q-(n-m)Q) f_{T_e}^{(n-m)}(y-mT,T,Q,\lambda_0)$$

b) We first point out the following observation:

$$Y^{(n)} = Y_d^{(n)} + KT$$

where K is a random variable corresponding to the number of review instances which do not result in an order trigger.

Suppose that there are $K = k$ review instances in $(0, y]$ and therefore, we have a total of $k+n$ decision epochs at which either an order is placed or only a review is carried out. The probability that $k+n$'th decision epoch is an order instance is given by $C(n+k-1,k)\phi_0^k(1-\phi_0)^n$.

$$f_{Y^{(n)},Q_0^{(n)}}(y,q) \;\; = \;\; \sum_k C(n+k-1,k)\phi_0^k(1-\phi_0)^n f_{Y_d^{(n)},Q_0^{(n)}}(y-kT,q)$$

The limits on k is determined using $0 < y - kT \leq nT$.

Proof of Equation (6.5):

First, suppose that $t \leq T$.

$$F_{T_e}^{(2)}(t,T,Q,\lambda_0) = \int_{y=0}^{t}\int_{x=0}^{t-y} f_{T_e}(x,T,Q,\lambda_0) f_{T_e}(y,T,Q,\lambda_0)\,dx\,dy$$
$$= \int_{y=0}^{t} \frac{F(t-y,Q,\lambda_0)}{F(T,Q,\lambda_0)} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} dy = F(t,2Q,\lambda_0)/F(T,Q,\lambda_0)^2$$

Now, suppose $T < t \leq 2T$.

$$F_{T_e}^{(2)}(t,T,Q,\lambda_0) = 1 - \int_{y=t-T}^{T}\int_{x=t-y}^{T} f_{T_e}(x,T,Q,\lambda_0) f_{T_e}(y,T,Q,\lambda_0)\,dy\,dx$$
$$= 1 - \int_{y=t-T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)^2}\left[F(T,Q,\lambda_0) - F(t-y,Q,\lambda_0)\right] dy$$
$$= 1 - \int_{y=t-T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} dy + \int_{y=t-T}^{T} \frac{f(y,Q,\lambda_0)F(t-y,Q,\lambda_0)}{F(T,Q,\lambda_0)^2} dy$$
$$= \frac{F(t-T,Q,\lambda_0)}{F(T,Q,\lambda_0)} + \int_{y=t-T}^{T} \frac{f(y,Q,\lambda_0)F(t-y,Q,\lambda_0)}{F(T,Q,\lambda_0)^2} dy$$

Proof of Equation (6.6):
First, take $t \leq T$.

$$F_{T_e}^{(3)}(t,T,Q,\lambda_0) = \int_{y=0}^{t}\int_{x=0}^{t-y} f_{T_e}^{(2)}(x,T,Q,\lambda_0) f_{T_e}(y,T,Q,\lambda_0)dxdy$$
$$= \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)dy$$
$$= \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)F(t-y,2Q,\lambda_0)}{F(T,Q,\lambda_0)^3}dy = \frac{F(T,3Q,\lambda_0)}{F(T,Q,\lambda_0)^3}$$

Now, let $T \leq t \leq 2T$.

$$F_{T_e}^{(3)}(t,T,Q,\lambda_0) = F_{T_e}^{(3)}(T,T,Q,\lambda_0) + \int_{y=0}^{T}\int_{x=T-y}^{t-y} f_{T_e}^{(2)}(x,T,Q,\lambda_0) f_{T_e}(y,T,Q,\lambda_0)dxdy$$
$$= \frac{F(T,3Q,\lambda_0)}{F(T,Q,\lambda_0)^3} + \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} \left[F_{T_e}^{(2)}(t-y,T,Q,\lambda_0) - F_{T_e}^{(2)}(T-y,T,Q,\lambda_0)\right] dy$$
$$= \frac{F(T,3Q,\lambda_0)}{F(T,Q,\lambda_0)^3} + \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)dy$$
$$- \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)F(t-y,2Q,\lambda_0)}{F(T,Q,\lambda_0)^3}dy = \int_{y=0}^{t} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)dy$$

Finally, suppose that $2T \leq t < 3T$.

$$F_{T_e}^{(3)}(t,T,Q,\lambda_0) = 1 - \int_{y=t-2T}^{T}\int_{x=t-y}^{2t} f_{T_e}^{(2)}(x,T,Q,\lambda_0) f_{T_e}(y,T,Q,\lambda_0)dxdy$$
$$= 1 - \int_{y=t-2T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} \left[F_{T_e}^{(2)}(2T,T,Q,\lambda_0) - F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)\right] dy$$
$$= 1 - \int_{y=t-2T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)}dy + \int_{y=t-2T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)dy$$
$$= \frac{F(t-2T,Q,\lambda_0)}{F(T,Q,\lambda_0)} + \int_{y=t-2T}^{T} \frac{f(y,Q,\lambda_0)}{F(T,Q,\lambda_0)} F_{T_e}^{(2)}(t-y,T,Q,\lambda_0)dy$$

Proof of Lemma 6.3.1: Observe that the inter order time $Y = T$ and the order quantity is $Q_0 = q \geq Q$ if q units have been demanded during the last T time units. Therefore,

$$f_{Y,Q_0}(T,q) \quad = \quad p_0(q,\lambda_0 T) \quad \text{if} \quad q \geq Q$$

Similarly, the inter order time is $Y = mT, m > 1$ and the order quantity is $Q_0 = q \geq Q$ if $k \leq Q - 1$ demands have arrived during $(m-1)T$ time units and $q - k$ units have been demanded during the last period of T time units. Then, we can write:

$$\begin{aligned} f_{Y,Q_0}(mT, q) &= \sum_{k=0}^{Q-1} p_0(k, \lambda_0(m-1)T) p_0(q-k, \lambda_0 T) \qquad (9.25) \\ &= \sum_{k=0}^{Q-1} \frac{e^{-\lambda_0(m-1)T}(\lambda_0(m-1)T)^k}{k!} \frac{e^{-\lambda_0 T}(\lambda_0 T)^{(q-k)}}{(q-k)!} \\ &= \frac{e^{-\lambda_0 mT}(\lambda_0 mT)^q}{q!} \sum_{k=0}^{Q-1} \frac{q!}{k!(q-k)!} \left(1 - \frac{1}{m}\right)^k \left(\frac{1}{m}\right)^{q-k} \\ &= p_0(q, \lambda_0 mT) B(Q-1, q, 1 - 1/m) \quad \text{if} \quad q \geq Q \end{aligned}$$

Proof of Corollary 6.3.1:

a) $P_{Q_0}(q)$ is obtained by summing $f_{Y,Q_0}(y, q)$ over all possible values of y and does not have a further closed expression.

b) First let $y = T$. Then,

$$f_Y(T) = \sum_{q=Q}^{\infty} p_0(q, \lambda_0 T) = \overline{P}_0(Q-1, \lambda_0 T)$$

Now, let $y = mT, m > 1$. To find $f_Y(mT)$, instead of using the final expression of $f_{Y,Q_0}(mT, q)$, we will use an intermediate expression provided in Equation (9.25).

$$\begin{aligned} f_Y(mT) &= \sum_{q=Q}^{\infty} \sum_{k=0}^{Q-1} p_0(k, \lambda_0(m-1)T) p_0(q-k, \lambda_0 T) \\ &= \sum_{k=0}^{Q-1} p_0(k, \lambda_0(m-1)T) \left[1 - P_0(Q-k-1, \lambda_0 T)\right] \\ &= P_0(k, \lambda_0(m-1)T) - \sum_{k=0}^{Q-1} P_0(Q-k-1, \lambda_0(m-1)T) p_0(k, \lambda_0 T) \\ &= P_0(Q-1, \lambda_0(m-1)T) - P_0(Q-1, \lambda_0 mT) \end{aligned}$$

The result follows from the fact that Poisson distribution is closed under convolution.

Proof of Corollary 6.3.2: If $t < T$, $D_0(0,t] = 0$ should hold. Therefore, $\varphi(t,0) = 1$ for $t < T$. Since the orders arrive in batches of a minimum size of Q, for $t \geq T$, $D_0(0,t]$ has a positive mass if $k = 0$ or $k \geq Q$.

Using Lemma 5.1.1 and Corollary 6.3.1, we can write:

$$\begin{aligned}
\varphi(t,0) &= \overline{F}_Y(t) = 1 - \sum_{m=1}^{\lfloor t/T \rfloor} f_Y(mT) \\
&= 1 - \left[1 - P_0(Q-1, \lambda_0 T) + \sum_{m=2}^{\lfloor t/T \rfloor} \left(P_0(Q-1, \lambda_0(m-1)T) - P_0(Q-1, \lambda_0 mT) \right) \right] \\
&= P_0(Q-1, \lfloor t/T \rfloor \lambda_0 T)
\end{aligned}$$

For $t \geq T$ and $k \geq 0$, Lemma 5.1.1 gives

$$\varphi(t,k) \;=\; \sum_{n=1}^{\lfloor k/Q \rfloor} \left[F_{Y^{(n)},Q_0^{(n)}}(t,k) - \sum_{m=1}^{\lfloor (t-nT)/T \rfloor} F_{Y^{(n)},Q_0^{(n)}}(t-mT,k) f_y(mT) \right]$$

Using Equation (6.8) for $F_{Y^{(n)},Q_0^{(n)}}(t,k)$ and $F_{Y^{(n)},Q_0^{(n)}}(t-mT,k)$ provides the result.

Proof of Lemma 6.4.1: Let $N_i(t)$ be the counting process of retailer i demands in $(0,t]$ where $t = 0$ is taken as the time of the last order where the inventory positions of all retailers are at their maximum levels.

Observe that the inter-order time is $Y \in (y, y+\delta y]$ and the order size is q if $(q-1)$ demands that do not trigger an order arrive in $(0,y]$, the q'th demand that arrives in $(y, y+\delta y]$ triggers the order. The order of size q is triggered by retailer i in $[y, y+\delta y)$ if $q \geq \Delta_i$, $N_i(y) = \Delta_i - 1$, $N_i(y+\delta y) = x_i = \Delta_i$ and for $j \neq i$ $N_j(y) = x_j < \Delta_j$ so that $\sum_{j=1}^{N} x_j = q$. Due to Poisson demands, we can write

$$P(Y \in [y, y+\delta y), Q_0 = q) = \\
= \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{c} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \leq x_j < \Delta_j \text{ for } j \neq i \end{array} \right\}} p_0(\Delta_i - 1, \lambda_i y) \left[\prod_{j \neq i} p_0(x_j, \lambda_j y) \right] [\lambda_i \delta y + o(\delta y)] \, I(q \geq \Delta_i) \tag{9.26}$$

Expand Poisson probabilities and multiply and divide the right hand side of Equation (9.26) by $(q-1)!/\lambda_0^q$. Then, we have

$$
\begin{aligned}
&P(Y \in [y, y+\delta y), Q_0 = q) = \\
&= \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{l} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \text{ for } j \ne i \end{array} \right\}} \left[\begin{array}{l} e^{-\lambda_0 y} \frac{\lambda_0 (\lambda_0 y)^{q-1}}{(q-1)!} (q-1)! \frac{(\lambda_i/\lambda_0)^{\Delta_i - 1}}{(\Delta_i - 1)!} \\ \left[\prod_{j \ne i} \frac{(\lambda_j/\lambda_0)^{x_j}}{x_j!}\right] [r_i \delta y + o(\delta_y)] \end{array} \right] \\
&= f(y, q, \lambda_0) \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{l} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \text{ for } j \ne i \end{array} \right\}} (q-1)! \frac{r_i^{(\Delta_i - 1)}}{(\Delta_i - 1)!} \left[\prod_{j \ne i} \frac{r_j^{x_j}}{x_j!} \right] [r_i \delta y + o(\delta_y)]
\end{aligned}
$$

Dividing by δ_y and letting $\delta_y \to 0$ gives the result.

Proof of Corollary 6.4.1:

a) $P_{Q_0}(q)$ is obtained by integrating $f_{Y,Q_0}(y,q)$ over y.

$$
\begin{aligned}
P_{Q_0}(q) &= \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{l} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \text{ for } j \ne i \end{array} \right\}} (q-1)! \frac{r_i^{x_i}}{(x_i - 1)!} \left[\prod_{j \ne i} \frac{r_j^{x_j}}{x_j!} \right] I(q \ge \Delta_i) \int_{y=0}^{\infty} f(y, q, \lambda_0) dy \\
&= \sum_{i=1}^{N} \sum_{\left\{ \begin{array}{l} \sum_{j=1}^{N} x_j = q, x_i = \Delta_i \\ 0 \le x_j < \Delta_j \text{ for } j \ne i \end{array} \right\}} (q-1)! \frac{r_i^{x_i}}{(x_i - 1)!} \left[\prod_{j \ne i} \frac{r_j^{x_j}}{x_j!} \right] I(q \ge \Delta_i)
\end{aligned}
$$

b) Using part (a), one immediately sees that $f_{Y,Q_0}(y,q)$ can be written as:

$$
f_{Y,Q_0}(y,q) = f(y, q, \lambda_0) P_{Q_0}(q)
$$

$f_Y(y)$ follows from summing $f_{Y,Q_0}(y,q)$ over all possible values of q.

Proof of Corollary 6.4.2:

Using Lemma 5.1.1 and Corollary 6.4.1(b), we can write:

$$
\varphi(t, 0) = \overline{F}_Y(y) = \sum_{q=\underline{Q}_0}^{\overline{Q}_0} \overline{F}(y, q, \lambda_0) P_{Q_0}(q)
$$

Now, suppose that $k > 0$. A total of k units can be demanded with a minimum of $\lfloor k/\overline{Q}_0 \rfloor$ and maximum of $\lfloor k/\underline{Q}_0 \rfloor$ demand arrivals. Using Lemma 5.1.1 and Equation (6.10), we have

$$\begin{aligned}\varphi(t,k) &= \sum_{n=\lfloor k/\overline{Q}_0 \rfloor}^{\lfloor k/\underline{Q}_0 \rfloor} F_{Y^{,}Q_0}(y,k,\lambda_0) P_{Q_0}^{(n)}(k) \\ &- \sum_{q=\underline{Q}_0}^{\overline{Q}_0} \int_{y=0}^{t} F(t-y,k,\lambda_0) P_{Q_0}^{(n)}(k) P_{Q_0}(q) f(y,q,\lambda_0) dy\end{aligned}$$

The result follows after observing that $F(y, k+q, \lambda_0) = \int_{y=0}^{t} F(t-y,k,\lambda_0) f(y,q,\lambda_0) dy$.

Bibliography

[1] V. Agrawal and S. Seshadri. Distribution free bounds for service constrained (Q, r) inventory systems. Naval Research Logistics 47 (2000), 635 – 656.

[2] Y. Aksoy and S. Erenguc. Multi-item models with coordinated replenishments: A survey. International Journal of Production Management 8 (1988), 63–73.

[3] D.R. Atkins and P.O. Iyogun. A lower bound for a class of inventory/production problems. Operation Research Letters 6 (1987), 63–67.

[4] D.R. Atkins and P.O. Iyogun. Periodic versus can-order policies for coordinated multi-item inventory systems. Management Science 34 (1988), 791–796.

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

[6] S Axsäter. Continuous review policies for multi-level inventory systems with stochastic demand, vol. 4. North Holland, Amsterdam, The Netherlands, 1993, pp. 175–197.

[7] S. Axsäter. Exact and approximate evaluation of batch ordering policies for two level inventory systems. Operations Research 41 (1993), 777–785.

[8] S. Axsäter. Approximate evaluation of batch-ordering policies for a one-warehouse, N non-identical retailer system under compound poisson demand. Naval Research Logistics 42 (1995), 807–819.

[9] S. Axsäter. Exact analysis of continuous review (R, Q) policies in two-echelon inventory systems with compound poisson demand. Operations Research 48 (2000), 686–696.

[10] S. Axsäter. A note on stock replenishments and shipment scheduling for vendor managed inventory systems. Management Science 47 (2001), 1306–1310.

[11] S Axsäter and Rosling K. Notes: Installation vs echelon stock policies for multi level inventory systems. Management Science 39 (1993), 1274–1280.

[12] S. Axsäter and Juntti L. Comparison of echelon stock and installation stock policies for two level inventory systems. International Journal of Production Economics 45 (1996), 305–312.

[13] S. Axsäter and Juntti L. Comparison of echelon stock and installation stock policies with policy adjusted order quantities. International Journal of Production Economics 48 (1997), 1–6.

[14] S. Axsäter and Zhang W.F. A joint replenishment policy for multi-echelon inventory control. International Journal of Production Economics 59 (1999), 243–250.

[15] J.L. Balintfy. On a basic class on inventory problems. Management Science 10 (1964), 287–297.

[16] G. Cachon. Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. Operations Research 49 (2001), 79–98.

[17] G. Cachon. Managing a retailer's shelf space, inventory and transportation. Manufacturing and Service Operations Management 3 (2001), 211–229.

[18] S. Çetinkaya and H.B. Bookbinder. Stochastic models for the dispatch of consolidated shipments. Transportation Research Part B 37 (2003), 747–768.

[19] S. Çetinkaya, Tekin E., and C.Y. Lee. A renewal theoretic model for joint inventory replenishment and outbound shipment release decisions. Working Paper, 2005.

[20] S. Çetinkaya and C.Y. Lee. Stock replenishments and shipment scheduling for vendor managed inventory systems. Management Science 46 (2000), 217–232.

[21] S. Çetinkaya, F. Mutlu, and C.Y. Lee. A comparison of outbound dispatch policies for integrated inventory and transportation decisions. To Appear in European Journal of Operational Research, 2005.

[22] F. Chen and Y.S. Zheng. One warehouse multi retailer systems with centralized stock information. Operations Research 45 (1997), 278–287.

[23] K.L. Cheung and H. Lee. The inventory benefit of shipment coordination and stock rebalancing in a supply chain. Management Science 48 (2002), 300–306.

[24] K.L. Cheung and K.F. Leung. Coordinating replenishments in a supply chain with quality control considerations. Production Planning and Control 11 (2000), 697–705.

[25] A.J. Clark and H.E. Scarf. Optimal policies for a multi-echelon inventory problem. Management Science 6 (1960), 475–490.

[26] P.D. Cousin. Supply base rationalisation:myth or reality? European Journal of Purchasing and Supply Management 5 (1999), 143–155.

[27] D.R. Cox. The analysis of non-markovian stochastic processes by the inclusion of supplementary variables. In Proceedings of the Cambridge Philosophical Society (1955), vol. 51, pp. 433–441.

[28] Tekin E., Ü. Gürler, and E. Berk. Age-based vs. stock level control policies for a perishable inventory system. European Journal of Operational Research 134 (2001), 309–329.

[29] F. Erhun and S. Tayur. Enterprise-wide optimization of total landed cost at a grocery retailer. Operations Research 51 (2003), 343–353.

[30] A. Federgruen. Centralized Planning Models for Multi-Echelon Inventory Systems under Uncertainty, vol. 4. North Holland, Amsterdam, The Netherlands, 1993, pp. 175–197.

[31] A. Federgruen, H. Greoenevelt, and H. Tijms. Coordinated replenishments in a multi-item inventory system with compound poisson demands and constant lead times. Management Science 30 (1984), 344–357.

[32] A. Federgruen and Y.S. Zheng. The joint replenishment problem with general joint cost structures. Operations Research 40 (1992), 384–403.

[33] M.L. Fisher, J.H. Hammond, W.R. Obermeyer, and A. Raman. Making supply meet demand in an uncertain world. Harvard Business Review 72 (1994), 83–93.

[34] R. Forsberg. Optimization of order-up-to-S policies for two-level inventory systems with compound poisson demand. European Journal of Operational Research 81 (1995), 143–153.

[35] R. Forsberg. Evaluation of (R,Q)-policies for two-level inventory systems with generally distributed customer inter-arrival times. European Journal of Operational Research 99 (1997), 401–411.

[36] R. Forsberg. Exact evaluation of (R, Q)-policies for two-level inventory systems with poisson demand. European Journal of Operational Research 96 (1997), 130–138.

[37] R.Y.K. Fung, X. Ma, and H.C.W. Lau. (T,S) policy for coordinated inventory replenishment systems under compound poisson demands. Production Planning and Control 12 (2001), 575–583.

[38] B. Golany and A. Lev-er. Comparative analysis of multi-item joint replenishment inventory models. International Journal of Production Research 30 (1992), 1791–1801.

[39] S.H. Goyal and A.T. Satir. Joint replenishment inventory control. European Journal of Operations Research 38 (1989), 2–13.

[40] S. Graves. A multi-echelon inventory model for a repairable item with one-for-one replenishment. Management Science 19 (1971), 761–773.

[41] M.Ç. Gürbüz, K. Moinzadeh, and Y.P. Zhou. Replenishment and allocation policies for supply chains with crossdocking. Working Paper, 2004.

[42] G. Hadley and T.M. Whitin. Analysis of Inventory Systems. Prentice Hall, New Jersey, 1963.

[43] T.M. Harland. Supply chain management:relationships, chains and networks. British Journal of Management 7 (1996), 63–80.

[44] R. Hill. Lead time behaviour in inventory modelling. In Proceedings of the Fourth ISIR Summer School on Inventory Modelling (1999), pp. 119–125.

[45] E. Ignall. Optimal continuous review policies for two product inventory systems with joint set-up costs. Management Science 15 (1969), 278–183.

[46] S.G. Johansen and P. Melchiors. Can-order policy for the periodic review joint replenishment problem. Journal of Operational Research Society 54 (2003), 283–290.

[47] K. Katircioglu. Essays in Inventory Control. PhD thesis, University of British Columbia, 1996.

[48] G.P. Kiesmüller and A.G. de Kok. Safety stock allocation in a multi-item multi-echelon inventory system with quantity-based order consolidation. Working Paper, 2005.

[49] W.K. Kruse. Waiting time in a continuous review (s, s) inventory system with constant lead times. Operations Research 29 (1981), 202–207.

[50] H. Lee and K. Moinzadeh. Operating characteristics of a two-echelon inventory system for repairable and consumable items under batch ordering and shipment policy. Naval Research Logistics 34 (1987), 365–380.

[51] H. Lee and K. Moinzadeh. Two parameter approximations for multi-echelon repairable inventory models with batch ordering policy. IIE Transactions 19 (1987), 140–149.

[52] L. Liu and X.M. Yuan. Coordinated replenishments in inventory systems with correlated demands. European Journal of Operations Research 123 (2000), 490–503.

[53] P. Melchiors. Calculating can-order policies for the joint replenishment problem by the compensation approach. European Journal of Operations Research 141 (2002), 587–595.

[54] K Moinzadeh. An improved ordering policy for continuous review inventory system with arbitrary inter-demand time distributions. IIE Transactions 33 (2001), 111–118.

[55] K. Moinzadeh and H. Lee. Batch size and stocking levels in multi-echelon repairable systems. Management Science 32 (1986), 1567–1581.

[56] C. Nielsen and Larsen C. An analytical study of the Q(s,S) policy applied to the joint replenishment problem. European Journal of Operational Research 163 (2005), 721–732.

[57] B.Y. Özkaya, Ü. Gürler, and E. Berk. The stochastic joint replenishment problem: A new policy, analysis and insights. Accepted for publication in Naval Research Logistics, 2005.

[58] P. Pantumsinchai. A comparison of three joint ordering policies. Decision Sciences 23 (1992), 111–127.

[59] L.Y. Quyang, K.S. Wu, and C.H. Ho. Integrated vendorbuyer cooperative models with stochastic demand in controllable lead time. International Journal of Production Economics 92 (2004), 255–266.

[60] B. Renberg and R. Planche. Un modéle pour la gestion simultanée des n articles d'un stock. Revue Francaise d'Informatique et de Recherche Opérationelle 6 (1967), 47–59.

[61] K. Rosling. Optimal inventory policies for assembly systems under random demands. Operations Research 37 (1989), 565–579.

[62] S. Ross. Stochastic Processes. Wiley, New York, 1983.

[63] İ. Sahin. Regenerative Inventory Systems Operating Characteristics and Optimization. Springer-Verlag Bilkent University Lecture Series, Ankara, 1990.

[64] C. Schmidt and S. Nahmias. (S-1,S) policies for perishable inventory. Management Science 31 (1985), 719–728.

[65] H. Schultz and S.G. Johansen. Can-order policies for coordinated inventory replenishment with erlang distributed times between ordering. European Journal of Operations Research 113 (1999), 30–41.

[66] S.S. Sherbrooke. Metric: A multi-echelon technique for recoverable item control. Operations Research 16 (1968), 122–141.

[67] E.A. Silver. Some characteristics of a special joint-order inventory model. Operations Research 13 (1965), 319–322.

[68] E.A. Silver. Three ways of obtaining the average cost expressions in a problem related to joint replenishment inventory control. Naval Research Logistics Quarterly 20 (1973), 241–254.

[69] E.A. Silver. A control system for coordinated inventory replenishment. International Journal of Production Research 12 (1974), 647–671.

[70] E.A. Silver. Establishing reorder points in the (S,c,s) coordinated control system under compound poisson demand. International Journal of Production Research 19 (1981), 743–750.

[71] R.M. Simon. Properties of a two-echelon inventory model for low demand items. Operations Research 19 (1971), 761–773.

[72] S. Stidham. Stochastic clearing systems. Stochastic Processes and Their Applications 2 (1974), 85–113.

[73] A. Svoronos and P. Zipkin. Estimating the performance of multi-level inventory systems. Operations Research 36 (1988), 57–72.

[74] R.J. Tersine. Principles of Inventory and Material Management. North Holland, New York, 1994.

[75] R.M. Thompstone and Silver E.A. A coordinated inventory control system under compound poisson demand and zero lead time. International Journal of Production Research 13 (1975), 581–602.

[76] F.A. Van Der Duyn Schouten, M.J.G. Van Eijs, and R.M.J. Heuts. Co-ordinated replenishment system with discount opportunities. International Journal of Production Research 32 (1994), 2879–2895.

[77] M.J.G. Van Eijs. On the determination of the control parameters of the optimal can-order policy. Zeitschrift of Operations Research 39 (1994), 289–304.

[78] G.J. van Houtum and W.H.M. Zijm. On the relation between cost and service models for general inventory systems. Statistica Neerlandica 54 (2000), 127–147.

[79] S. Viswanathan. Note. periodic review (s,S) policies for joint replenishment inventory systems. Management Science 43 (1997), 1447–1454.

[80] Y.S. Zheng. Optimal control policy for stochastic inventory system with markovian discount opportunities. Operations Research 42 (1994), 721–738.