Analysing Service Level Agreements with Multiple Customers

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Osama Abdulaziz Almughamisi Alamri

Bachelor of Science, Mathematics (Taiba University)

Master of Science, Mathematics (Queensland University of Technology)

School of Science

College of Science, Engineering and Health

RMIT University

July 2017

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research programme; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Osama Abdulaziz Almughamisi Alamri

July 2017

Acknowledgements

In the name of Allah the most gracious and the most merciful

In praise is for Allah the almighty, who granted me the strength and the ability to proceed successfully.

This thesis appears in its current form due to the assistance and guidance of several people. I would therefore like to offer my sincere thanks and appreciation to all of them.

First and foremost, my sincere thanks go to my supervisors, Professor Panlop Zeephongsekul and Professor Babak Abbasi, for their extraordinary support and patience during my candidature, as well as for their insightful advice and valuable suggestions that brought this thesis to its current form. Without their assistance, advice and effort during the analysis and writing, this thesis would not have been achieved.

My heartfelt thanks also to Dr James Mines for his previous supervision. I would also like to thank Tabuk University for providing the scholarship funding that made my PhD study possible.

There are always people behind the scenes, and I would like to acknowledge them for their support and help. A special thanks to my wife Wejdan, and my kids Meshari, Gala and Abdulaziz. I appreciate them for their constant encouragement and support. Also, my sisters Rawya, Abeer and Noof and my big brothers Sami, Abdullah and Ahmed. Thanks for their patience and for giving me the potential to complete this research by creating a suitable study atmosphere.

My deepest gratitude to my father Mr Abdulaziz, and my mother Mrs Fozia, for everything they gave me. I seek for their forgiveness for neglecting their rights during my study overseas, and I hope they accept this thesis as a gift.

Contents

List of Tables

List of Figures

List of Abbreviations

CDF	Cumulative Distribution Function
DKF	Dvoretzky-Kiefer-Wolfowitz Inequality
FCFS	First Come, First Served
FR	Fill Rate
PBD	Performance-Based Contracts
pdf	Probability Density Function
PLFR	Prioritised Lowest Fill Rate
PS	Probability of Success
Q5	5th Quintile
SC	Supply Chain
SCM	Supply Chain Management
SD	Standard Deviation
SLA	Service Level Agreement
US	United States

Abstract

Within numerous production and distribution environments, maintenance of effective customer service is central to securing competitive benefits. Globalised industries are becoming more commonplace as well, further increasing the competitive pressure. Companies, as a result, are forced to expand product availability and deliver to the demand on schedule.

As part of a supply chain, service levels are an important measure of performance in operations management and are widely used to evaluate and manage supplier performance.

This thesis examines the SLA for the supplier under two types of contracts to guarantee the agreed customer service level. Specifically, this dissertation will shed light on the two most important (SLA) measurements for inventory systems: fill rate and ready rate. Both SLA measurements are commonly used as performance measures in SLAs between customers and suppliers. Throughout this thesis, we examine performance-based contracts in which the supplier has either: a single customer with a large demand, or multiple customers with a smaller demand. Our experiments were designed so that the demand distribution for the single customer case was similar to the aggregated demand distribution in the multiple customer case. The thesis primarily focused on four main questions, with each question being examined in its own chapter.

The first research problem is addressed in Chapter 3. Earlier studies of finite horizon fill rate only consider the situation in which there is a single customer in the supply chain. In Chapter 3, we develop a model to analyse the fill rate distributions for a supplier that has multiple customers, each with its own SLA. In particular, we examine the impacts of performance review period length and the correlation between customer demands on the average fill rate and the probability of overreaching the target fill rate when a supplier has multiple customers. Under the multiple customer contracts, two service policies for demand fulfilment. In the first policy, First-Come-First-Served (FCFS), demand is filled with no prioritization (e.g., in the case of two customers, there is a 50% chance that the first customer is served first). In the second policy, Prioritized Lowest Fill Rate (PLFR), customers are prioritized so that the customer with the highest

negative deviation from its target fill rate in the current performance review period is served first. The results and findings in Chapter 3 provide insights that can assist suppliers in the design and negotiation of SLAs.

The second research problem is addressed in Chapter 4. Previous studies on the finite horizon fill rate are limited and assume a zero lead time for the supplier. We create a model to examine the impact of different supplier lead times on the finite horizon fill rate, considering either single customer or multiple customers. As lead time exists in real-life supply chains, we explore the effect of various lead times on the fill rate distribution and required base stock over finite horizons with a variety of review period lengths. The results revealed that to fix the long-run fill rate, as the lead time increases, more stock is required; however, the probability of exceeding the target fill rate (the probability of success) increases as the lead time increases. The results indicate that the increase in the probability of success as the lead time increases is higher when the review period is shorter.

For the third research problem Chapter 5 presents further results related to the fill rate, an important measure of supply chain performance, specifically ensuring that a customer's service need is met with maximum reliability. These results mainly concentrate on variability, an aspect that is largely ignored in the literature on fill rate. Related results concerning consistency and asymptotic normality extend the range of application of the fill rate in evaluating reliability and determining the optimal stock level of a supply chain.

Chapter 6 explores the fourth research problem which considers the ready rate, a widely used performance measure in SLAs. The ready rate considered in this study is defined as the long-run fraction of periods in which all customer demand is filled immediately from on-hand stock. Previous studies of SLAs have been solely concerned with one supplier serving one customer, whereas in practice, a supplier usually deals with more than one customer. In multiple customer cases, the supplier has an SLA with each customer, and a penalty is incurred whenever the agreement is violated. In this chapter, we create a model to examine the impacts of various factors such as the base-stock level, the type of penalty (lump-sum and linear penalty), and the review period duration on the supplier's cost function when the supplier deals with multiple customers. The results show that dealing with more customers is preferable for a supplier (assuming the overall demand is the same) and that under a lump-sum penalty contract, a longer performance review is beneficial.

Finally, Chapter 7 closes with a brief review, discussion on the models constructed and suggests areas for future studies.

Chapter 1: Introduction

1.1 Introduction and Background

Services at logistics level significantly impact customer satisfaction, which in turn has a main effect on profits (Ghiana, Laporte & Mussammo 2004). In an increasingly globalised industry environment, it is challenging for companies to compete without strong inventory management strategies (Schwartz & Rivera 2010). A supply chain (SC) can take the form of goods or products, moving from manufacturers to suppliers to retailers, then finally to customers through a chain, to realise customer demand (Gong, Lai & Wang 2008). In recent years, supply chain management (SCM) has been of interest in the management field. The SC design process, is the organisation and combination of key business activities undertaken by an enterprise, from obtaining raw materials to the distribution of the finished product to the customer. Good SC design is a key driver of SCM performance and positive customer perceptions (Gupta & Maranas 2003; Sieke, Seifert & Thonemann 2012).

Today, in a global marketplace, companies providing excellent customer service will remain competitive (Larsen & Thornstenson 2008). As a result of competition, companies try to provide faster responses to customer requirements (Persson & Olhager 2002). Kumar and Sharman (1992) have stated: ‘We love your product, but where is it?’ The timely delivery of a product is often more important than the features of a product in determining customer satisfaction

1.2 Supply Chain (SC) Definition

Over the last decade there has been an increasing realisation of the importance of SCM. Operations management has been a focal area for many researchers. Several definitions of SC have appeared, yet it retains the same meaning with different terminology. Lee and Billington (1992) defined an SC: ‘As a set of relationships among suppliers, manufacturers, distributors, retailers and customers that facilitates the transformation of raw materials into final products.’ Ganeshan et al. (1999) defined an SC as: ‘A system of suppliers, manufacturers, and customers where materials flow downstream from suppliers to customers and the information flows in both directions.’

In real-life systems, an SC is a network of suppliers, manufacturers, warehouses and channels of distribution structured to first obtain raw materials, the raw materials are then transformed into completed products, which are sent to customers on time (Ganeshan et al. 1999). Strategic-level SC planning involves deciding the configuration of the network, that is, the number, location, capacity and technology of the facilities. Tactical-level planning of SC operations involves deciding the aggregate quantities and material flows for purchasing, processing and distribution of products.

1.3 Supply Chain Management (SCM)

Companies now pay attention to SCM, as effectively managing SCs can save them billions of dollars in inventory and logistics costs (Julka, Srinivasan & Karimi 2002). It is challenging to define the full scope of SCM, as it is a complex endeavour. Researchers and practitioners have examined the individual stages and individual functions of SCM, such as inventory control, manufacturing planning, product design, transportation and distribution, purchasing and marketing etc. individually. They found that optimizing individual elements of SCM often does not improve general performance, as the benefits and aims of an individual stage or function in the SC commonly conflict with further stages or functions. For example, many operations are prepared to decrease inventory levels to reduce holding costs, however this can increase the risk of a stock-out and affect distribution of products to retailers and customers. Now, however, increasing attention is being paid to the performance, design and analysis of the SC as a whole. (Beamon 1999, P. 281) defined SCM:

> “as an integrated process wherein, a number of various business entities (i.e., suppliers, manufacturers, distributors, and retailers) work together in an effort to: (1) acquire raw materials, (2) convert these raw materials into specified final products, and (3) deliver these final products to retailers”.

In SCM it is important to minimise costs while meeting a certain service level.

SCs are purposely designed and value-driven to maximise efficiency and achieve strategic priorities for chain members (Melnyk, Narasimhan & DeCampos 2014; Sieke, Seifert & Thonemann 2012). The resources that members are prepared to contribute to the chain are fundamental to its success. Melnyk, Narasimhan and DeCampos (2014) modelled an SC at three levels: influencers, which are the regulatory, economic and industrial opportunities and constraints; design decisions, which are the physical elements and the human resources along the chain; and building blocks or products, which are the inventory, transport, technology and capacity decisions that firms must make to operate.

SC agreements can be between two members of the chain or along the whole chain. Importers, manufacturers, distributors and transport firms can network through their industries as members of other chains, further complicating control. Therefore, contracts tend to be performance-based with a focus on outcomes. Service level agreements (SLAs) therefore contain targets, quantities, replenishment times and contingency methods (Halldórsson, Larson & Poist 2008; Liang & Atkins 2013). They may also contain penalties for non-performance or rewards for providing services within time, quantity or quality constraints (Hefley & Loesche 2010).

A standard SLA establishes objectives set by the customer. For the supplier of goods and services, non-performance or under-performance may have severe consequences (Selviaridis & Norrman 2015). For the purposes of this research, agreements include time of delivery, periodic review of performance and monitoring. Agreements are often established for the long term and thus include contract review dates for agreement conditions, and the resetting of financial penalties or incentives. However, there is increasing evidence that environmental and social responsibility measures are appearing as quality constraints for SC entities. Responding to European emission standards, Jaber, Glock and El Saadany (2013) developed a coordinated SC model based on suppliers and customers that included allowable gas emission rates along the chain. Within the model, Jaber, Glock and El Saadany (2013) compared the various permitted emissions trading schemes to establish levels before authorities financially penalised chain members. Yan, Chien and Yang (2016) produced a measure for environmental sustainability in a Taiwanese electronics chain, where members developed processes for reducing waste and achieving cleaner production along the chain. Further, Webster (2015) added reputational damage to firms who did not take international labour conditions into account along their SCs, such as Bangladeshi sweat shops. While these factors are beyond the focus of this research, they are important aspects of SC operations and point to the increasingly complex social and regulatory environments for producers, wholesalers and retailers. This thesis seeks to model SC agreements between one supplier and one or more customers, with constant demand and a finite horizon over variable lead times and variable review and replenishment periods. Measures are for fill rate, ready rate, and a required base-stock measure as developed from a model by Thomas (2005).

Established in the mid-20th century, SC research intensified due to the complexity of global trade and the increasing use of offshore manufacturing and assembly by countries that had previously been manufacturing nations. As the pace of international trade has grown, SCs diversified, first into industry-specific chains such as processed food, and then into time-sensitive chains, such as international relief organisations' responses to natural disasters (humanitarian chains). Later, computerisation and data management facilitated logistics for inventory in warehousing, transport and replenishment measures. Finally, online retail firms (such as Amazon) opened up the original catalogue concept into the new online retail industry, which perhaps represents the peak of SCM, where goods such as food can be delivered within an hour of ordering.

For goods retailers, supply management is a critical factor in business growth, and the success of a firm may depend on the strength of the agreements along the chain. This introductory chapter, therefore, begins with a description of the nature of SCs and their contracts (SLA), followed by an explanation of the elements of those contracts.

1.4 Supply Chain Management Benefits

From successful SCM, various benefits can be estimated:

- ***Output developments:*** these make it easy to coordinate materials, and avoid under-utilisation of capacity due to delays in delivery of parts.
- ***Decrease in inventory level:*** the visibility of demand and supply lowers the requirement to maintain high inventory levels to protect against uncertainty. The ability to understand when to buy products based on customer demand, allows for coordination with logistics planning.
- ***Optimised shipping:*** enhanced logistics coordination reduces the number of truck loads required, and directly lowers transportation costs.
- ***Increase in customer services:*** improving response to customer demand allows for on time product delivery.

1.5 Performance-Based Contracts

It is important to measure the service level because its association with managing of stock-levels, can affect performance and the resultant relationship with customers (Constantin, Ioan & Romania 2016). Services provided in SCs are numerous and often difficult to classify or measure, hence the need for a unified service level framework (Ellram, Tate & Billington 2004). To drive improvements in SC processes, performance must be measured. Further, if it is not possible to manage the process, performance cannot be measured (Neely, Adams & Kennerley 2002). In SC analysis and design, an important object is the establishment of convenient SLA performance measures. Performance measures in SLAs are applied to evaluate or compare existing systems, and are utilised to design proposed systems by setting the values of decision variables that introduce the best desired level of performance (Beamon 1999).

Recently, the design of performance contracts has drawn the attention of operations researchers. In contemporary business, performance-based contracts (PBD) are widely used because of their focus on outcomes rather than processes. In an SC with performance-based contracts, Liang and Atkins (2013) examined a customer SLA using principle-agent theory. Their results found benefits for the supplier from an SLA with strategic dynamic behaviour. Guajardo et al. (2012) also study performance-based contracts. They examine SC performance and compare the after-sales services between time-and-materials contracts and PBD. Also under PBD, Mirzahosseinian and Piplani (2011) inspected SC performance in relation to fixed-parts services.

By outsourcing support services, sustainment agreement is being redesigned using a unique PBD approach which controls agreements between suppliers and customers (Mirzahosseinian & Piplani 2013). In 1998, an examination of PBD application in the defence sector of the USA was initiated, with a team of 60, consisting of the Logistics Agency and the office of the Secretary of Defense, who assessed the transfer from traditional contracts to PBD. A target of 50% migration was set to PBD by the end of 2005 in the United States (US) Army, Air Force and Navy (Rievley 2001).

The most common type of PBD are SLAs, which evaluate supplier performance (Larson 2008; Liang & Atkins 2013). SLAs are established between a service provider and a client, based on the quality of service that the client expects from the service provider (Hefley & Loesche 2010). A survey by Oblicore (2007) found that more than 90% of organisations use SLAs to manage their suppliers. SLAs are mostly used to manage supplier performance in long-term business relationships. Typically, a performance (service level) target is set by the customer, and the supplier incurs a financial penalty if they fail to achieve this target. This type of contract can result in high penalty costs for underperforming suppliers of goods or services (Selviaridis & Norrman 2015). To realise a preferred performance outcome, an appropriate motivation should be in place to encourage the supplier (Mirzahosseinian & Piplani 2013).

In SLAs, the performance desired by the customer is recognised and a related target service level is specified. The US Office of Federal Procurement Policy described PBD as 'an acquisition structured around the results to be achieved as opposed to the manner in which the work is to be performed (on Performance-Based Service Acquisition 2008)', (see, e.g., Acquisition Central website).

1.6 Service Level Agreement (SLA)

SLAs are one of the most common types of contracts used to manage supplier relationships (Larson 2008; Liang and Atkins 2013). For instance, *Hewlett-Packard and UPS* Supply Chain Solutions are applying an SLA to improve management of their partnership (Lewis et al. 2007). In industrial companies, suppliers and retailers normally decide on a service level that a supplier is expected to realise (Thonemann et al. 2005). Typically, a performance (service level) target is set by the customer, and the supplier incurs a financial penalty if they fail to achieve the target. This type of contract can result in high penalty costs for underperforming suppliers of goods or services (Selviaridis & Norrman 2015). For example, in Indonesia, penalties for failing to deliver coal on time were costing the Nobel Group tens of millions of dollars each year (Fragkos & De Reyck 2016). Alternatively, an SLA may specify a bonus for meeting and/or exceeding the performance target, and service levels can also form part of a scorecard used to evaluate supplier performance (Katok, Thomas & Davis 2008). In a survey of 190 manufacturers, suppliers and retailers, the common response of those questioned was that improving their service level was the principal goal of their inventory management (Kanet, Gorman & Sto¨ßlein 2010). In the 1990s, many manufacturers were required to cooperate with their suppliers to improve their organisational performance and competitiveness (Ittner et al. 1999; Shin, Collier & Wilson 2000).

1.7 Inventory Management Performance

This section identifies the measures used in this research.

1.7.1 Performance Measure

Fill rate: This refers to inventory levels at a distribution point in the supply chain. Disney et al. (2015, p. 501) define the item fill rate as 'the proportion of demand fulfilled directly from inventory'. The inventory fill rate is a measure that has been used extensively over time, and is popular in the management press (Johnson & Scudder 1999; Lee & Billington 1992; Tempelmeier 2000). This measure is calculated over a time interval, the replenishment review period, and is discussed further in the literature review.

Ready rate: The ready rate was established by Schneider (1981) 'as the long-run average cumulative satisfied demand per replenishment cycle divided by the average demand per replenishment cycle' (Goetschalckx 2011, p. 438). The ready rate has variously been described by Larsen and Thorstenson (2014, p. 13) as 'the fraction of time during which the on-hand stock is positive', and Axsäter (2006), who used the term 'net inventory level' instead of 'on-hand stock'. For the purposes of this research, Liang and Atkins' (2013, p. 1104) explanation is used: 'the long run fraction of periods that demands are filled from stock, which is equal to 1–stock-out rate'. This is further discussed in the literature review.

1.7.2 Important parameters in Inventory Management

Replenishment review period: Depending on the service or product, the review period in the service contract may be a day, a week or a year, but is generally monthly or quarterly (Syntetos, Babai & Gardner 2015). Further, SC agreements may be established for short-term (finite planning horizon), or long term (infinite planning horizon) (Zhang 2013).

Lead time: There are several interpretations of lead times. Christopher (2016) stated that the customer's view of lead time is from order to delivery. From the supplier's viewpoint, lead time may extend from the time when working capital is committed, to when the customer's cash is received. In this research, lead time is defined as the delay between the time customers place an order and the time they receive it. Rushton, Croucher and Baker (2014) explained that lead times can vary according to the environment; logistics lead times include manufacturing and supply of a product; customer's order cycle time refers to the period that customers are willing to wait for their goods. SC lead time refers to the period between customer order and delivery time. The lead time gap is the point where inventory must be held. Lead times are further discussed in the literature review.

Stock-out event; backorder: Stock-out occurs when there is unexpectedly high demand, perhaps with a longer lead time. Backorder may mean a delay to a customer until the next delivery, but it may also mean lost sales, especially in retail or food chains. Jain, Rudi and Wang (2014, p. 134) advised that 'optimal order quantity with timing observations is greater than the optimal order quantity with full demand observations'. Table 1.1 shows an inventory review result using these concepts.

Table 1.1: Sample of inventory review

Order cycle	Demand (units)	Stocked-out units
1	120	0
2	120	0
3	120	0
4	120	30
5	120	0
6	120	0
7	120	0
8	120	80
9	120	0
10	120	0

Table 1.1 shows 10 cycles with expected demand that resulted in a stock-out event in each of two cycles. Demand over the replenishment review period was 1200 items, and demand for 110 items was unfulfilled, thus the fill rate was as = $\frac{1090}{1200}$ = 0.91, or 91%. The ready rate for the 10 cycles showed stock-out for two periods, so was = $\frac{8}{10}$ = 0.8 or 80%. The base stock policy occurs when the supplier replenishes to a level stipulated in the agreement each order cycle. In a typical agreement, the supplier is responsible for reaching a target fill rate or ready rate, as measured at each performance review point.

While agreements contain idealised conditions, these rarely occur in practice and anomalies occur. Suppliers with a long-run or infinite horizon contract may overfill an agreement to ensure compliance. In the example, if a supplier must exceed a 90% fill rate, a 93% long-run fill rate may be used as a buffer against penalties. It well known that base stock inventory policies minimise holding and shortage costs for an integrated SC (Zipkin 2000).

1.8 Performance Incentives and Penalties

In most SLAs, the partners agree that there is a target that should be achieved by the supplier, or else the supplier can face a financial penalty. Financial incentives and penalties for SLAs are typically linear (a rate) or fixed (an amount) and both forms are commonly used in service agreements (Larsen & Thorstenson 2014; Sieke, Seifert & Thonemann 2012). For example, Fragkos and De Reyck (2016) reported that in delivering coal outside of the contract conditions, the Nobel Group in Indonesia was losing a significant amount of money each year.

The first type of penalty is a lump sum, in which a supplier incurs a penalty if the supplier's review phase performance is beneath the performance threshold. The second type is a linear penalty SLA, where the supplier incurs a linear penalty of the quantity of deviation from a performance threshold. For instance, the Acquisition Central website (on Performance-Based Service Acquisition 2008); describes two types of SLA penalty:

- ***Example 1.*** 'The firm-fixed-price for this... shall be reduced by 2% if the performance standard is not met' (a lump-sum penalty).
- ***Example 2.*** 'For each 5% degradation in... Performance observed..., the firm fixed- price... will be reduced by 1%' (a linear-based penalty).

For example, the large US retail chain CVS imposes a fixed penalty on suppliers who fail to supply at least 92% of purchase transactions (orders) within one month (CVS, 2013). Similarly, Dm-Drogeriemarkt (one of the biggest drugstore chains in Germany) applies fixed financial penalties to suppliers who do not achieve pre-specified thresholds (Mostberger 2006). Thonemann et al. (2003) found that in Germany, 70% of retailers measure the performance of their suppliers based on service levels. KPMG (2010) identified food, drinks and consumer goods companies (FDCGs) as frequently having service level commitments with retailers regarding product availability, the value of the product and on-time delivery. Toys “R” Us, on the other hand, applies a linear penalty that increases as the realised fill rate decreases (Toys “R” Us 2013). Drugstore chain Rite Aid is an example of an organisation with a mixed penalty system, consisting of a fixed flat fee and a linear percentage-based penalty (Rite Aid, 2014). Supplier SLAs that incorporate either fixed (flat) or linear penalties can affect a supplier’s base stock level decision (Liang & Atkins 2013). Sieke, Seifert and Thonemann (2012) show that SLAs with both penalties (flat or unit penalty), and can organise the SC through supplier and manufacturer. Holmes and France (2002) reported that, in 1997, late delivery resulted in enormous fees for the Boeing Company. Because of delivery delays, the company lost customers to Airbus. Similarly, in the US, grocery chain Kroger applies a flat penalty every time a supplier cannot deliver a complete order on time (Kroger, (2011)). An observed study by Bensaou (1999) found that Japanese companies often divide their purchases between multiple suppliers, then subsequently use service level performance measures to evaluate and select the suppliers for long-term relationships.

Considerable research has been done on the effect of performance on supply chains using either incentives or penalties, or both. Liang (2009) stated that ‘Monetary compensation is used for an incentive’ . Returning to the social responsibility and firms’ reputations of social responsibility, Porteous, Rammohan and Lee (2015) noted the increasing potential for supply chain disruption through regulatory or social media influences. They set a model of incentives and penalties for suppliers’ social and environmental transgressions, which included customers’ operational costs. Incentives proved more efficient among 334 US companies and 17 industries with reduced supplier violations and reduced customer costs. Incentives were not necessarily financial; they included environmental training for staff and offers of increased business.

In assessing the efficiency of using performance incentives or penalties, Aktas and Ulengin (2016) established a service agreement with a durable goods firm and its logistics provider. A newsvendor model was used, whereby the state of the inventory cannot be known until the customer places an order; thus, the order is random, the inventory is established before the order is made and economic results are bound by insufficient inventory or oversupply. They found that use of the reward and penalty system enhanced Supply chain performance, reaching 96.1% Aktas and Ulengin (2016). Alwan et al. (2016) noted that the newsvendor model, in which demand is assumed to be independent from one period to the next, proved to be superior to the exponential smoothing method and moving-average method under certain conditions.

Less research has been conducted on the effects of penalties alone on SC performance. Qi, Ni and Shi (2015) used game theory to explore consumers' search behaviours when confronted with a stock-out (backorder). Using one distributor and two retailers, Qi et al (2015) found that if the wholesale price is exogenous, their model predicted more retailers and higher chain efficiency; if the wholesale price was endogenous, the opposite occurred, as the wholesaler could raise its prices. In a stock-out situation in which demand was such that the wholesaler was unable to supply the retailers, the penalty was that the chain was confronted with higher competition and reduced influence. That is, unless SC members develop the capacity to compete, the chain itself may disintegrate.

1.9 Inventory and Demand Fulfilment Policies

Policies refer to the business models that firms use, and the policies and practices they employ to achieve their objectives. Policies are therefore part of the assumptions in SC modelling.

Base stock policy: Base stock inventory policies minimise storage costs for an integrated SC (Song, Dong & Xu 2014). Tempelmeier and Bantel (2015) stated that an inventory base stock policy is commonly used in modelling and practice. A base stock policy has random demand and is filled one unit at a time; times between orders are independent and identically distributed, stock is constantly reviewed and backorders occur when stock is out.

First come, first served policy: demand is filled with no prioritization (e.g. in the case of two customers, there is a 50% chance that customer 1 is served first). Also, known as 'queue state dependent order acceptance policy', order receipt is random, which has an effect upon stock-out. Orders cannot be stratified by the priority of customer status, or item quantity if it is available, so other customers may miss out if replenishment of inventory does not occur in time.

Prioritised lowest fill rate: This refers to prioritising customers based on current measured fill rate performance. Prioritised Lowest Fill Rate (PLFR) customers are prioritised such that the customer with the highest negative deviation from their target fill rate in the current performance review period is served first. For example, if service level is reviewed every 5 days, and on day 4 the supplier faces a shortage and is unable fill the demand of all customers, then the fill rate of each customer for the past three days is calculated, and customers are prioritised for service according to the difference between their computed fill rate (for the last three days) and their target fill rate. Note that the PLFR policy is not necessarily optimal, as the optimal policy must be designed based on the penalty structure outlined in the SLAs.

1.10 Research Scope and Objectives

This research aims to evaluate the performance of an SC with one supplier and one or more customers. The performance is measured using fill rate, and ready rate. The models have static demand across replenishment review periods (finite horizon). There are two policies used (see Chapter 3): FCFS and PLFR. Other chapters focus only on the FCFS models. Due to issues with multiple parameters, the models are based on simulation approach.

Objective 1: The first objective is to model fill rate distributions for a supplier with a base stock policy and a single customer, and then to compare this to multiple customers each with their own agreement. Total demand in all scenarios is constant; The first objective is then extended by using a pooled inventory, multiple-customer model, with varying review period lengths. The effect of correlated customers' demands are then considered and tested for the probability of overreaching the target fill rate. Service levels for a single supplier are examined for customers with different levels of demand.

Objective 2: The second objective is to investigate the effect of various lead times on the fill rate performance of a base stock model system. This thesis explores cases with one customer with high demand and multiple customers with small demands.

Objective 3: The third objective is to assess the variability of these results on consistency and asymptotic normality, extending the range of this measure to determine an optimal stock level and evaluate the reliability of the SC.

Objective 4: The fourth objective is to model ready rate distributions for a supplier with a base stock policy and a single-customer case, and then to compare this to multiple customers each with their own agreement. Then we examine the impact of the base stock level, the type of penalty (lump-sum and linear penalty) and differences in ready rate thresholds on the supplier's costs. This is tested with single and multiple customers.

Objective 5: The last objective is to identity possible future research directions that may improve and build on the work in this research.

1.11 Contribution, Significance of Thesis

It is well known that supply chain management is regarded as an essential element for success and customer satisfaction in most industries and businesses. In a supply chain, reaching customer satisfaction is dependent upon measuring and evaluating the performance of supplier with customers, and the two most important measures of this evaluation are the fill rate and ready rate. However, earlier research of finite-horizon fill rate and ready rate only inspect situations involving a single customer in the supply chain, whereas in practice, a supplier often deals with many more. Taking this perspective, this thesis introduces new models under one and multiple customers as well as under fill rate and ready rate measures to fully evaluate supplier performance.

The components of this dissertation are as follows:

1) Chapter 3 introduces two service policies for demand fulfillment to investigate finite-horizon fill rate behaviour when there are multiple customers for a supplier that uses a base stock policy. In this situation, customers' demands are fulfilled from a pooled inventory according to an established service policy. Therefore, the base stock policy for a supplier with multiple customers signifies that there are stock level (S) items available in each replenishment period to fulfil the demand of all customers. The inventory is replenished to S at the beginning of each replenishment period. Altering the time frame of the performance review period and tracking the correlation between customers' demands allows for an examination of the changes in the average fill rate and the probability of exceeding the target fill rate of each customer. The findings in this thesis provide insights that can assist suppliers in the design and negotiation of SLAs. The results of this model have been published in Omega Journal.
2) The thesis explores the impact of lead times on the finite-horizon fill rate of a base-stock inventory system. Specifically, new models were developed to examine the effect of lead time on the distribution of the fill rate over finite horizons with various review period lengths for single and

multiple customers. Findings suggest that when the lead time increases, the supplier needs to stock more units to reach a defined average fill rate; however, an increase in the lead time leads to a higher probability of exceeding the target fill rate when the expected fill rate is fixed. The insights provided in Chapter 4 can have a significant influence on crafting an SLA and assist suppliers and customers both in deciding on the review period length as well as bonus and penalty amounts. The results of this model have been published in the Manufacturing and Service Operations Management (MSOM) conference, New Zealand.

3) The third component concerns reliability measures for the fill rate so as to assess consistency and asymptotic normality and determine an optimal base stock level for single or multiple customers. Reliability was established with a set of equations, numerically tested first for one customer, and then extended to two customers. Results revealed that the upper bounds for reliability increased with increasing performance review duration and decreased as confidence levels rose, reaching symmetry over the long-run contract FR. Utilising these findings, the optimal stock levels for reliability in the supply chain were established. The results of this model have been published in the International Society of Science and Applied Technologies (ISSAT) conference, New Jersey.

4) Ultimately, a new model was developed to evaluate ready-rate behaviour in a multiple customer setting. The thesis examines the impact of contributing factors, such as the base stock level; two important forms of SLA penalties,lump-sum and linear; and the review period duration on the supplier's cost function when the supplier deals with single and multiple customers. Given that the ready-rate is a random variable, a numerical approach is employed to systemically investigate cost outcomes under different SLA contracts. The findings provide useful insights that can assist suppliers in the design and negotiation of future SLAs and in devising strategies for achieving compliance, thus avoiding penalties once

an SLA is in place. The results of this model have been published in the Operational Research Society Journal.

1.12 Main Research Questions

This research examines the performance of supplier for both measures (fill rate and ready rate) over different review period phases, when suppliers faced two different types of SLA contract (a single customer with a large demand, and multiple customers with a smaller demand, in which both cases have the same demand).

Here, we present main research questions for this thesis, whereas, each chapter has numerous of sub-questions.

Chapter 3 will answer the first two questions, related to the average fill rate measure for a single and multiple customers; the answer to the second question tries to explain the different stocking levels required for the supplier to achieve the fill rate target:

- What is the change in the average fill rate when the review period is increased for single and multiple customers?
- What are the changes in the average fill rate and the probability of exceeding the target fill rate of each of multiple customers when compared to a single customer?

These questions examined the fill rate measure for a supplier with zero lead time. Chapter 4 will address the next question, which investigates the impact of lead time for the supplier and how many units should be stocked to achieve the fill rate target.

- What is the impact on base stock levels of single and multiple customers' cases over various review lengths and lead times?

Chapters 3 and 4 are primarily concerned with fill rate—assessing its mean value and simulating sampling distribution under various supply and demand regimes. However, Chapter 5 considers the variance of the fill rate and uses this to consider its consistency as an estimator as well as its asymptotic normality. Results are then extended to multiple customers.

The previous questions were concerned about fill rates that SLAs measure. The next questions, addressed in Chapter 6, examine other important measures:

- What base stock level is required to minimise the supplier's expected cost?
- How is the supplier's expected cost affected by differing performance review phase durations?

1.13 Thesis Structure

This thesis is organised into seven chapters. Chapter 1 has introduced the background of the research. Chapter 2 is a literature review on the fill rate, ready rate lead time and simulation analysis. It explains the model selected for the research.

Next, Chapter 3 examines the impact of the performance review period duration on the shape of the fill rate distribution, which is not well understood. Past studies on finite horizon fill rate only consider the situation where there is a single customer in the SC. In Chapter 3, we analyse fill rate distributions for a supplier that exercises a base stock policy and has multiple customers, each with their own SLA. In particular, we examine the impact of performance review period length and correlation between customers' demands on the average fill rate, and the probability of overreaching the target fill rate, when a supplier has multiple customers and their demands are fulfilled from a pooled inventory. Further, we explore the realised service level for customers with different levels of demand dealing with a single supplier. Customers' demands on the average fill rates are then correlated and tested for the probability of overreaching the target fill rate (0.95); these are compared to independent customers and a single customer.

Chapter 4 investigates the impact of different supplier lead times on the finite horizon fill rate, considering one customer or multiple customers. Previous studies on the finite horizon fill rate are limited and assume a zero lead time for the supplier. Because lead time exists in firms, we study the effect of various lead times on the fill rate distribution and base stock over finite horizons with a variety of review period lengths.

Chapter 5 moves to the third objective, we will consider the variance of the fill rate and use this to consider its consistency as an estimator as well as its asymptotic normality. We also extend the results to multiple customers but will concentrate mainly on the case of two customers since extension beyond this case is straightforward but algebraically cumbersome. Measures of supply chain reliability for both the single and multiple customers are introduced, and finally, several numerical examples will be provided to illustrate the foregoing results.

The fourth objective is to evaluate ready rate performance in a multiple customer setting. In Chapter 6, previous studies in ready rate SLAs performance have been solely concerned with one supplier serving one customer, whereas in practice a supplier usually deals with more than one customer. In multiple-customer cases, the supplier has an SLA with each customer, and a penalty is incurred whenever the agreement is violated. In Chapter 6, we examine the impact of various factors, such as the base stock level, the type of penalty (lump-sum or linear) and the review period duration on the supplier's cost function when the supplier deals with multiple customers. Finally, Chapter 7 includes a review, a discussion on the models constructed and suggests areas for future studies.

1.14 Publications

The following are publications based on the contents of this thesis:

Conference proceedings

- O Alamri, B Abbasi & Z Hosseinifard 2016, 'The impact of lead time on the finite horizon fill rate in single and multiple-customer cases', 6 July, *MSOM Annual Conference*, Auckland University, New Zealand.
- P Zeephongsekul, O Alamri & B Abbasi 2016, 'Fill rate: ensuring reliability in a supply chain', H Pham (ed), *Proceedings of the 22nd ISSAT International Conference Reliability and Quality in Design*, *International Society of Science and Applied Technology*, New Jersey, pp. 1–5.

Conference presentations

- B Abbasi, Z Hosseinifard, O Alamri, D Thomas & J Minas 2016, 'Service level agreement between a supplier and multiple customers', December, *NZSA-ORSNZ Annual conference*, Auckland University, New Zealand.
- O Alamri, B Abbasi, JP Minas & P Zeephongsekul 2016, 'Comparing ready rate performance for a supplier in single and multiple customer cases', 6 May, *POMS Annual Conference*, Orlando, FL.
- B Abbasi, Z Hosseinifard, O Alamri, D Thomas & J Minas 2015, 'One customer with large demand or multiple customers with smaller demands: a service level agreement perspective', 10 May, POMS Annual Conference, Washington, DC.

Journal articles

We submitted four journal papers, of which two are under review and two has been accepted.

- O Alamri, B Abbasi, JP Minas & P Zeephongsekul 2017, 'Service level agreements: ready-rate analysis with lump-sum and linear penalty structures', *Journal of the Operational Research Society* (forthcoming).

- O Alamri, B Abbasi & Z Hosseinifard 2017, 'The impact of lead time on the finite-horizon fill rate in the case of a supplier with multiple customers', *Australasian Journal of Information Systems* (under review).
- B Abbasi, Z Hosseinifard, O Alamri, D Thomas & J Minas 2017, 'One-customer with large demand or multiple-customers with smaller demands: an SLA perspective', *Omega Journal* (forthcoming).
- P Zeephongsekul, O Alamri and B Abbasi 2016, 'Fill rate: ensuring reliability in a supply chain', *ISSAT Journal* (under review).

Chapter 2: Background and Literature Review

2.1 Introduction

Performance is critical in SCs, and the dynamic nature of supply has resulted in many possible directions for research since the field developed in the 20th century (Schneider 1981). In inventory control, where undersupply results in penalties and oversupply inflates costs, research attention has been directed towards service constraints, SC optimisation, production performance measures and monitoring inventory control (Beamon 1999). This research investigates suppliers' inventory service level measures, established by Schneider (1981) as the fill rate or ready rate, which are measured over either a finite or infinite horizon. Over time, these terms have diverged; for example, Liang and Aitkens (2013) use a ready rate or "stock-out rate" as measures for an inventory optimisation model. Thomas (2005) investigated the effect of varying the length of the performance review phase on both the optimal base-stock level and on the fill rate distribution. These elements are the focus of this literature review, as well as fill rate, ready rate, and lead time. In this chapter, an overview of the literature on fill rate and ready rate is provided but technical review related to the topic of each chapter is provided in the related chapter.

2.2 Fill Rate

For inventory measures, a periodic replenishment review inventory system, the 'traditional' or standard approach for fill rate is to assess stock levels by unfilled orders and approximate future demand. In calculations, high standard deviations (SD) for this system lead to a negative value for fill rate. Second, total shortages are assessed for individual periods. Thus, shortages in series (continuous review) lead to an overestimation of the actual shortages (backorders), therefore underestimating the actual system fill rate (Nahmias & Olsen 2015; Vollmann, Berry & Whybark 1997). To address these anomalies, Hadley and Whitin (1963) used a Poisson-distributed random demand function. In this method, the expected shortage of units each period equals the average quantity that demand exceeds the base stock level with the lead time, plus the review period, less the average quantity that demand exceeds the base stock level with the lead time. This approach is accurate when the review period is smaller than the lead time and there is a small SD in demand.

Several attempts have been made to address the underestimating and overestimating anomalies in stochastic inventory systems with service level constraints. Johnson et al. (1995) observed that the traditional expression for line item fill rate, once under 95%, underestimates the exact fill rate and this underestimation becomes significant below 90%. Extrapolating from Hadley and Whitin's (1963) inventory system, Johnson et al. (1995) provided an exact expression for the fill rate in a periodic inventory system to account for high demand variability and customer returns. Tempelmeier (2007) argued that as well as the fill rate, the extent of the shortfall (backorder) in supply is important. Adding a new measure, Larsen and Thorstenson (2014) tested the customer order fill rate with order fill rate and volume fill rate, and found the new measure superior. Guijarro, Cardós and Babiloni (2012) approximated the fill rate, outperforming Johnson et al.'s (1995) estimation when there is probability of zero demand. Moving from the product, Milner and Olsen (2008) and Hasija, Pinker and Shumsky (2008) investigated supply agreements for call centres and estimated shortfall as percentages of delay time for waiting customers.

In a restricted supply space, Mapes (1993) determined the service level fill rate of a restricted capacity periodic review safety equipment system to provide an agreed service level. Dubois, Allaert and Witlox (2013) used retail for their calculation of the fill rate of a periodic review order-up-to inventory system with capacitated replenishments, lost sales and zero lead time. Again in retail, Wan, Evers and Dresner (2012) examined the impact of product variety on fill rate, finding that it decreased by increasing the product variety; however, as product variety increases, the rate of decrease diminishes.

In the standard periodic review, the interval of review is one unit of time; in an actual review the interval is ($R \geq 1$). In calculating fill rate, multistage models could assist extended global SCs. Radner (1985) initially used multiple periods in a repeated principal-agent game. Radner (1985) found that the principal's decision on an agent's action for one period could affect performance on a second game, as both principal and agent became more efficient (skilled). These procedures were independent and identically distributed. Choi, Dai and Song (2004) examined contract performance for a supplier; next, a buyer firm was contracted to an end customer. Choi, Dai and Song (2004) argued that the end customer cannot be guaranteed delivery because of variables in replacement of inventory, despatch or receipt and storage issues by either the supplier or the distributor. Given this risk, contracts can be based on long-run expectations of inventory availability. Zhang and Sobel (2012) solved the fill rate problem under single and multistage period reviews, concluding that higher fill rates occur with shorter SCs. Zhang (2010) also offered an exact expression for the fill rate for a general periodic two-stage inventory system, expanding the expression to N inventory system stage, optimising the standard SC performance.

Using a base stock policy, Chen, Lin and Thomas (2003) showed that the expected value of the fill rate for a finite horizon is higher than that for an infinite horizon. They also proved that the expected fill rate value is highest when the length of the finite horizon (review period) is one. Larsen (2011) showed that for an infinite horizon, the estimation of the (volume) fill rate is more accurate than estimating the order fill rate. However, for a finite horizon model, the estimation of the order fill rate is more accurate than estimation of the (volume) fill rate. Thomas (2005) conducted a simulation study considering one customer in the SC, an Erlang distribution for the demand and a static periodic review base stock model to examine the behaviour of the fill rate distribution when the performance review period changes. He found that the average fill rate decreases when the performance review period length is increased.

To minimise holding costs, Boyaci and Gallego (2001) and Shang and Song (2006) developed infinite horizon continuous review inventory models with fill rate constraints. Shang and Song derived a closed-form approximation for an optimal base stock model, imputing a shortage cost as the fill rate constraint. Banerjee and Paul (2005) found the average fill rate decreased when the review period increased in a base stock inventory system. Further confirmation came from an inventory experiment to determine optimal stock levels by Katok, Thomas and Davis (2008), used a controlled laboratory setting to study both performance review period length and magnitude of the bonus for meeting or exceeding the service level target in a supply chain with a single customer. They concluded longer performance review periods may be more effective than shorter ones because of enhanced feedback and the opportunity to improve service.

Using the reward and penalty performance variables, Yin and Ma (2015) considered fixed and linear bonuses for an SC of two, a manufacturer and a retailer. They showed the retailer could achieve higher service level and higher profits with a fixed bonus rather than a variable bonus, if their results were based on modelling a single performance review phase. Conversely, Liang and Atkins (2013) found that penalties that aligned closely with performance were preferable to either bonuses or set penalties. Following the penalties finding, Bijvank et al. (2014) used a lost sales approach, where a missing retail item is substituted and not purchased later. They tested three approaches and found Archibald's (2007) solution optimal. The studies reviewed in this section examined the problem of a fill rate with a single customer, and assumed zero lead time.

2.3 Ready Rate

There has been some research on the ready rate in the early literature, discussing its part in stochastic modelling (Feeney & Sherbrooke 1966; Rosling 2002). In a US Navy context, Silver (1972) used the term to describe the inventory of useable spares for assembly. Others have assumed a meaning similar to the fill rate (Chen & Krass 2001; Schneider 1981). Larsen and Thorstenson (2008, 2014) consistently use the measure in their modelling, and Liang and Atkins (2013) used ready rate in their study. Other authors that have used the ready rate as a standard measure are Srivathsan and Kamath (2012), Rossetti et al. (2013) and Rossetti and Xiang (2010). Rossetti and Xiang (2010) commented that if the distributor experiences unpredictable (lumpy) demand and reorder quantities vary, a Poisson process would probably not apply and the ready and fill rate could markedly differ.

Wang, Chen and Feng (2005, p. 667) define the time-window ready rate as 'the percentage of periods in which orders are completely fulfilled within a pre-specified window'. Wang, Chen and Feng (2005) used a single-item single-location min-max (s, S) inventory system due to the complexity of their modelling, finding that changing the dimensions of the time-window to have considerable impact on inventory costs. Liang and Atkins (2013) confirmed that 'time-window fulfilment' was commonly used by industry, although in logistics rather than inventory.

2.4 Lead Time

Fulfilling customers' orders promptly is a measure of a firm's efficiency and productivity. A stock-out may not only cancel the order but also it will affect the probability of future customer demand (Bertazzi et al. 2013). In inventory management studies, lead time has been a point of interest (Priyan & Uthayakumar 2014). Heydari (2014) claimed that ordering delay, transport or information issues between line members may disrupt the entire chain. Inappropriate lead time planning can therefore disrupt customer service or cause costly overstocking (Louly & Dolgui 2013).

Lead time has a direct impact on the performance of inventory control systems (Berling & Marklund 2014). Lead time, the duration between order and delivery, may refer to the part of the SC or all of it. Li and Liu (2013) noted that the length of lead time can be affected by size and batching of orders, transport distances and delays, pricing prioritisations and adequate communications. Minimising lead time for suppliers can reduce high inventory levels and increase customer satisfaction.

In SC operations lead time can have variety of effects on its members, ranging from lead times that are too long, too short, or to unpredictability in inventory control. Heydari (2014) reported on a retailer experiencing unpredictable delivery times. A model of a supply agreement was constructed that, as an incentive to the customer, stabilised lead time fluctuations. Improving transport (shipping) reliability leading to allow for SC continuity and improve its profitability.

Disturbances along the SC lead to increased lead times. Spiegler, Naim and Syntetos (2016) developed a chain model, testing various disturbances. As the resulting model became nonlinear, Spiegler, Naim and Syntetos (2016) used simulation and nonlinear control theory to establish that inventory and order rate, as well as work-in-progress, were input dependent. As the lead time increases, the inventory level should be increased to compensate the inventory orders (the customer's demand). Expectations were that as lead times decreased, services improved, the chain became more efficient, flexibility improved and costs decreased. However, when the lead time decreases abruptly, Speigler, Naim and Syntetos (2016) found that operations along the chain were disrupted, while oversupply of inventory was reduced. The authors cautioned that reduced lead times should be stepped so that the chain nodes could adapt to the new rate. Muckstadt and Sapra (2010) advocated for establishing continuous reviews of lead times.

In modelling, stochastic lead time relates not to a precise period, but to a statistically probable period (Wang & Disney 2017). An assumption of a constant lead time refers to classic safety stock policy (Kouki, Jemaï & Minner 2015). Ren, Li and Che (2016) modelled key points along a SC of perishable goods, established an inventory level to minimise waste and a change point. Stock-out occurs after the change point comprised of the retailer's inventory and ordering costs, and this is further reliant on the length of the lead times, particularly for perishable goods.

2.5 Simulation and Analysis

Simulation can be useful to identifying options when designing SCs in a complex and uncertain decision environment for suppliers and customers. To estimate the fill rate along a SC with a manufacturer or distributor and one or more customers, a simulation approach may be used, similar to Thomas (2005) or Katok, Thomas and Davis (2008) for service agreements, or Mapes (1993) for fill rate. Ingalls (1998) and Lee et al. (2002) advised on modelling and simulation for designing SCs. Tzafestas and Kapsiotis (1994) used simulation to examine their SC model. In a multistage SC, Zhang and Zhang (2007) simulated information sharing on SC performance. The simulation approach is the method of designing a model of a real-life system to realise the behaviour of the system or/and evaluate strategies for the processes of the system.

Modelling is used to test stochastic SC designs over time, illustrate SC variables, and isolate the problem under discussion (Kulkarni, 2016). Brandenburg et al. (2014) surveyed the recent literature on SCs, noting formal modelling studies for closed-loop SCs, reverse logistics, environmental effects and social factors in forward SCs. They found that a focus on multiple variable decision makers was emerging, and techniques included analytical hierarchy processes, analytical networking and life cycle analysis. Using both supply and time targets within a review period, Thomas (2005) explained that if the contract stipulates a fixed number of items then the target time must vary; if the time is fixed, then the number of items varies. Nevertheless, a standard (finite horizon) contract has both a set number of items and a due date. Thomas (2005) then constructed a static periodic review base stock model to illustrate the fill rate distribution for different review periods.

In this dissertation, we use a simulation approach similar to Thomas' (2005) to analyse the fill rate and ready rate SLAs measures in a SC with multiple customers. This is due to the complexity arising from the numerous parameters in the model. Thomas (2005) noted that researchers standardised the inventory fill rate as stationary demand, which is independent over replenishment review periods for finite time (finite horizon). Further, with a base stock policy and zero lead time, the value of the long-run fill rate is less than the expected value for a periodic fill rate. The fill rate is a random variable across a finite time horizon, the incentive for studying this random variable derives from the extensive practical use of fill rates as the performance measure in SLAs. Understanding the processes for replenishment is of value to managers due to their targets for performance incentives and penalties. Using a stationary base stock policy, Thomas examined the effect of varying review periods on demand distribution and penalty costs with the fill rate SLA measure.

Further research has been undertaken on this concept. Katok, Thomas and Davis (2008) simulated an SC agreement, where a supplier is committed to a minimum fill rate for a given period with a single customer. Two factors were studied in the simulation by Katok, Thomas and Davis (2008), during different review periods, and had different incentives for meeting contract conditions. Under these conditions, the supplier's bonus rate may increase the fill rate, whereas the effect on varying review periods is unknown. The simulation showed that longer review periods were more effective due to more time for communication and response between the parties.

The concept of 'a single-item, single-stage, continuous review inventory system with backordering and constant lead times controlled by a base stock policy', as Larsen and Thornsten (2014, p. 13) commented, has been extended in many different directions. This was to be expected, given the range of SCs and contractual obligations that developed over time.

2.6 Research Gap

This chapter contains a review of the literature on modelling fill rates, ready rates, lead times, and different service policies for demand fulfilment for each research component. Under fill rate SLA measures, some studies developed fill rate with a single customer (Thomas, 2005); (Katok, Thomas, and Davis, 2008), while others considered fill rate with a single review period (Yin and Ma, 2015). However, despite many suppliers dealing with more than one customer, all the research on fill rate and SLAs focuses on cases with only one customer; from what can be ascertained, there has yet to be research investigating the fill rate in multiple customer situations. Furthermore, in the multiple-customer case, two different policies for fulfilling customers' demands are considered: a First-Come-First-Served (FCFS) policy and a second policy (called PLFR throughout the thesis) which prioritises customers based on current measured fill rate performance. Chapter 3 addresses this gap in the literature by studying the effect of having multiple customers on the fill rate with various performance review period lengths.

Moreover, the majority of previous studies on the finite-horizon fill rate assumed zero lead time, even though it is unrealistic in real-world businesses. Hence, in Chapter 4, this thesis attempts to fill a gap in the literature: analysis of the finite-horizon fill rate in an SC with a positive lead time.

From what can be gathered, all literature dealing with fill rate is mainly concerned with its mean value and with simulating its sampling distribution under various supply and demand regimes. In Chapter 5, the variance of the fill rate is assessed and used to consider its consistency as an estimator, as well as its asymptotic normality. The results are also extended to multiple customers, however this thesis concentrates mainly on the case of two customers, as extension beyond this case is straightforward but algebraically cumbersome. Measures of SC reliability for both single and multiple customers are introduced, and finally, several numerical examples are provided to illustrate the foregoing results.

Finally, under ready rate SLA measures, our research fills a gap in the existing literature; this thesis is the first to evaluate ready rate performance in a multiple customer setting. Since the ready rate is a random variable in a finite time horizon, a numerical simulation approach is employed to systemically investigate cost outcomes under different SLA contracts. The research uses a periodic-review, base-stock model with zero lead time. In the analysis, an evaluation is carried out on the effect of different penalty types (fixed and linear) over varying performance review phase durations. Also investigated are the effects of varying base-stock and demand levels. Throughout, the thesis compares and contrasts the multiple customer case to the well-studied single customer case. The results have direct managerial implications for customers and inventory suppliers for both current and future SLAs.

Chapter 3: Finite Time Horizon Fill Rate Analysis for Multiple-Customer Cases

3.1 Introduction

In inventory theory, the management of inventory, its mechanisms and related service levels (such as fill rate) have been an active area of study since the 1950s (Johnson et al. 1995). A popular performance measure of service level in inventory management is the item fill rate. Lee and Billington (1992) and Johnson and Scudder (1999) reported that the item fill rate is a popular and preferred service level measurement applied in many firms. The fill rate is defined as 'the long run average fraction of demand satisfied immediately using on-hand stock' (Zhang & Zhang 2007). In this chapter, we study the item fill rate (also referred to as the 'fill rate') as our service level measure. This choice from the extensive practical use of fill rates as the performance measure in SLAs.

In this chapter, we try to clarify the 'replenishment period'. In practice, a supplier may replenish the inventory to a base stock level at a specified time interval, called the replenishment period; for example, weekly, monthly or daily. Supplier performance is then evaluated at a regular interval, referred to as the 'performance review period'; for example: monthly, quarterly or six-monthly. In a given performance review period, a supplier may incur a financial penalty for not meeting a target fill rate that is specified in the terms of the SLA. In some SLAs, the supplier may receive a bonus for exceeding the target fill rate level. In a finite period, inventory level and supplier performance can be monitored. This information can be used to assist a supplier in meeting or exceeding the terms of the SLA.

In regular SLAs, when the supplier signs the contract, they are responsible for reaching a target fill rate at each performance review point. Unfortunately, it is not easy to calculate the probability of achieving a specific fill rate target for a given base stock level and performance review cycle. In practice, some suppliers with SLAs use an infinite horizon model, then 'overshoot' to increase their probability of success. For instance, if a supplier must exceed a 90% fill rate, they might establish their stock level based on a 93% long-run fill rate. This 'overshooting' is done to improve their probability of success and avoid any related penalties.

3.2 Procedure for Fill Rate Analysis

The fill rate can be measured over various performance review periods (finite time horizons). The problem is to establish the effects of review period length and the correlations between customers' demands on average fill rate, and the probability of overreaching the target fill rate. In fill rate analysis a probability of exceeding the target is of interest, especially if the problem is compounded with multiple individually contracted customers. The multiple-customer demands are filled from the supplier's pooled inventory, with constant demand. Inventory policies for multiple customers' cases considered in this thesis are FCFS and PLFR. These constraints are examined under both demand Erlang distribution and normal distribution. Taking the problem from the customers' perspective, the realised service level is calculated for customers having different levels of demand, all with the same supplier.

3.3 Related Research Questions for this Chapter

The research questions for this chapter are:

- Q3.1 What is the change in the average fill rate when the review period is increased?
- Q3.2 What is the change in the average fill rate and the probability of exceeding the target fill rate of each of multiple customers, compared to a single customer?
- Q3.3 How does the correlation structure of demand affect the average fill rate and the probability of overreaching the target fill rate of each customer?
- Q3.4 Who receives the better service (higher average fill rate and higher probability of exceeding the target fill rate): the customer with higher demand or the customer with lower demand?

3.4 Model and Notations

The notations for this model are:

T: duration of review period measured by the number of replenishment periods. For example, when T = 5 it means the performance is reviewed after 5 replenishment periods.

S: base stock level;

I: set of customers, cardinality of set **I** is *n;*

C_i: customer index; the customer index is deleted in cases of a single customer or with multiple identical customers;

$\alpha_t^{C_i}(S,T)$: the fill rate for customer *i* for the period *t*,….*t*+*T* where *T* is the review period, any *S* is the base stock level and $\alpha_t^{C_i}(S,T)$ is a random variable;

Y_{it}: customer *i* demand at time *t*, where *t* is expressed as discrete time (e.g., a day);

X_{it}: demand satisfied from inventory for customer i at time *t*, $X_{it} = \min(Y_{it}, \text{available inventory})$:

$$\sum_{i=1}^{n} X_{it} = \min(\sum_{i=1}^{n} Y_{it}, S)$$

$\bar{\alpha}^{C_i}(S,T)$: the average fill rate for customer *i*, when the base stock level is *S* and review period is *T;*

$P_{\pi_i}^{C_i}(S,T)$: the probability of exceeding the target fill rate of π_i for customer *i*, when the base stock level is *S* and the performance review period is *T*.

We assume a discrete time situation, where at each time *t* there are *S* units in stock and stochastic demands arrive from customers. All unmet demand is lost. Inventory is replenished to *S* for period *t*. An algorithm is used for multiple customers under the FCFS constraint; for two simultaneous customers, there is a 50% chance that customer 1 is served first. For the prioritised constraint with multiple customers, the customer with the highest negative deviation from their target fill rate in the current period is served first. For example, if $T = 5$ days and on day 4 the supplier is unable fill all demand, each customer's fill rate for the previous three days is calculated and customers are prioritised according to the difference between their computed three-day fill rate (for the previous three days) and their target fill rate. Note that the PLFR policy is not necessarily optimal, as the optimal policy must be designed based on the penalty structure outlined in the varying agreements. Accordingly, the fill rate for the period t…., $t+T$ for customer i is:

$$\alpha_t^{C_i}(S,T) \equiv \frac{X_{i(t+1)}+\cdots+X_{i(t+T)}}{Y_{i(t+1)}+\cdots+Y_{i(t+T)}} \tag{3.1}$$

The average fill rate for customer i as a function of T and S is defined as:

$$\bar{\alpha}^{C_i}(S,T) = E\left(\frac{X_{i1}+\cdots+X_{iT}}{Y_{i1}+\cdots+Y_{iT}}\right) \tag{3.2}$$

then the long-run fill rate is defined (for aggregated customers' demand) by:

$$\bar{\alpha}(S,\infty) = \lim_{T\to\infty} \mathrm{E}\left(\frac{\sum_{i=1}^{n}(X_{i1}+\cdots+X_{iT})}{\sum_{i=1}^{n} Y_{i1}+\cdots+Y_{iT}}\right) \tag{3.3}$$

and the probability of exceeding the target fill rate (π_i) for customer i is:

$$P_{\pi_i}^{C_i}(S,T) = \Pr(\alpha_t^{C_i}(S,T) \geq \pi_i) \tag{3.4}$$

In the next section, a simulation model is used to address research related questions 3.1 to 3.4.

3.5 Fill Rate Analysis Approach and Review Periods

Simulation models are used to study real-life processes that change stochastically over time. Simulation can be useful to determine possible alternatives when designing an SC, where supplier and customers face a complex and uncertain decision environment. To evaluate the fill rate measure in an SC with multiple customers, due to difficulties from the several parameters in the model, a simulation approach similar to Thomas' (2005) is applied in this chapter. Scenarios are then presented on customer numbers, service policy (FCFS or PLFR) and duration of the review period. In the single and multiple-customer case comparisons, in all our experiments we assume that the overall demand distributions are equal, with the exception of one experiment in Section 3.3, where the demand distribution in both one-customer and multiple-customer cases is Gamma, and the customers' demands are correlated in the multiple-customer case. This is due to the fact that the distribution of the sum of two correlated variables with Gamma distribution is not Gamma. In that experiment, we fixed the overall mean and variance.

When addressing Q3.1, regarding the change in the average fill rate when the review period is increased, all other parameters were fixed and review period lengths are set as

T= {5, 10, 20, 100}. Q3.2 concerned on changes in the average fill rate and the probability of exceeding the target fill rate of multiple customers when compared to a single customer. This is run via controlling T and S, and changing the number of customers in the SC. Q3.3 (regarding the correlation structure of demand impacting the average fill rate and the probability of overreaching the target fill rate of each customer) was run, controlling all parameters and changing the correlation coefficient of demand. The final question referred to better service, and this was run with all parameters controlled and the average and variance of demand changed for the customer with higher demand.

From our simulation model's output, we indicate the average fill rate (FR) in the plots, and some probabilities over the unknown distribution of the fill rate, such as the probability of exceeding the target FR, referred to here as the probability of success (PS).

3.5.1 Calculating simulation runs

The required number of simulation runs (sample size *n*) was obtained from the Dvoretzky-Kiefer-Wolfowitz (DKF) inequality. Kosorok (2008) stated that the DKF inequality for a two-sided estimate of the empirical cumulative distribution is:

$$\Pr(\sup_{x \in \mathrm{R}}(F_n(x) - F(x)) > \varepsilon) \leq 2e^{-2n\varepsilon^2} \tag{3.5}$$

where ε is the error in estimation ($\varepsilon > 0$), $F_n(.)$ is the estimated empirical cumulative distribution and $F(x)$ is the true cumulative distribution. Therefore, to have an error less than or equal to ε in estimating $F_n(.)$ with $(100 \times \theta)\%$ confidence (e.g., 0.95), the required n is obtained as:

$$n = ln\left(\frac{1-\theta}{2}\right)/(-2\varepsilon^2) \tag{3.6}$$

In this dissertation, we set the targets for FRs by $\theta = 0.95$ and $\varepsilon = 0.01$, thus 18,445 was the optimum number of simulation runs, representing 18,445 review periods (e.g. 18445× 10 days when the performance review period T = 10). Then, output for the simulation model for each scenario was 18,445 FR values, to approximate the distribution of the FR.

3.5.2 Distributions of customer demand

To validate the model in our simulation experiments, for accurate results, two distributions of customer demand were examined. The first was the Erlang distribution, with a shape parameter of k and scale parameter of 1 denoted by $Erlang\ (k)$. The Erlang distribution has been used to study the finite horizon fill rate in previous studies (Hira 2009; Thomas 2005). The second distribution for customer demand employed in this study is the normal distribution $N\ (\mu, \sigma^2)$. When demand distribution is normal, the parameters are selected such that the probability of having a negative demand is very small; however, when the demand is negative it is assumed to be zero. When there are multiple identical customers, the demand distribution of each customer is Erlang($\frac{k}{n}$) and normal($\frac{\mu}{n}, \sqrt{\frac{\sigma^2}{n}}$),. Situations in which customers are not identical are analysed later in this chapter.

We can design and employ our experiments for any base stock-level S. However, to be consistent with previous studies, for some of the experiments we select S, such that the long-run FR for the aggregated demand is 0.95. To calculate the value of S that yields a given long-run FR, the relationship between the unfulfilled demand and the long-run FR was used.

Let μ denote the expected value of demand and α denote the long-run FR. Then $1-\alpha$ is the average percentage of lost demand. Hence, the expected value of unfulfilled demand is $(1-\alpha)\mu$. Therefore:

$$\mu(1-\alpha) = \int_S^{\infty} (y-S) f(y) dy$$

$$\mu(1-\alpha) = \int_S^{\infty} y f(y) dy - S(1 - F(S))$$

where $f(.)$ and $F(.)$ are the probability density function (pdf) and cumulative distribution function (CDF), respectively. By simplifying the above formula, we obtain:

$$\alpha = 1 + \left(S\big(1-F(S)\big) - \textstyle\int_S^{\infty} y f(y) dy\right)/\mu \tag{3.7}$$

For Erlang(k), (3.7) is simplified to:

$$\alpha = 1 + \frac{S}{k}\big(1-F(S)\big) - \left(\frac{1}{k}\int_S^{\infty} \frac{y^k e^{-y}}{(k-1)!} dy\right) = 1 + \frac{S}{k}\big(1-F(S)\big) - (1 - F''(S))$$

where $F''(.)$ is the CDF of Erlang($K+1$).

We finally obtain:

$$\alpha = F"(S) + \frac{S}{k}(1 - F(S)) \tag{3.8}$$

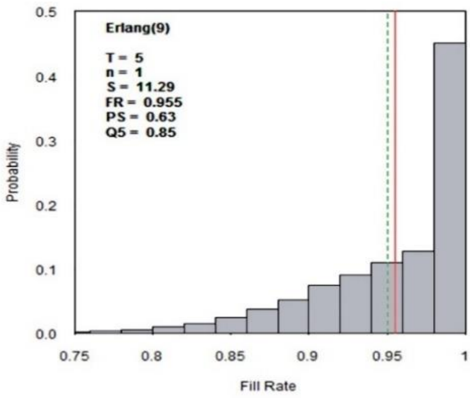
By setting the long-run FR to 0.95 for Erlang (6) and Erlang (9), base stock levels are reached at 8.11 (S = 8.11) and 11.29 (S = 11.29), respectively.

3.5.3 The FR distribution

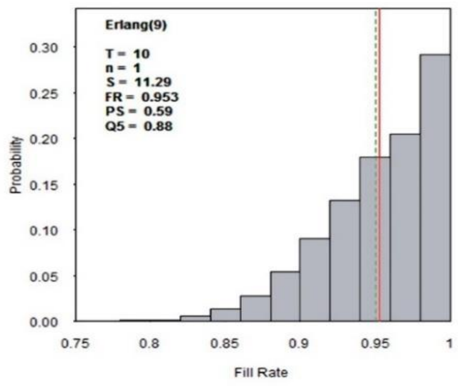
For a supplier with multiple customers, we examine the effect of performance review period length on the average FR and the probability of exceeding the target FR. We run the simulation for the one and three-customer cases, applying two different customer demand distributions (Erlang and normal), with four different performance review periods (i.e., T = 5, 10, 20, 100), and two different service policies (FCFS and PLFR). In total, we simulate 24 scenarios. Figures 3.1 and 3.2 present the FR distributions when $T = 5$ and *10, 20, 100* and the demand distribution is Erlang. In each plot there are two vertical lines, the green dashed line indicates the target FR threshold of 0.95, while the solid red line shows the average FR for each scenario. The average FR, the probability of exceeding the target fill rate of 0.95 (PS), and the fifth quintile (Q5) are reported. Figures 3.3 and 3.4. present the plots for Erlang and normally distributed demand with T = 5, 10, 20 and 100.While the FR is a continuous measure between 0 and 1, the output of each simulation scenario is 18,445 FR values. These values are used to construct the density histograms in Figures 3.1 and 3.4.

All experiments were performed on an Intel Core i7 CPU with a 2.2 GHz processor and 16 GB RAM. The codes were written in Python programming language (Python 3.4 version). To solve these models, we applied other packages to solve methods, such as Numpy, Sicipy, Matplotlib and Anaconda. When solving simulation algorithm codes in Figures 3.5 to 3.10, it took a long time to reach the results (typically two to five days, especially when we increased the length of the performance review period) due to many loops and the desired accuracy of the results.

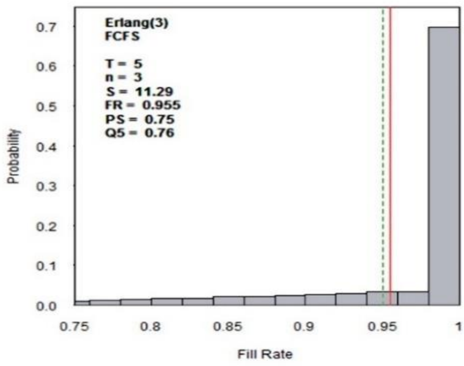
The comparison of the results in Figures 3.1.a, 3.1.b, 3.2.a and 3.2.b shows that in the single-customer case, when T is increased, the average FR decreases. The same impact is detected in the multiple-customer case in Figures 3.1.c-f and 3.2.c-f for both service policies (FCFS and PLFR). The probability of exceeding the target FR is higher when the FCFS policy is used, whereas Q5 is higher for the PLFR policy. This is due to the fact that in the PLFR policy the customer with the lowest (current) FR is always prioritised, and this serves to cut the left tail of the FR distribution. However, at the same time, it may squeeze the right tail of the FR distribution and thus cause a reduction in PS compared to the FCFS.



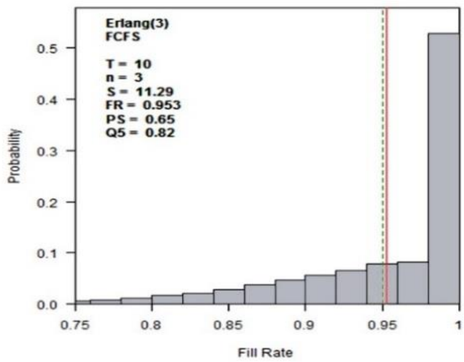
a. One customer with demand distribution of Erlang(9) and T = 5



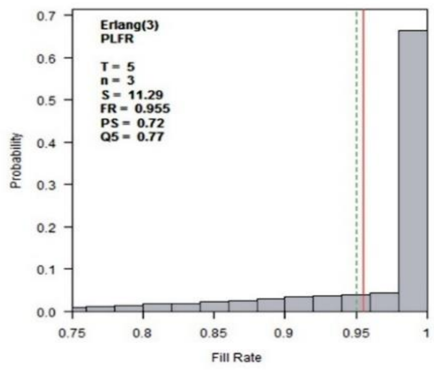
b. One customer with demand distribution of Erlang(9) and T = 10



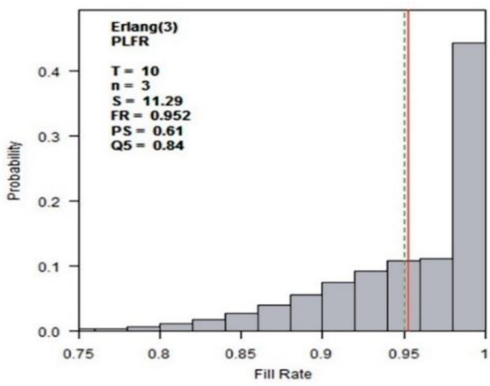
c. Three customers with demand distribution of Erlang(3), T = 5 and Service Policy is FCFS



d. Three customers with demand distribution of Erlang(3), T = 10 and Service Policy is FCFS



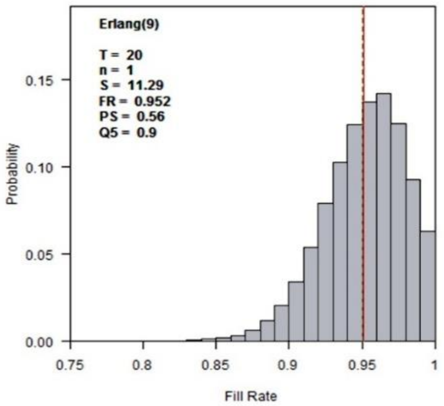
f. Three customers with demand distribution of Erlang(3), T = 5 and Service Policy is PLFR



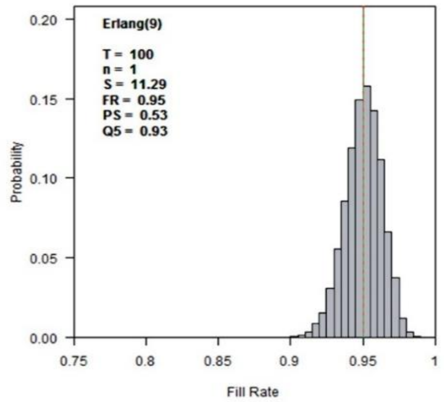
e. Three customers with demand distribution of Erlang(3), T = 10 and Service Policy is PLFR

Note: FCFS 2nd row, PLFR 3rd row

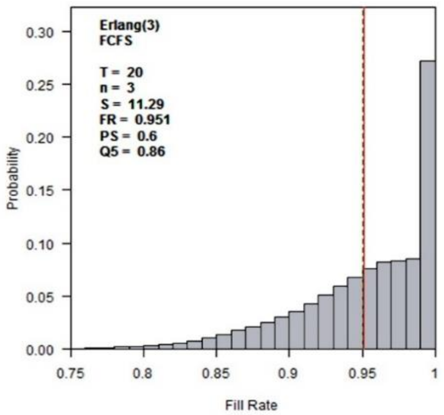
Figure 3.1: The distribution of FR when the distribution of the overall demand is Erlang(9) and T is 5 or 10. In each plot, n denotes the number of customers, FR denotes the average fill rate and PS denotes the probability of exceeding the target fill rate of 0.95 The vertical red solid line and green dashed line show the average FR and the target FR, respectively.



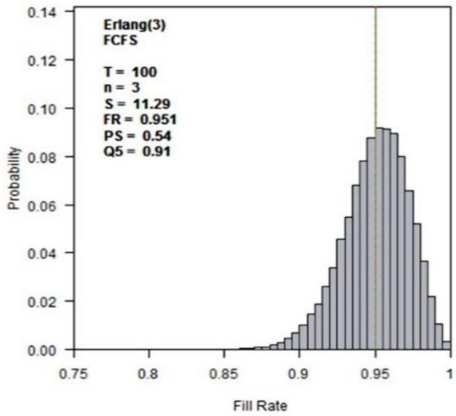
a. One customer with demand distribution of Erlang(9) and T = 20



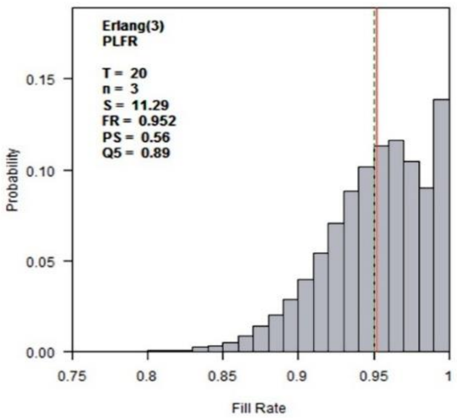
b. One customer with demand distribution of Erlang(9) and T = 100



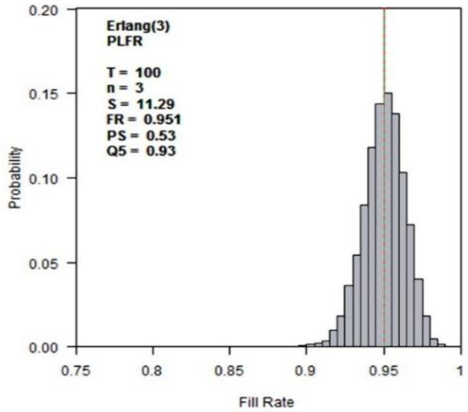
c. Three customers with demand distribution of Erlang(3), *T* = 20 and the service policy is FCFS



d. Three customers with demand distribution of Erlang(3), *T* = 100 and the service policy is FCFS



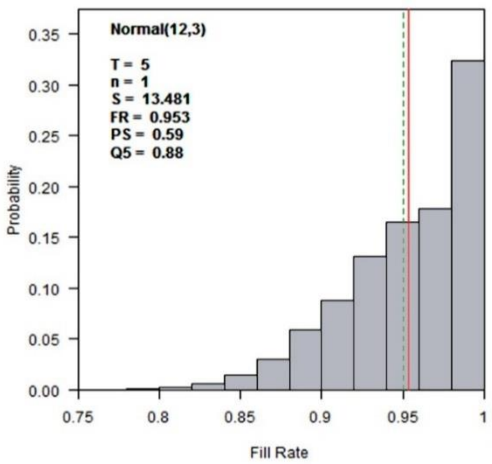
e. Three customers with demand distribution of Erlang(3), *T* = 20 and the service policy is PLFR



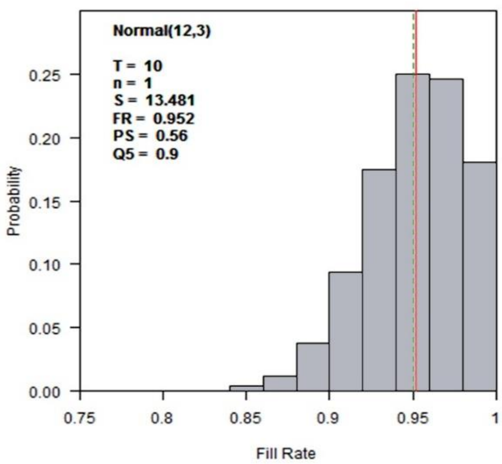
f. Three customers with demand distribution of Erlang(3) , *T* = 100 and the service policy is PLFR

Note: FCFS 2nd row, PLFR 3rd row

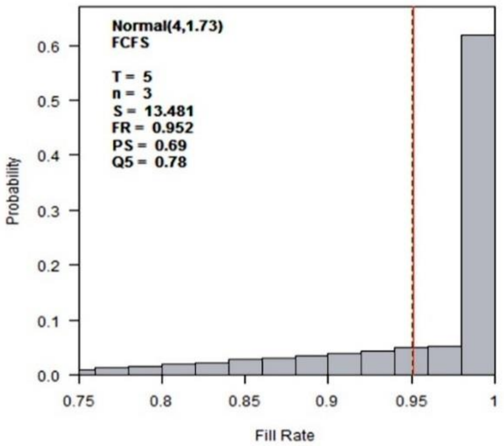
Figure 3.2: The FR distribution when the distribution of the overall demand is $\mathbf{Erlang(9)}$ and *T* is 20 and 100. The plots in the second row show the result when the FCFS policy is used, and the plot in the third row shows the results when the PLFR service policy is used.



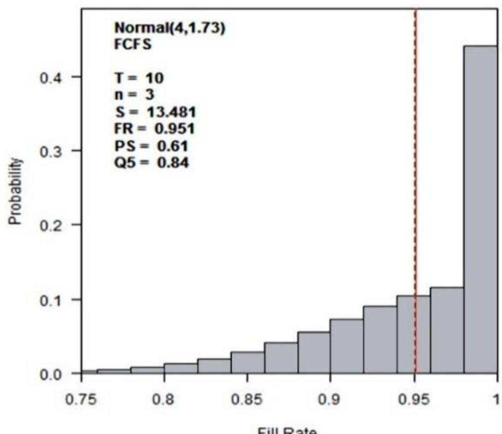
a. One customer with demand distribution of N(12,3) and T=5T= 20



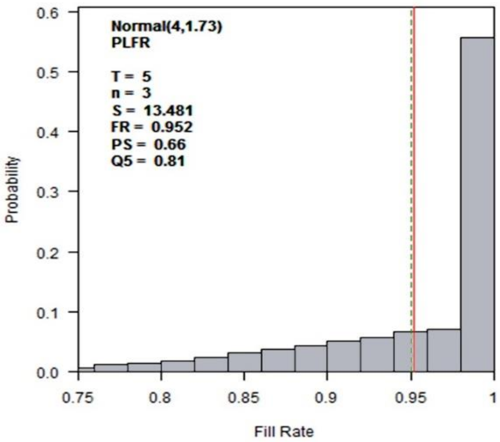
b. One customer with demand distribution of N(12,3) and T=10



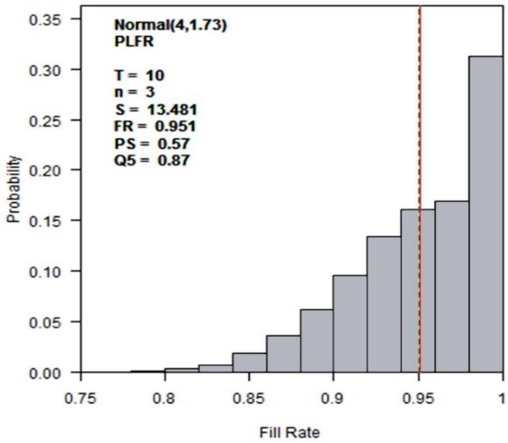
c. Three customers with demand distribution of N(4,1.73), T=5 and the service policy is FCFS



d. Three customers with demand distribution of N(4,1.73), T=10 and the service policy is FCFS

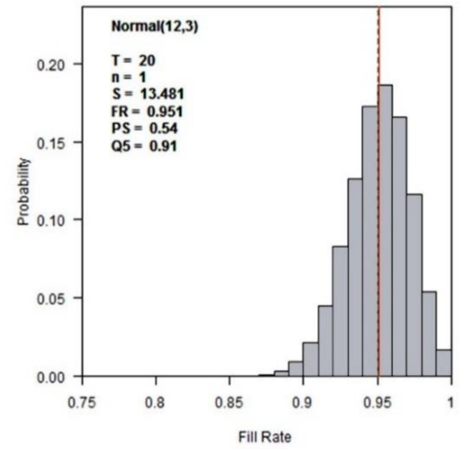


e. Three customers with demand distribution of N(4,1.73), T=5 and the service policy is PLFR

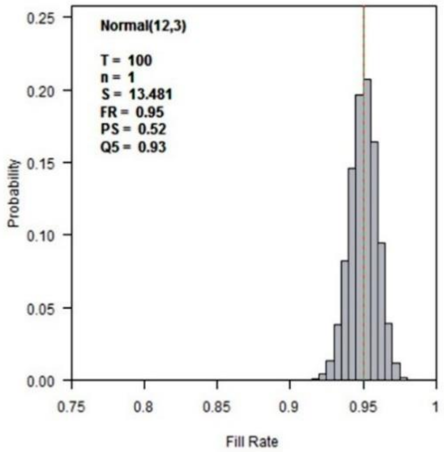


f. Three customers with demand distribution of N(4,1.73), T=10 and the service policy is PLFR

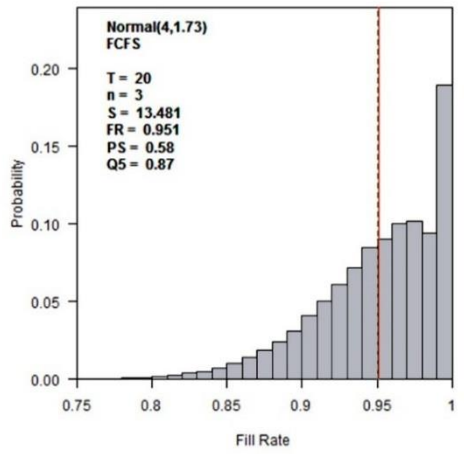
Figure 3.3: The FR distribution when the distribution of the overall demand is $\mathrm{Normal}(12, 3)$ and T is 5 and 10. The plots in the second row show the result when the CFS policy is used, and the plot in the third row shows the results when the PLFR service policy is used.



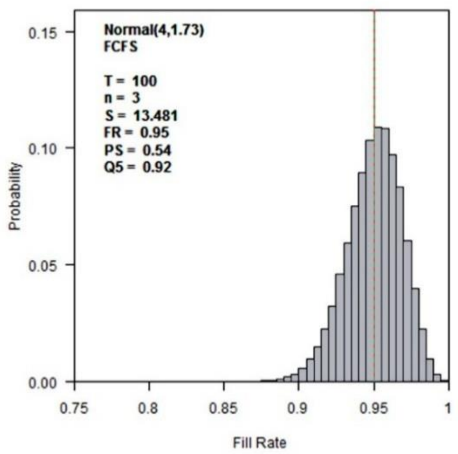
a. One customer with demand distribution of N(12,3) and T=20



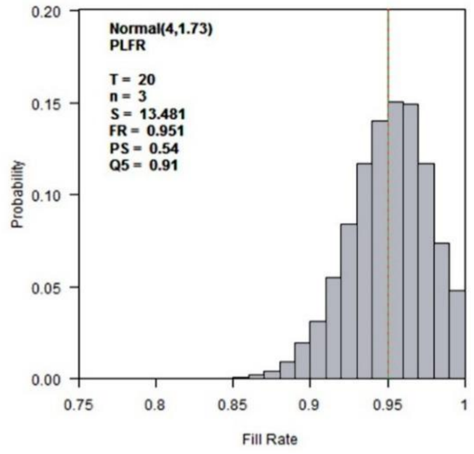
b. One customer with demand distribution of N(12,3) and T=100



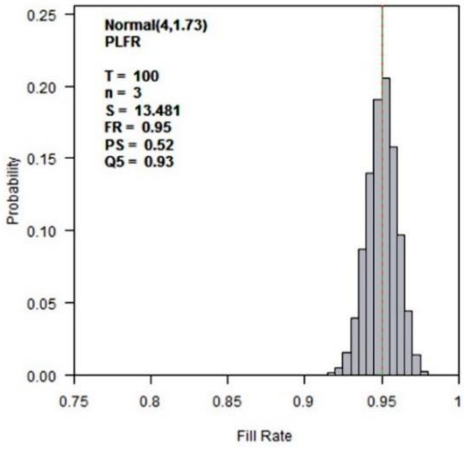
c. Three customers with demand distribution of N(4,1.73) , T=20 and the service policy is FCFS



d. Three customers with demand distribution of N(4,1.73), T=100 and the service policy is FCFS



e. Three customers with demand distribution of N(4,1.73), T=20 and and the service policy is PLFR



f. Three customers with demand distribution of N(4,1.73), T=100 and the service policy is PLFR

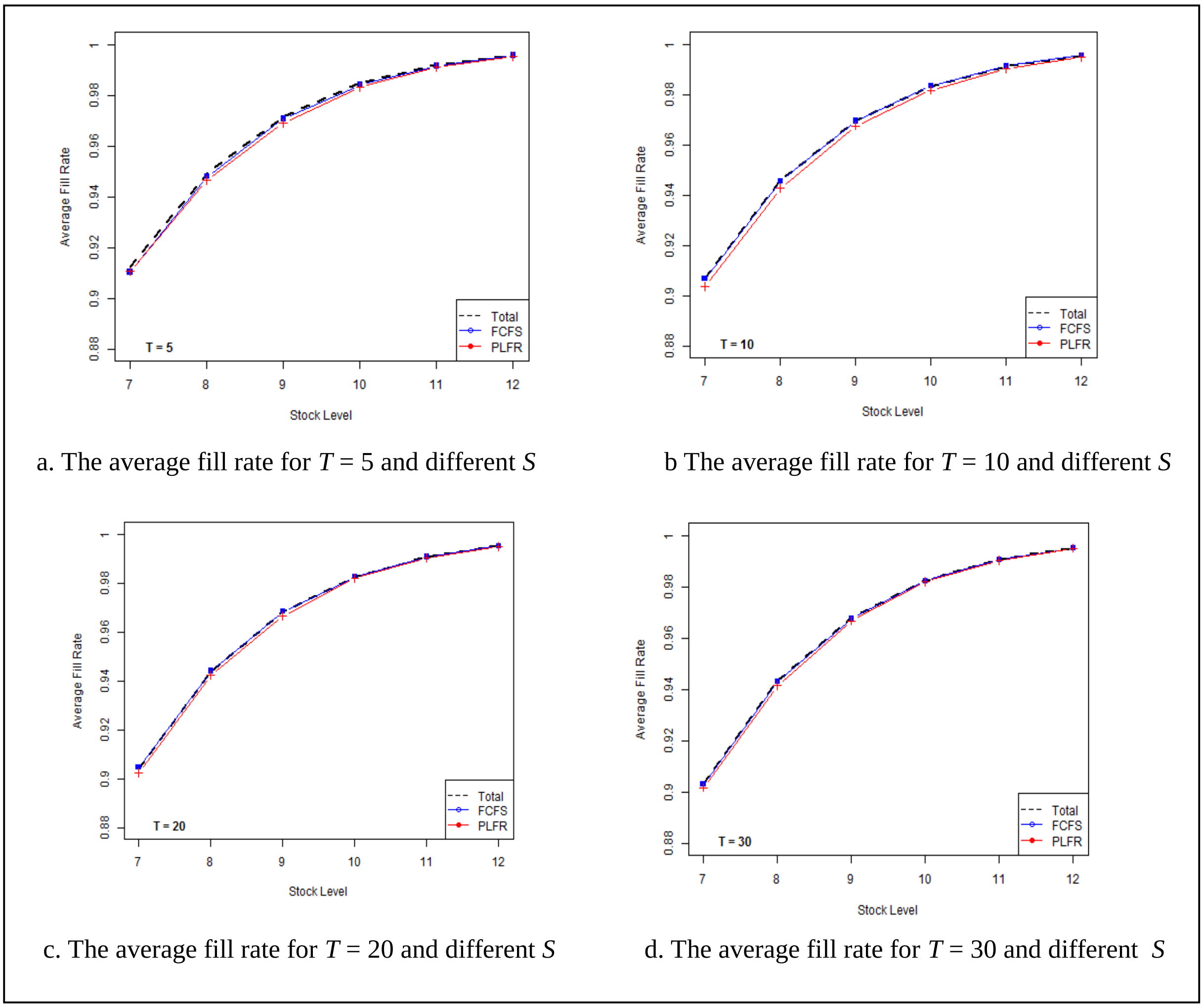
Figure 3.4: The FR distribution when the distribution of the overall demand is $\mathrm{Normal}(12, 3)$ and *T* is 20 and 100. The plots in the second row show the result when FCFS policy is used, and the plot in the third row shows the results when PLFR service policy is used.

Figures 3.3 and 3.4 display the FR distributions for one customer and three independent customers when the performance review period is T = 5, 10, 20 and 100, and the distribution of the overall demand is normal $(12{,}3)$. In the case with three customers the distribution of demand for each customer is normal $(4, \sqrt{3})$. The base stock level in this scenario is S = 13.481, to have a long-run FR of 0.95. The results in Figures 3.1, 3.2, 3.3 and 3.4 show that when increasing the performance review period the average FR decreases. They also show that the average FR when T is fixed is the same for the one-customer and multiple-customer case (for both the FCFS and PLFR policies). The latter point will be verified by the results presented in Figures 3.5, 3.6, 3.7 and 3.8.

In the next step we compare the FR and PS for the single versus multiple-customer cases, considering several base stock-level (S) and review period length (T) values. In the single-customer case the demand distribution is $\text{Erlang}(6)$, while in the multiple-customer case there are two customers each with $\text{Erlang}(3)$ demand. Thus, in all cases the overall demand distribution is $\text{Erlang}(6)$. We also compare the FR and PS for single versus multiple-customer cases, where the demand distribution is normal (12, 2). The distribution of demand for each customer under multiple-customer case is normal $(6, \sqrt{2})$.

To answer question Q3.1 (*What is the change in the average FR when the review period is increased?*) the following results were obtained. Figure 3.5 shows that the average FR is fixed for a given S in all cases (one customer; two customers under either FCFS or PLFR policies). Figure 3.6 is similar to Figure 3.5, but shows cases with normal demand distribution. The Figure 3.6 results using either FCFS or PLFR policies are consistent with Figure 3.5 results; that is, average FRs are the same for single or multiple customers.

Figure 3.7 displayed a convex shaped plot in both single and multiple customer cases, where s T rose probability of success (PS) first decreased, and then increased. For instance, with a short review period (lower than a threshold (θ), e.g., $T = 5$), a supplier would prefer multiple customers due to the increased PS. However, longer review periods result in a preference for a single customer. Figure 3.8 is similar to Figure 3.7, but shows cases with normal demand distribution. The results of Figure 3.8 verify the findings from Figure 3.7 that indicate for shorter T, the probability of exceeding the target FR is higher for multiple customers, and for longer T, the probability of exceeding the target FR is higher for a single customer.



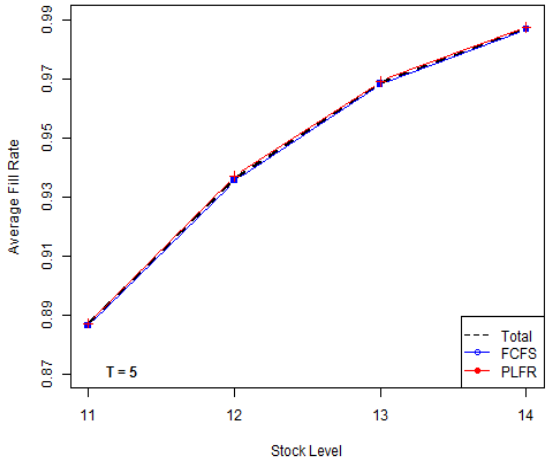
a. The average fill rate for $T = 5$ and different S

b The average fill rate for $T = 10$ and different S

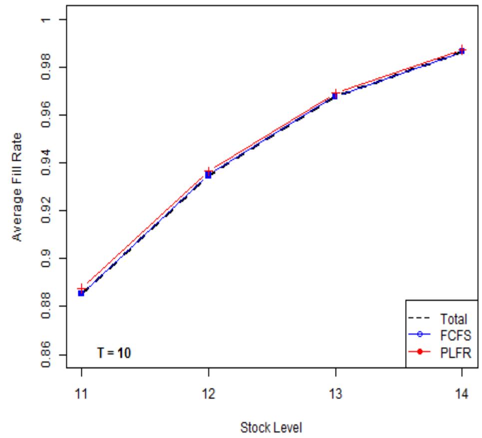
c. The average fill rate for $T = 20$ and different S

d. The average fill rate for $T = 30$ and different S

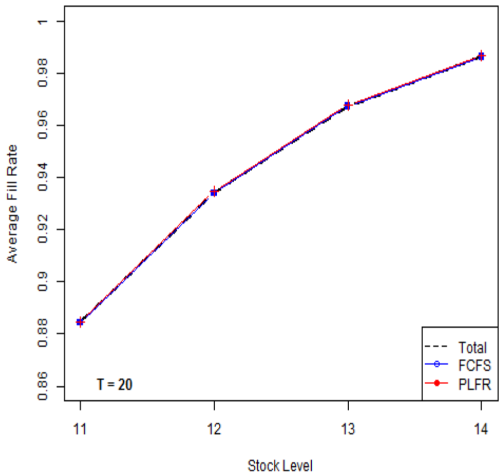
Figure 3.5: Comparison of the average FR for the single and multiple-customer cases. In the single-customer case the demand distribution is Erlang(6). In the multiple-customer case there are two customers each with an Erlang(3) demand distribution.



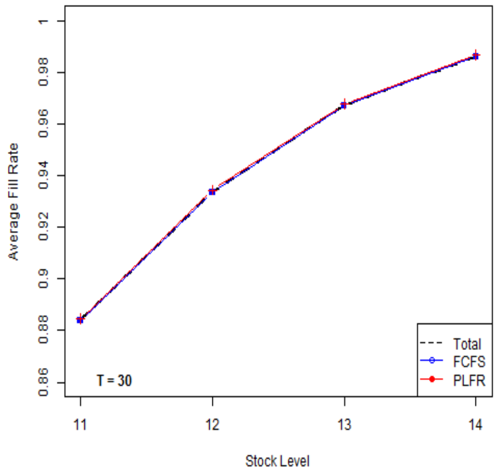
a The average fill rate for T=5 and different S



b. The average fill rate for T=10 and different S

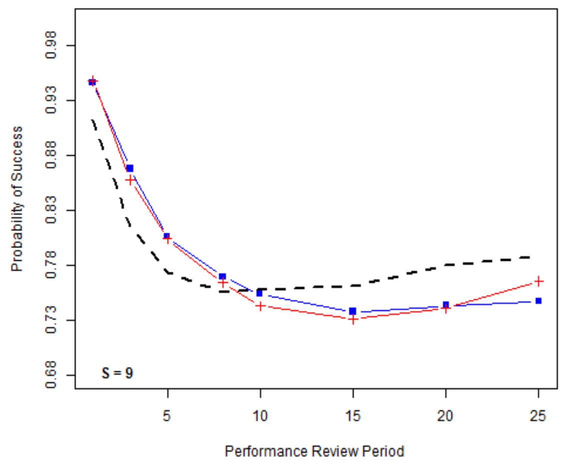


c. The average fill rate for T=20 and different S

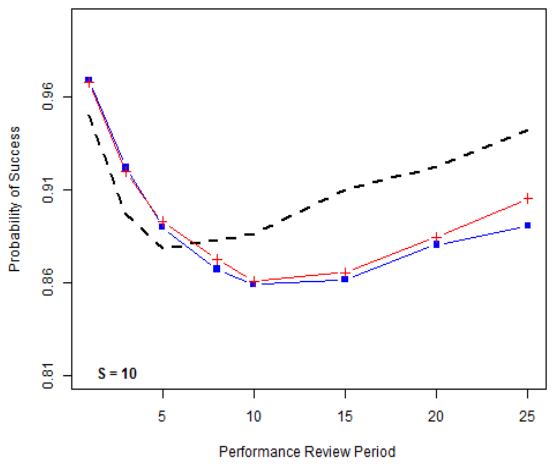


d. The average fill rate for T=30 and different S

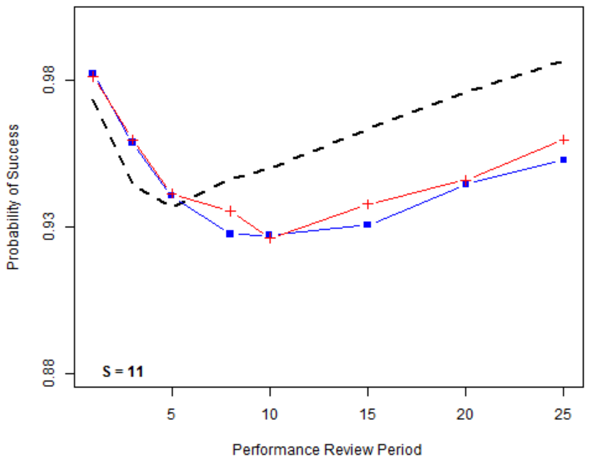
Figure 3.6: The change in the average FR. The overall demand distribution is $\mathrm{Normal}(12, 2)$. The distribution of demand for each customer in multiple-customer case is $\mathrm{Normal}(6, \sqrt{2})$.



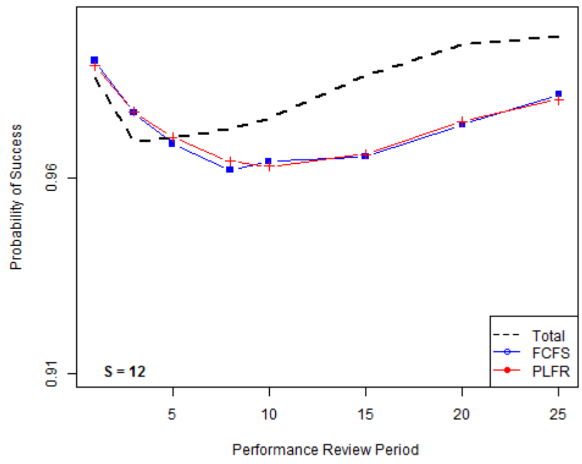
a. PS for *S*=9 when *T* changes from 1 to 25



b. PS for *S*=10 when *T* changes from 1 to 25

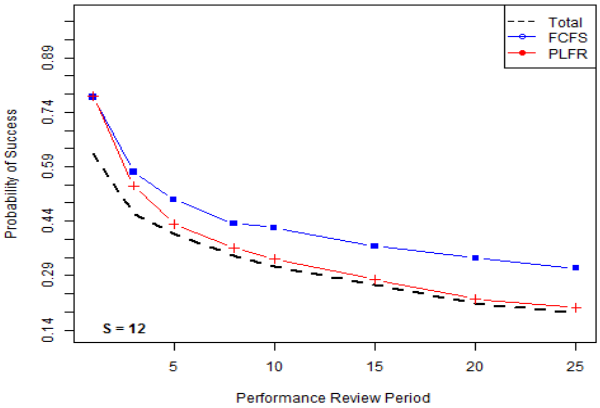


c. PS for *S*=11 when *T* changes from 1 to 25

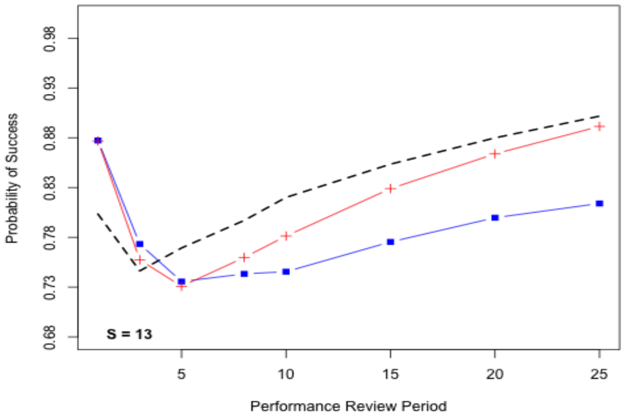


d. PS for *S*=12 when *T* changes from 1 to 25

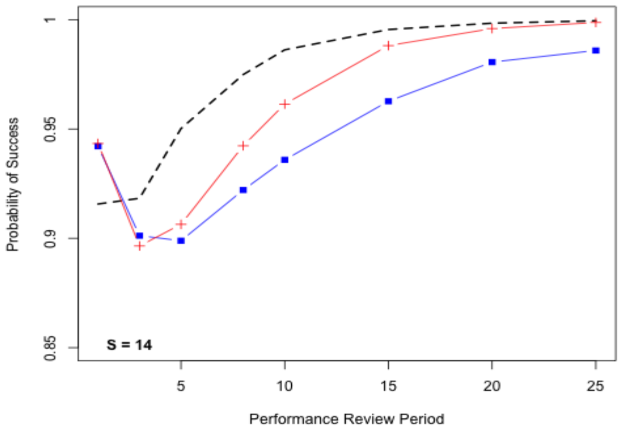
Figure 3.7: Comparison of the probability of exceeding the target FR (PS) of 0.95 for the single and multiple-customer cases. In the single-customer case the demand distribution is Erlang(6). In the multiple-customer case there are two customers each with an Erlang(3) demand distribution.



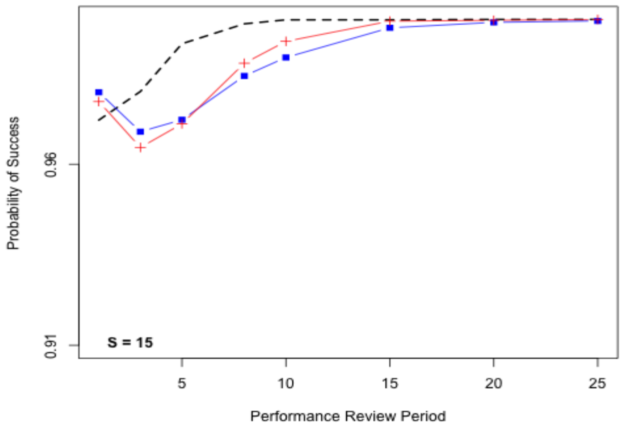
a. PS for S = 12 when T changes from 1 to 25



b. PS for S = 13 when T changes from 1 to 25



c. PS for S = 14 when T changes from 1 to 25



d . PS for S = 15 when T changes from 1 to 25

Figure 3.8: The change in the probability of exceeding the target FR (PS) of 0.95. The overall demand distribution is $\text{Normal}(12, 2)$. The distribution of demand for each customer in multiple-customer case is $\text{Normal}(6, \sqrt{2})$.

3.6 Probability of Exceeding Target FR

Question 3.2 is: *What is the change in the average FR and the probability of exceeding the target FR of each of multiple customers when compared to a single customer?*

In this section, the probability of exceeding the target FR for each customer is considered through various scenarios. We answer research question 3.2 by analysing the required base stock level (*S*) to realise a given probability of exceeding the target FR in the multiple-customer context.

In each case, there is a numerical search for the lowest base stock level (*S*) that can satisfy the given PS. This is undertaken by applying a simulation-optimisation approach to solve:

$$\text{Min S} \tag{3.9}$$

Subject to:

$$P_{\pi_i}^{C_i}(S,T) \geq \gamma_i, \text{ For all } i,\ i= 1,..., n$$

$$S \geq 0$$

where π_i is the target FR for customer i and γ_i is the required probability of exceeding the target FR for customer i.

3.6.1 Procedure for probability analysis

In the simulation-optimisation approach, a systematic heuristic search method is used, similar in concept to the bisection method for numerically finding the root of $f(x) = 0$. Initial base stock values are, S_1 and S_2, where S_1 is a low number so that $S = S_1$ results in the violation of at least one constraint in Equation (3.9); while S_2 is higher so that $S = S_2$ satisfies all constraints in (3.9).

A new base stock level is defined as $S_{new} = \frac{S_1+S_2}{2}$ and if by setting $S = S_{new}$ all the constraints in Equation (3.9) are satisfied, then $S_2 = S_{new}$,otherwise $S_1 = S_{new}$. The procedure continues until $S_2 - S_1 \leq \xi$ or the number of iterations exceeds ψ (ψ can be a large number or a constraint that can be removed). On running the calculation, S_2 is the solution obtained for Equation (3.9). ξ is the highest acceptable error for *S* that guarantees that the optimal S is not less than $S_2 - \xi$. In this study $\xi = 0.01$ and $= 50$. The pseudo code for this simulation-optimisation approach is shown below:

Results for Q3.2

Table 3.1: Pseudo code of the simulation-optimisation method used for solving (9)

Initialization

Set Initial S_1 and S_2 such that $S = S_1$ violates at least one of the constraints in (9) and $S = S_2$ satisfies all of the constraints in (9). Set $\omega = 0$. The simulation model should be run for S to check if any of the constraints are violated.

While $S_2 - S_1 \geq \xi$ and $\omega \leq \psi$

$S_{new} = \frac{S_1+S_2}{2}$

Run the simulation model for $S = S_{new}$

If $S = S_{new}$ satisfies all constraints in (9) then

$S_2 = S_{new}$

otherwise $S_1 = S_{new}$

$\omega = \omega + 1$

End While

$S = S_2$

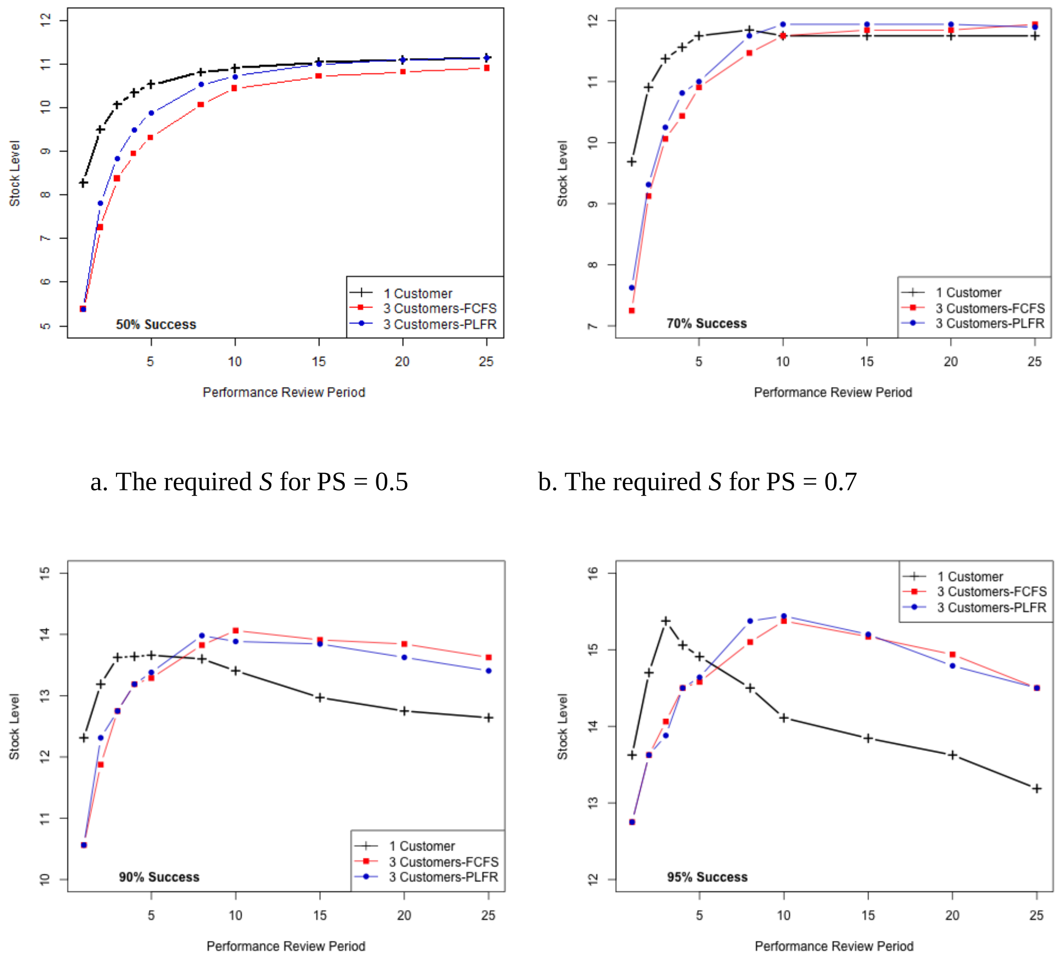
Results for Q3.2

The results of the model (3.9) for different parameter values are presented in Figure 3.9. The overall distribution of demand is Erlang (9), and each multiple customer has a demand distribution of Erlang (3). The figure shows the stock levels for one or three customers to exceed a 95% target FR with probabilities of success (50%, 70%, 90% and 95%), when the distribution of the overall demand is Erlang(9).

In Figure 3.9, when we decrease the performance review period, the required *S* to achieve a given PS is lower when there are multiple-customer, regardless of the service policy used (FCFS or PLFR). In each of the plots in Figure 3.9, we can observe a threshold *T* value (θ), where the single and multiple-customer curves intersect. This is the 'point of indifference', where the required *S* to achieve a given PS is equal in the single and multiple-customer cases. For *T* values greater than θ, a lower *S* is required in the single-customer case.

Likewise, it can be seen that this threshold value (θ) decreases as the required PS increases. For instance, for the PLFR policy, when the required PS is 0.7 then $\theta = 8$, meaning that if $T < 8$, a lower *S* is needed in the multiple-customer case. However, when the required probability of success is increased to 0.9, the threshold θ below which a lower *S* is required in the multiple-customer case is reduced to 6, and when PS =0.99 the threshold θ is below 5. This finding is valid for both the FCFS and PLFR policies.

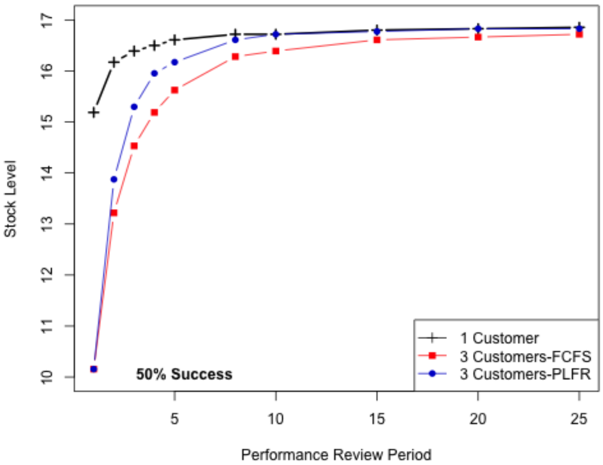
The phenomena described above can be understood by looking at the FR distributions presented in Figures 3.1, 3.2, 3.3 and 3.4. When *T* is short, the spread of the distribution is less in the multiple-customer case, and hence a lower base stock level is required to achieve a given PS when compared to the single-customer case. Conversely, when *T* is long, the spread of the distribution is less in the single-customer case, and so to reach a given PS a lower base stock level is needed when servicing a single customer. This analysis was repeated for normally distributed demand, and the results are presented in Figure 3.10.



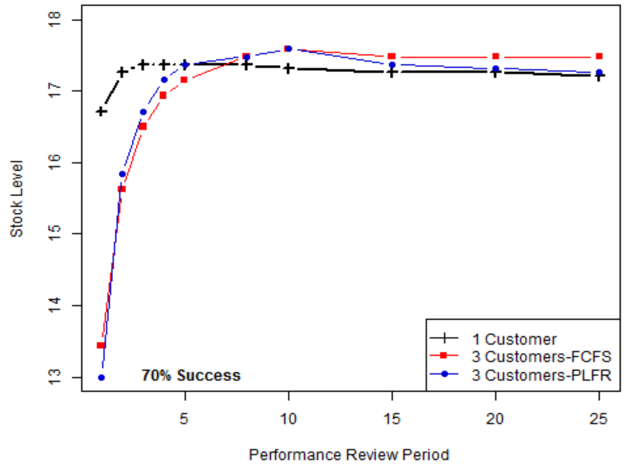
a. The required S for PS = 0.5 b. The required S for PS = 0.7

c. The required S for PS = 0.9 d. The required S for PS = 0.95

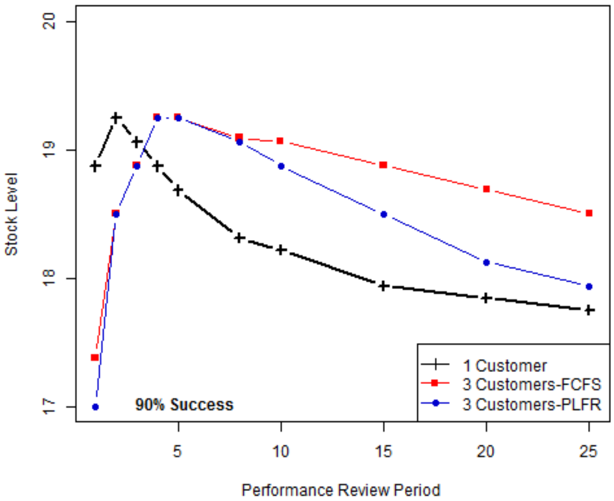
Figure 3.9: The comparison between one and three customers in terms of the base stock levels needed to meet a 95% target FR with various probabilities of success (50%, 70%, 90% and 95%), when the distribution of the overall demand is Erlang(9).



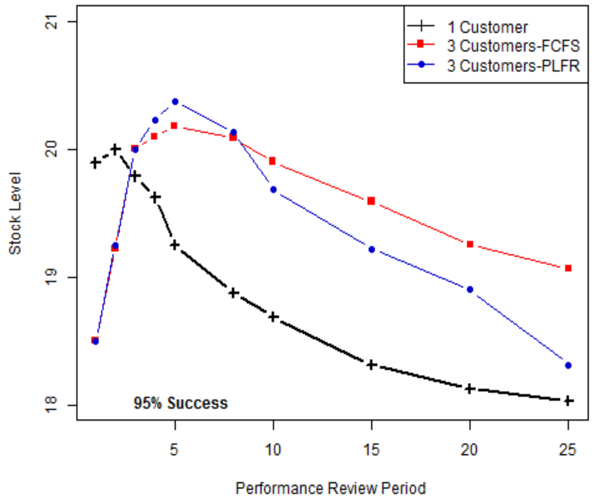
a. S required for PS = 0.5



b. S required for PS = 0.7



c. S required for PS = 0.9



d. S required for PS = 0.95

Figure 3.10: The required S to achieve a predefined probability of exceeding the target FR of 0.95 (PS) in the case of one and multiple customers. The demand distribution for the single-customer case is Normal$(16, 3)$ and it is Normal$(\frac{16}{3}, \sqrt{3})$ for each customer in the multiple-customer case.

Figure 3.10 shows the *S* levels required for various PS values (i.e., 0.5, 0.7, 0.9 and 0.95) in overreaching the target FR of 0.95 for the cases where the demand distribution is normal. The results of Figure 3.10 are consistent with those of Figure 3.9.

The answer to the question (*What are the changes in the average FR and the probability of exceeding the target FR of each of multiple customers when compared to a single customer?)* follows. With short review periods, the supplier needs lower stock levels when there is more than one customer. As review period duration lengthens, a supplier with a single customer can hold lower stock levels than if there are multiple customers. This finding is consistent with the finding for the first research question.

3.6.2 Correlated customer demands

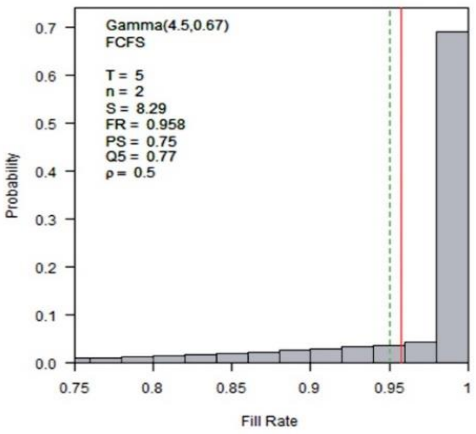
Question 3.3 is *How does the correlation structure of demand affect the average FR and the probability of overreaching the target FR of each customer?*

This simulation run was conducted with an FR for a finite horizon, to establish correlation between customers. The first design analysis establishes whether the (marginal) demand distribution of each customer case remains the same for both correlated and independent demand; thus, individual mean, individual variance and the total mean do not change and the total (sum of) variance of demand is higher (lower) in respect to positive (negative) correlation compared to independent demand. The next scenario has correlated demands, but the total variance of demands, as well as individual mean and total mean of demand, are the same as the independent demand case. With the NORTA algorithm (Cario & Nelson 1997; Niaki & Abbasi 2006) we can generate multivariate random data for given marginal distributions and correlation matrices.

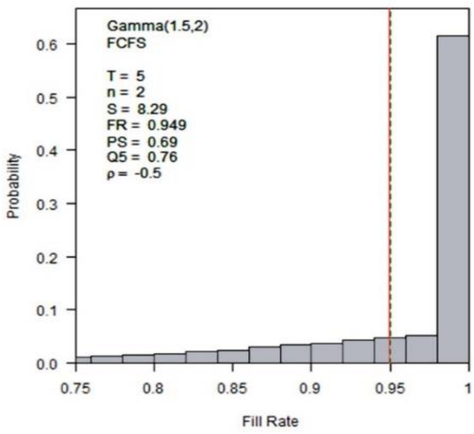
The Erlang distribution has one parameter, and it is not possible to control the aggregated mean and variance of demand when the customers' demands are correlated. It is necessary to inspect the impact of correlation on the FR distribution, and to eliminate pooling effects the aggregated demand should have the same mean and variance in all cases that must be compared. Therefore, two types of demand are applied: Gamma and normal distribution.

For the Gamma distribution scenarios, the marginal demand distribution of each customer in the correlated demand cases is Gamma (k $(1+\rho)$,1/ $(1+\rho)$), and the distribution of the demand in the one-customer case is Gamma(2k,1), which is equivalent to Erlang(2k).

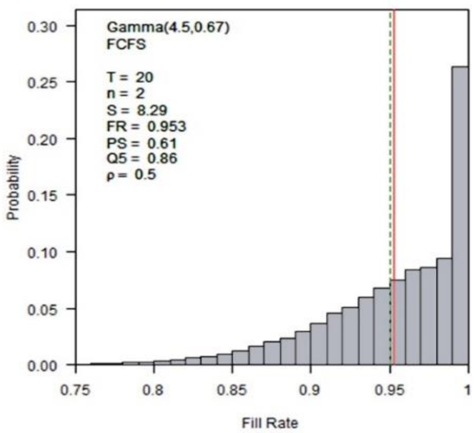
Here, ρ denotes the correlation coefficient between the demands of two customers. The total variance is 2k in both the multiple customers, with correlated demand case and the single-customer case. For the normal distribution scenarios, the marginal demand distribution of each customer in the correlated demand cases would be normal $N\left(\frac{\mu}{2}, \sqrt{\frac{\sigma^2}{1+\rho}}\right)$ and the demand distribution for the one-customer case is normal $N(\mu, \sqrt{2}\sigma)$. Figures 3.11 and 3.12 display the FR distributions of the two-customer cases when the demand is Gamma distributed. Figures 3.13 and 3.14 show the FR distributions of the two-customer cases case when the demand distribution is Normal.



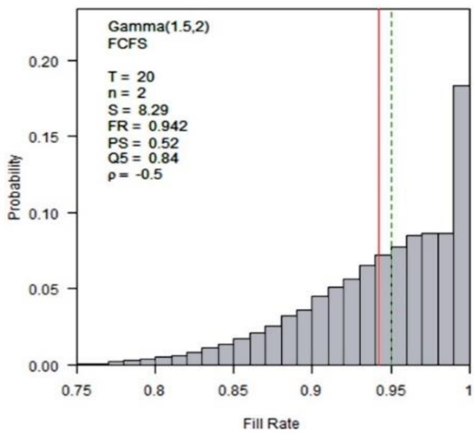
$a. D_i$~Gamma$(4.5, \frac{2}{3})$, T= 5 and $\rho = 0.5$



c. D_i~Gamma$(1.5, 2)$, T= 5 and $\rho = -0.5$

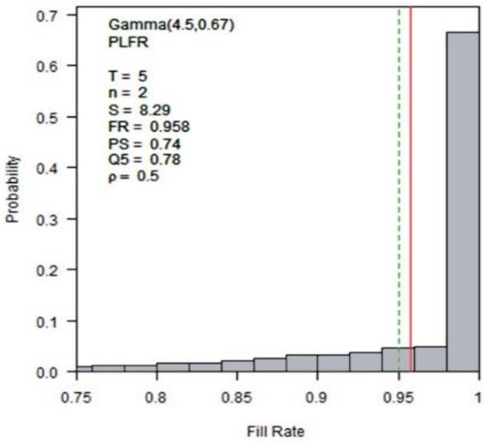


b. D_i~Gamma$(4.5, \frac{2}{3})$, T= 20 and $\rho = 0.5$

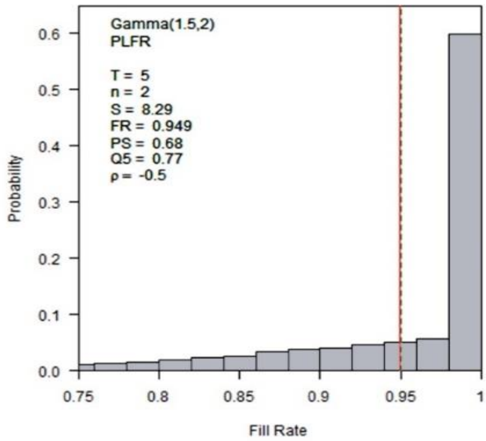


d. D_i~Gamma$(1.5, 2)$, T= 20 and $\rho = -0.5$

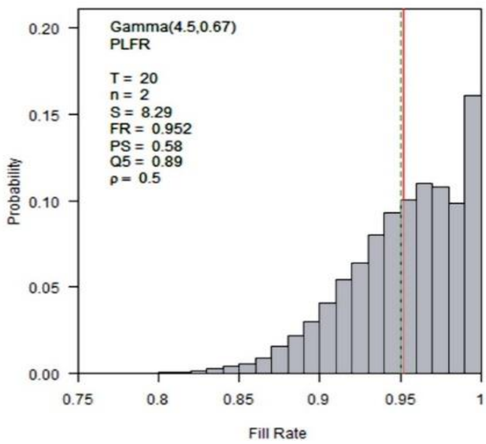
Figure 3.11: The FR distribution when customers' demand are correlated, the demand distribution for each customer is $\mathbf{Gamma}(3(1+\rho), 1/(1+\rho))$, FCFS is applied and $\rho = -0.5$ or 0.5.



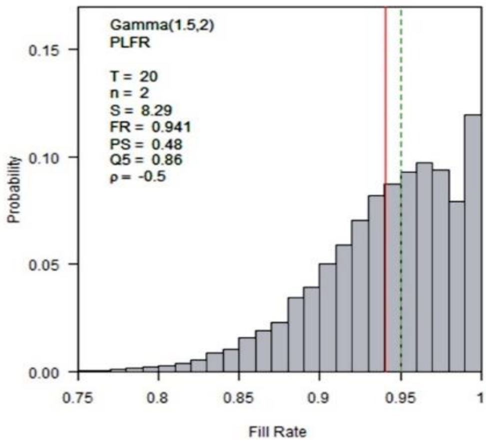
$a. D_i \sim \text{Gamma}(4.5, \frac{2}{3})$, T= 5 and $\rho = 0.5$



$b.\ D_i \sim \text{Gamma}(1.5, 2)$, T= 5 and $\rho = -0.5$

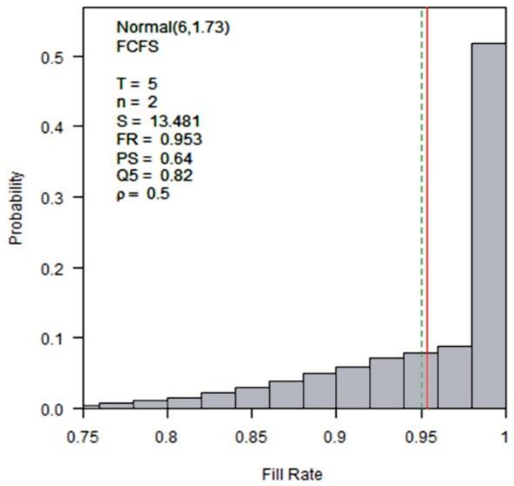


$c.\ D_i \sim \text{Gamma}(4.5, \frac{2}{3})$, T= 20 and $\rho = 0.5$

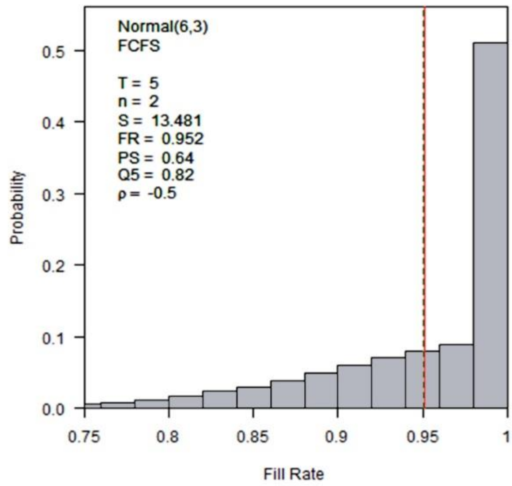


$d. D_i \sim \text{Gamma}(1.5, 2)$, T= 20 and $\rho = -0.5$

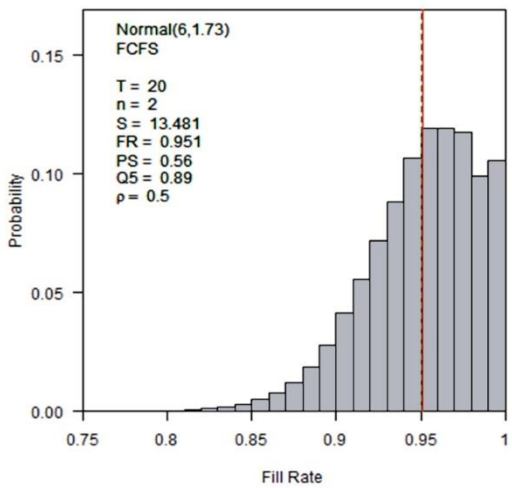
Figure 3.12: The FR distribution when customers' demands are correlated, the demand distribution for each customer is $\mathbf{Gamma}(3(1+\rho), 1/(1+\rho))$, PLFR is applied and $\boldsymbol{\rho = -0.5}$ or $\mathbf{0.5}$.



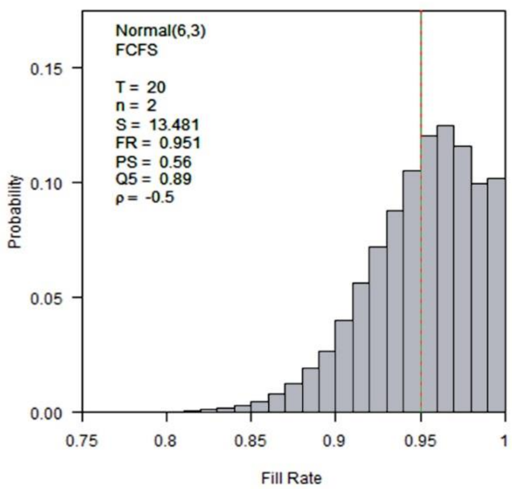
$a. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1+0.5}}$, T=5 and $\rho = 0.5$



$b. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1-0.5}}$, T=5 and $\rho = -0.5$

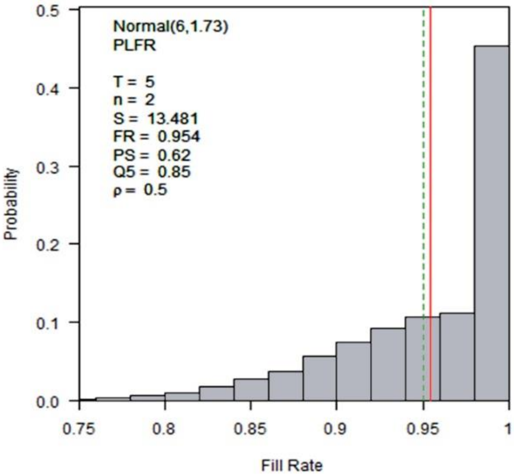


$C. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1+0.5}}$, T=20 and $\rho = 0.5$

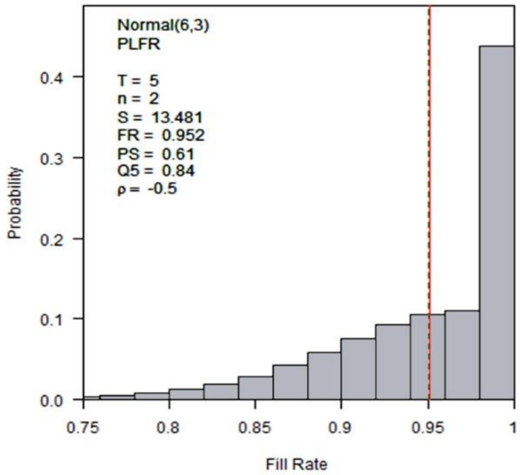


$d. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1-0.5}}$, T=20 and $\rho = -0.5$

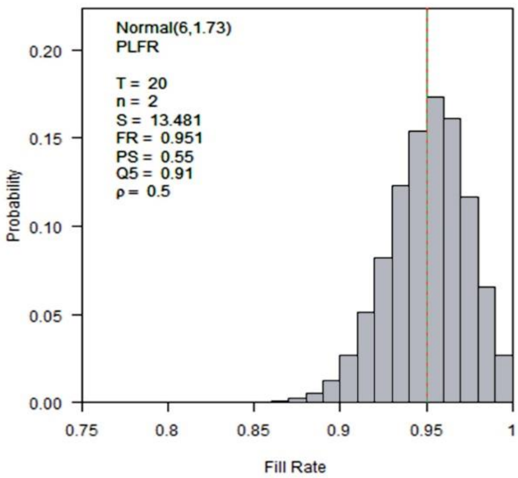
Figure 3.13: The FR distribution when customers' demands are correlated, the demand distribution for each customer is Normal$(6, \sqrt{\frac{4.5}{1+\rho}})$, FCFS is applied and $\rho = -0.5$ or 0.5.



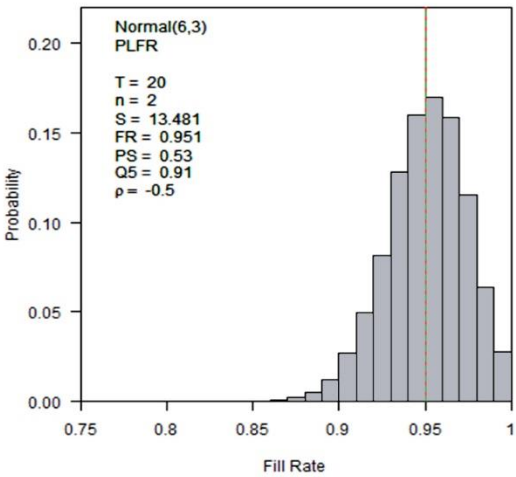
$a. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1+0.5}}$, T=5 and $\rho = 0.5$



$b. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1-0.5}}$, T=5 and $\rho = -0.5$



$C. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1+0.5}}$, T=20 and $\rho = 0.5$



$d. D_i \sim \text{Normal}(6, \sqrt{\frac{4.5}{1-0.5}}$, T=20 and $\rho = -0.5$

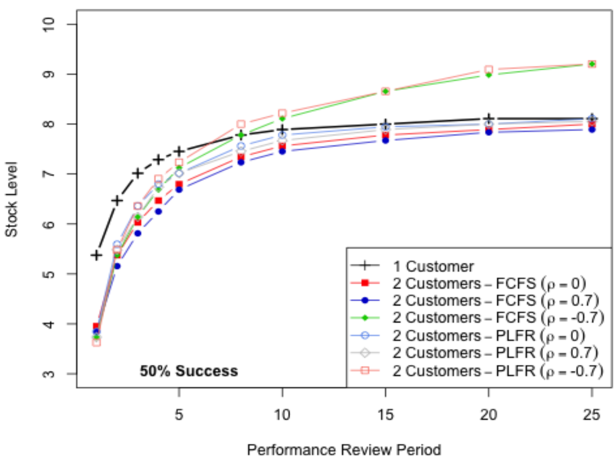
Figure 3.14: The FR distribution when customers' demands are correlated, the demand distribution for each customer is Normal$(6, \sqrt{\frac{4.5}{1+\rho}})$, PLFR is applied and $\boldsymbol{\rho = -0.5}$ or $\mathbf{0.5}$.

Figures 3.11–3.14 display the FR distribution for two-customer cases with correlated demand. We examined two different demand distributions to validate the results. In Figures 3.11 and 3.12, the demand distribution is Gamma, and in Figure 3.13 and 3.14 it is normal. Similarly, the service policy in Figures 3.11 and 3.13 is the FCFS policy, while it is the PLFR policy in Figures 3.12 and 3.14. In these experiments, the mean and variance are fixed while the correlation coefficient changes. Hence, for Gamma, the distribution of demand for each customer is Gamma(3(1+ ρ),1/ (1+ ρ)) and for normal, the distribution of demand for each customer is normal $N(6, \sqrt{\frac{4.5}{1+\rho}})$, where ρ is set to 0.5 and -0.5 in the results presented in Figures 3.11, 3.12, 3.13 and 3.14. Thus, the results of the first part of question 3.3 (*How does the correlation structure of demand affect the average FR?)* show that the average FR is higher when customers' demands are positively correlated.

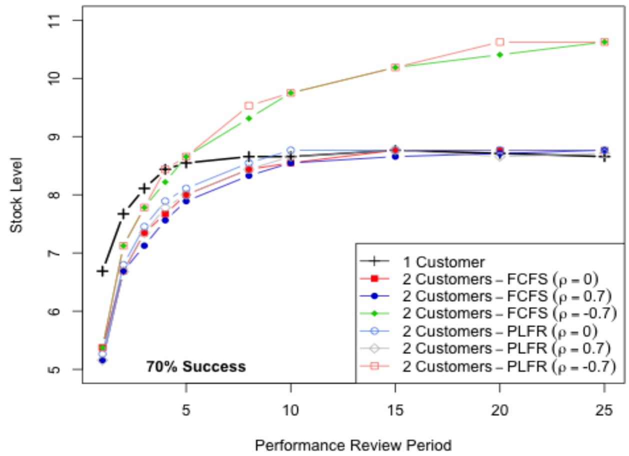
Comparison of the positive and negative correlated demand is important to see the effect of positive and negative correlations. Next, we compared the effect of positive and negative correlated demands by the seeking the required base stock level (*S*) to achieve a target PS.

In Figure 3.15, we presented the required base stock level to achieve a given probability of exceeding the target FR of 0.95 in the case of having one customer with demand distribution of Erlang (6), and in the cases of having two correlated customers each with demand distribution of Gamma (3 (1+ ρ),1/ (1+ ρ)). The correlation coefficient (ρ) is set to 0.7 and -0.7 to analyse the impact of positively and negatively correlated demand. Note that the aggregate average and variance of demand are fixed when ρ is -0.7, 0 or 0.7.

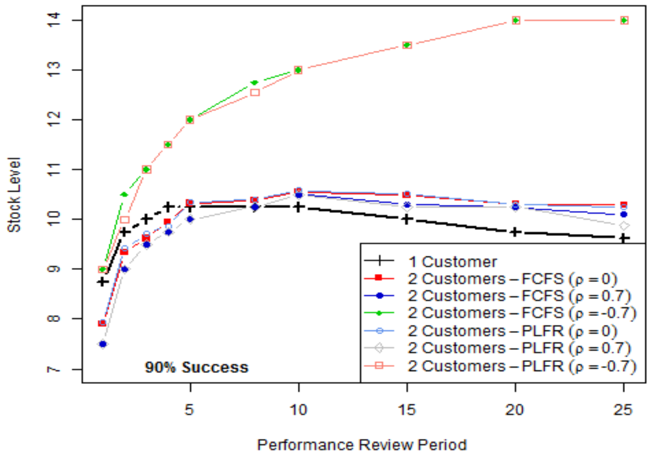
The second part of the related research question concerned the impact of correlation in demands on the probability of exceeding the target FR, and this is presented in Figures 3.15 and 3.16. In Figure 3.15, the results display two different policies. FCFS and PLFR behave similarly in terms of the base stock level required to achieve a given PS. Additionally, a supplier must hold a higher base stock level to achieve a given PS when the customers' demands are negatively correlated, compared to cases when demands are independent. Conversely, when customers' demands are positively correlated, the base stock level required to achieve a given PS is lower than in cases with independent demands. The reason for this observation is that when correlation is negative, there is a high probability that in a given period the demand for one customer will be large, while the demand for the other customer will be small. In turn, there is a low chance that in a given period both customers will have a small demand. Therefore, the PS reduces when the demands are negatively correlated. Results similar to those in Figure 3.15, but for normally distributed demand, are presented in Figure 3.16. Figure 3.16 is similar to Figure 3.15 in the main text, where the demand distribution is normal. It investigates the impact of correlated demand on the probability of exceeding the target FR in multiple-customer cases. The findings are consistent with those of Figure 3.15.



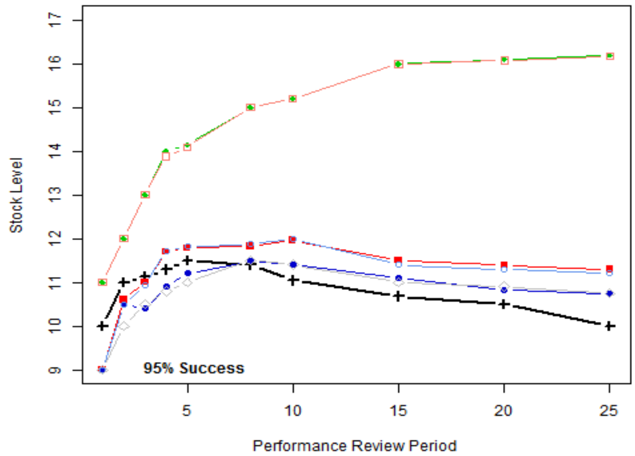
a. The required S for achieving PS = 0.5



b. The required S for achieving PS = 0.7

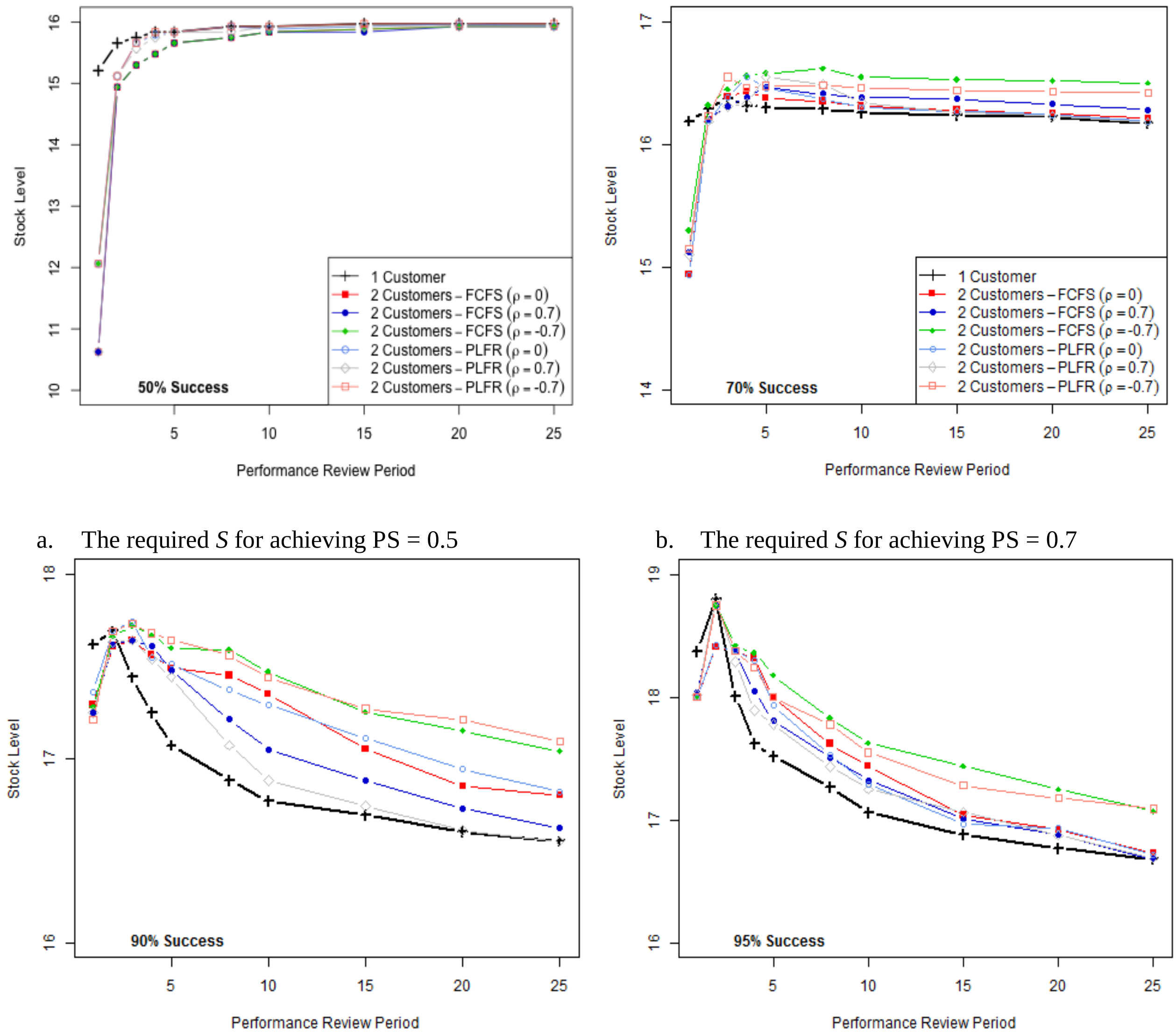


c. The required S for achieving PS = 0.9



d. The required S for achieving PS = 0.95

Figure 3.15: The comparison between one customer (with demand of Erlang (6)), two independent customers (each with demand of Erlang(3)) and two customers with correlated demands (each with Gamma(3(1+ ρ),1/(1+ ρ))) in terms of the stock levels needed to meet a 95% target FR with various probabilities of success. The total mean and variance of demand are fixed in all cases.



a. The required *S* for achieving PS = 0.5

b. The required *S* for achieving PS = 0.7

c. The required *S* for achieving PS = 0.9

d. The required *S* for achieving PS = 0.95

Figure 3.16: The required S to achieve a predefined probability of exceeding the target FR of 0.95 in the multiple-customer case when the customers' demands are correlated. The demand distribution each customer is Normal $\left(8, \sqrt{\frac{4}{1+\rho}}\right)$, and where ρ is the correlation coefficient between customers' demands. The overall demand distribution is Normal$(16, 2\sqrt{2})$.

The impact of correlation in customer demands was found that negative correlation gives a higher probability of one customer's demand exceeding that of a second customer. Thus, the probability of FR exceeding a set target is reduced. In answer to the question, "*How does the correlation structure of demand affect the average FR and the probability of overreaching the target FR of each customer?*", when customers' demands are positively correlated, the probability of exceeding a FR level of 0.95 is higher; with a negative correlation, the probability is lower.

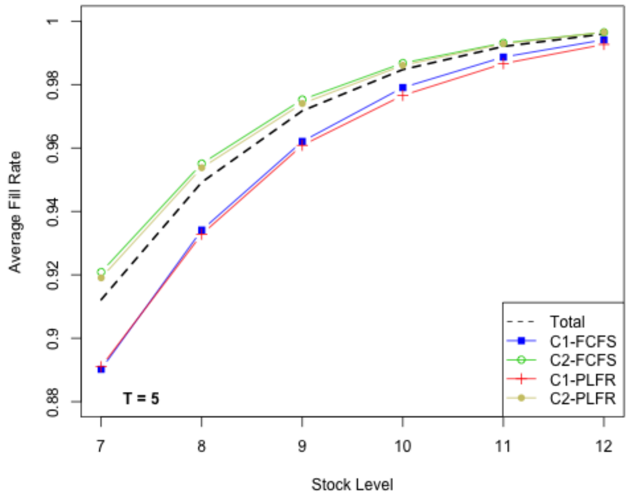
3.7 What Achieves Better Service for Customers: Smaller or Larger Demand?

The final question for this chapter, Q3.4, was: *Who receives the better service (higher average FR and higher probability of exceeding the target FR): the customer with higher demand or the customer with lower demand?*

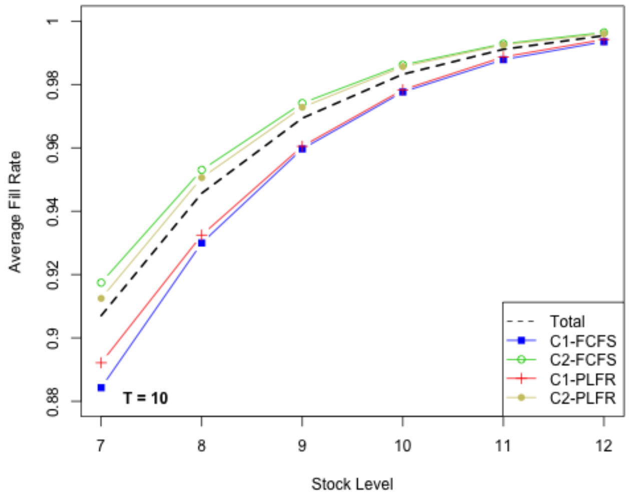
To answer this question, we consider the case where there are multiple customers with different demand distributions and; the supplier's base stock level is set to achieve an average FR for overall demand. Of the two policies, in FCFS the sequence of customer arrivals is random, so on day *t* the chance that customer 1 arrives after customer 2 is 50%. The results are examined for two situations. In the first scenario, the larger customer has a larger mean and variance of demand. In the second scenario, the larger customer has a larger mean of demand, but the variance of demand is same as the smaller customer.

3.7.1 Larger mean and larger variance of demand

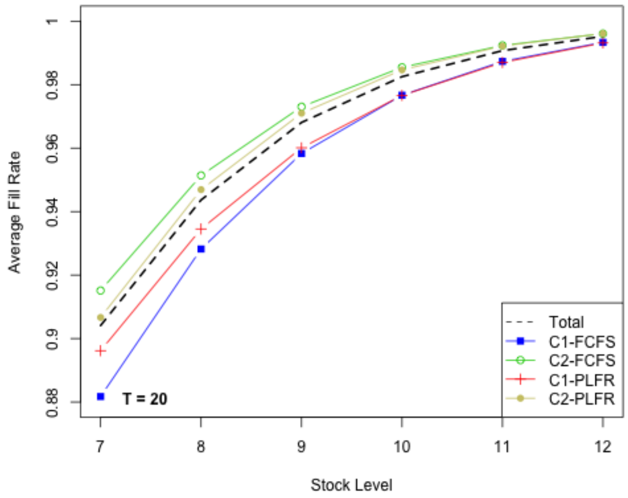
In this scenario, suppose that the supplier is facing two customers and a total demand distribution of $\mathrm{Erlang}(6)$. Customer 1 has a demand of $\mathrm{Erlang}(2)$ and customer 2 has a demand of Erlang (4). In Figure 3.17, for different values of *S* and *T*, the expected FR value for each customer is displayed. Both service policies (FCFS and PLFR) show that the customer with higher demand will have a higher expected FR value. The difference between the average FR of the larger customer and the smaller customer is higher for shorter *T*. Figure 3.18 displays the probability of exceeding a target FR of 0.95 for different values of *S* and *T*. The results in Figure 3.18 determine that in both service policies, the customer with the larger demand has a higher probability of exceeding the target FR than the customer with the smaller demand. Between the larger and smaller customers, the difference in the PS is higher for longer *T*. This is because when satisfying the demand of the small customer first (either with the FCFS or PLFR policy), due to the smaller customer's lower variance of demand, it is still highly likely that there is sufficient stock available for the larger customer. Similar plots for the case when the overall demand distribution is $\mathrm{Normal}(12,2)$ and the demand distribution of the customer 1 and 2 are $\mathrm{Normal}(5,1)$ and $\mathrm{N}(7,\sqrt{3})$ respectively, are presented in Figures 3.19 and 3.20.



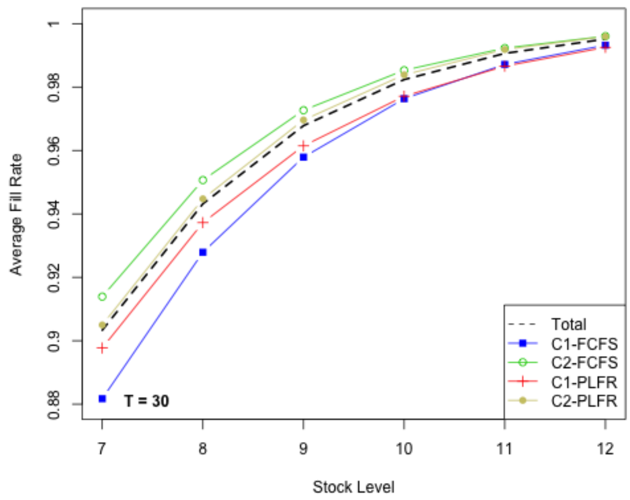
a. FR when *T*=5 and *S* varies from 7 to 12



b. FR when *T*=10 and *S* varies from 7 to 12

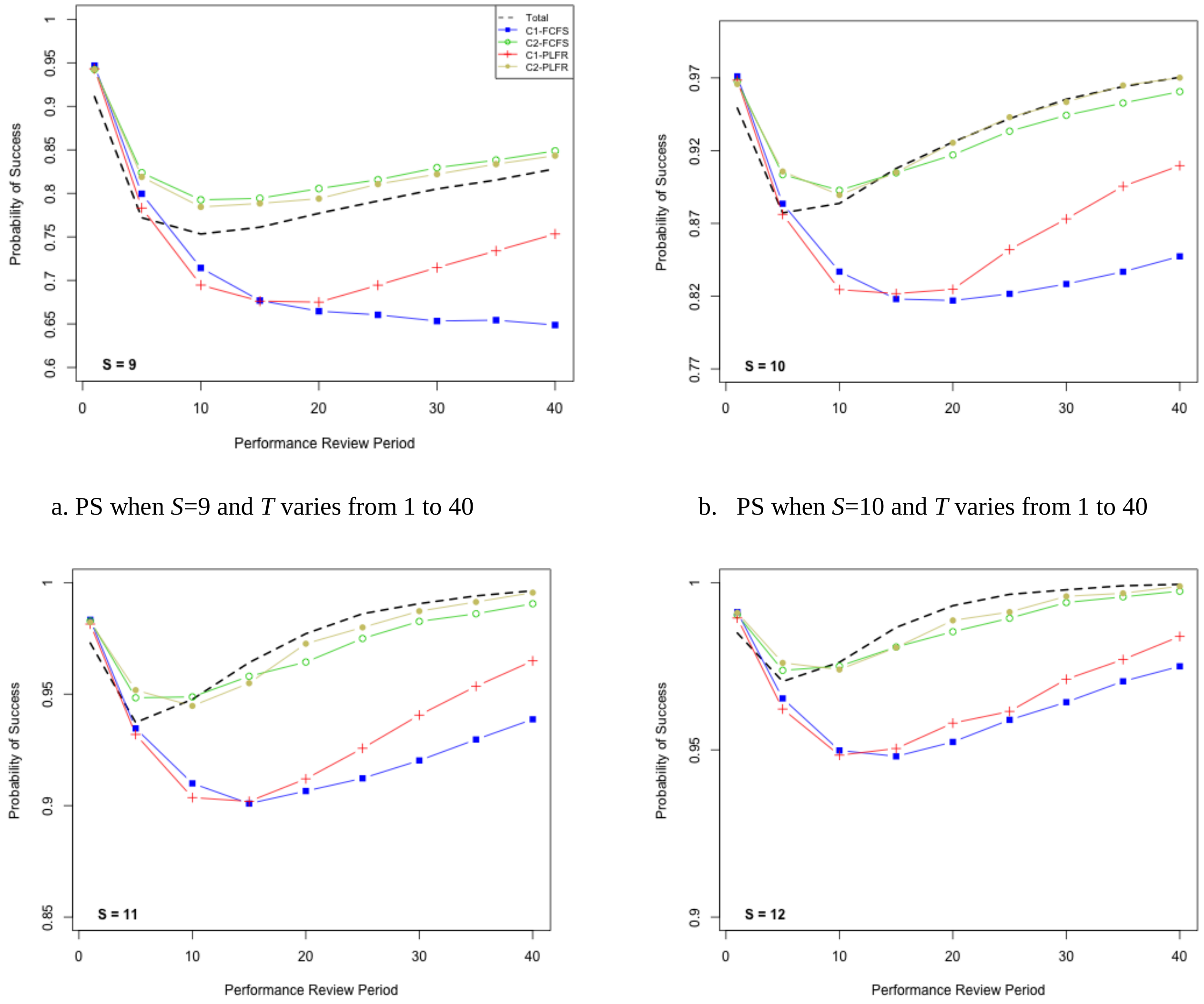


c. FR when *T*=20 and *S* varies from 7 to 12



d. FR when *T*=30 and *S* varies from 7 to 12

Figure 3.17 – The average fill rate for customers 1 and 2 (C2 is the larger customer). The demand distribution for customers 1 and 2 are Erlang(2) and Erlang(4) respectively. The base stock level (x-axis) changes from 7 to 12 and four review periods T=5, 10, 20 and 30 are considered. The dashed line shows the average fill rate for a single customer with an Erlang(6) demand distribution.



a. PS when S=9 and T varies from 1 to 40

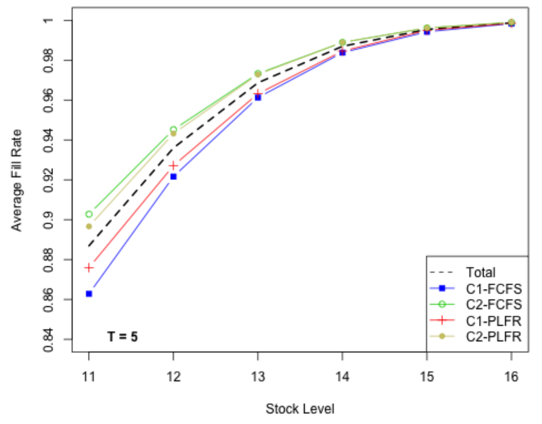
b. PS when S=10 and T varies from 1 to 40

c. PS when S=11 and T varies from 1 to 40

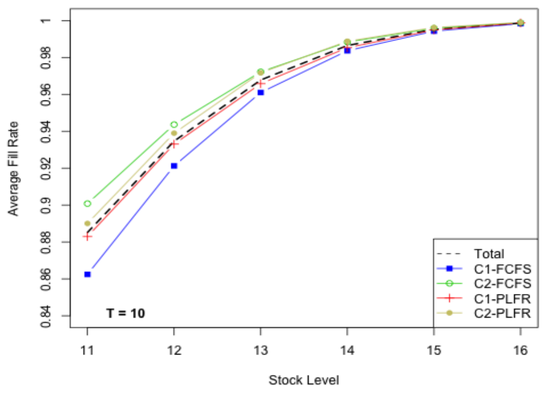
d. PS when S=12 and T varies from 1 to 40

Figure 3.18: The probability of exceeding the target FR of 0.95 for customers 1 and 2 (C2 is the larger customer). The demand distribution of customers 1 and 2 are Erlang(2) and Erlang(4), respectively. The dashed line shows the average FR for a single customer with an Erlang(6) demand distribution.

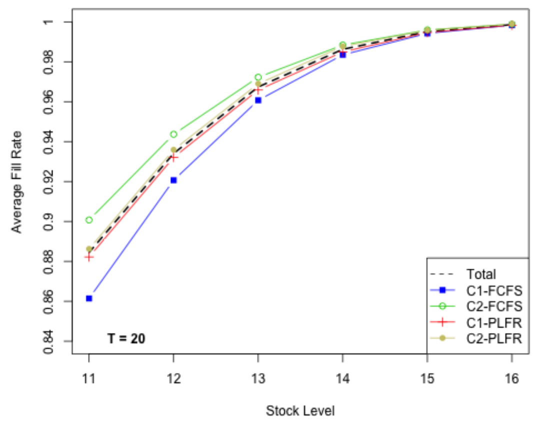
Figures 3.19 and 3.20 show the average FR and the probability of overreaching the target FR (0.95) for single and multiple-customer cases, in the multiple-customer case there are two customers, one of them with a higher demand. Figures 3.19 and 3.20 are similar to Figures 3.17 and 3.18 but here the demand distribution is normal. The results and remarks are consistent with those of Figure 3.17 and 3.18.



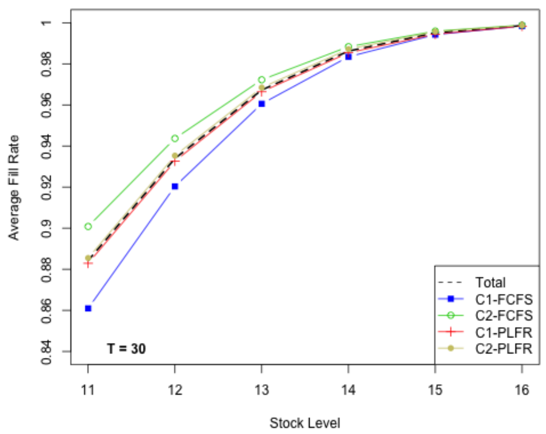
a. FR when $T = 5$ and S varies from 11 to 16



b. FR when $T = 10$ and S varies from 11 to 16

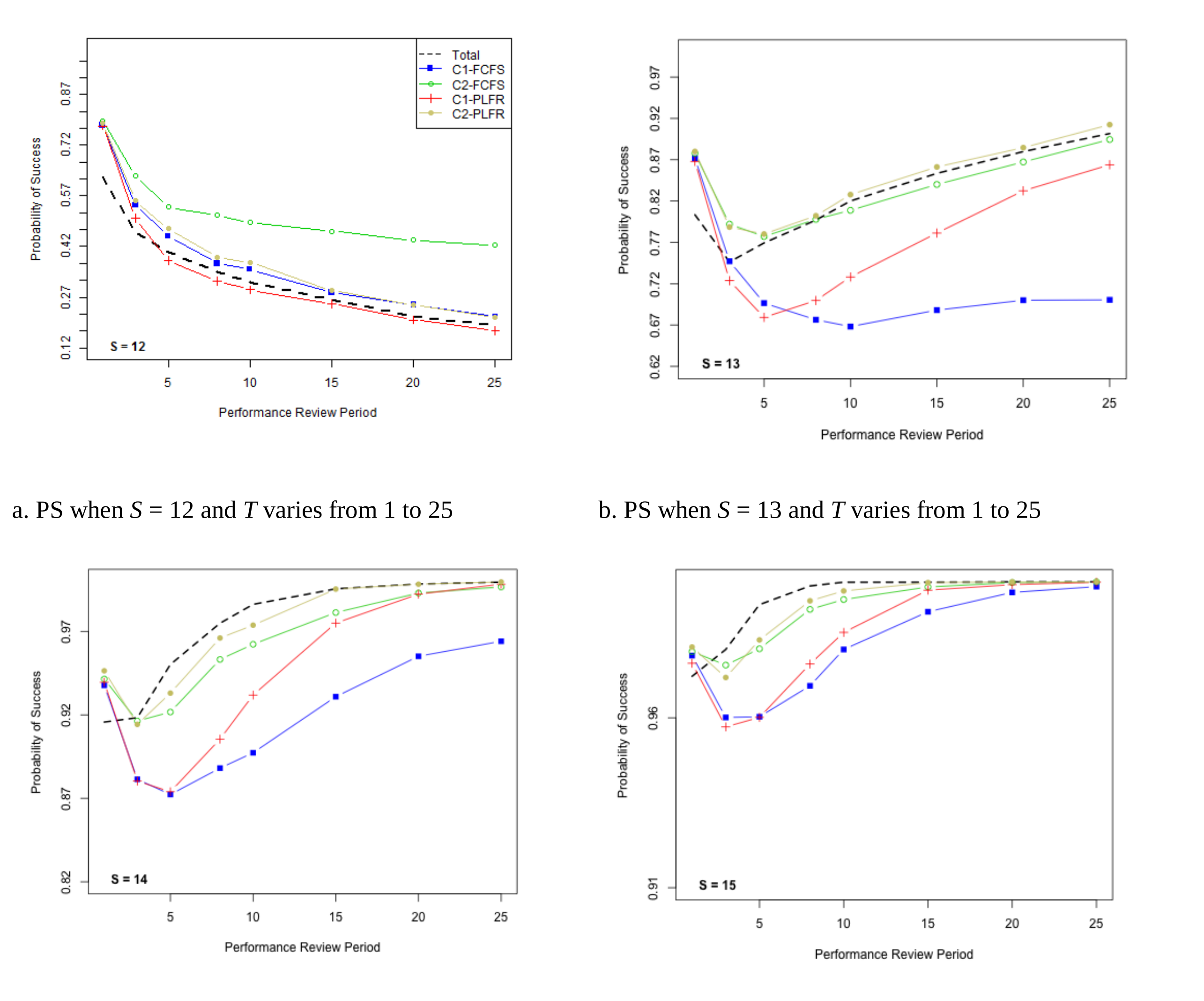


c. FR when $T = 20$ and S varies from 11 to 16



d. FR when $T = 30$ and S varies from 11 to 16

Figure 3.19: The average FR for customers 1 and 2 (customer 2 has the larger average and variance of demand). The demand distribution for customers 1 and 2 are $\mathrm{Normal}(5, 1)$ and $\mathrm{Normal}(7, \sqrt{3})$ respectively. The dashed line shows the average FR for the system with a single customer with demand distribution of $\mathrm{Normal}(12, 2)$.



a. PS when *S* = 12 and *T* varies from 1 to 25

b. PS when *S* = 13 and *T* varies from 1 to 25

c. PS when *S* = 14 and *T* varies from 1 to 25

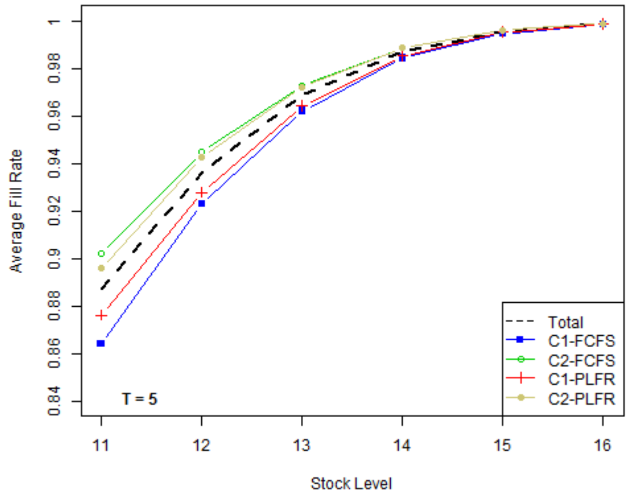
d. PS when *S* = 15 and *T* varies from 1 to 25

Figure 3.20: The probability of exceeding the target FR of 0.95(PS) for customers 1 and 2 (customer 2 has larger average and variance of demand). The demand distribution for customers 1 and 2 are $\mathbf{Normal}(5, 1)$ and $\mathbf{Normal}(7, \sqrt{3})$, respectively. The dashed line shows the average FR for the system with a single customer with demand distribution of $\mathbf{Normal}(12, 2)$.

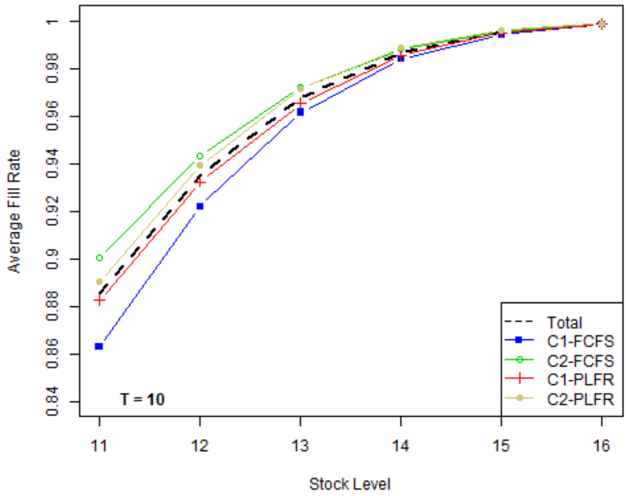
Overall, evident from the results two customers have different demands, the supplier achieves a higher FR and a higher PS for the larger customer. The difference in PS becomes greater as the review period is lengthened. This occurs because when serving the smaller customer first, there is a still a good probability that there is sufficient stock for the larger customer's order to be satisfied.

3.7.2 Customers with high demand (mean) but same variance as customer with lower demand.

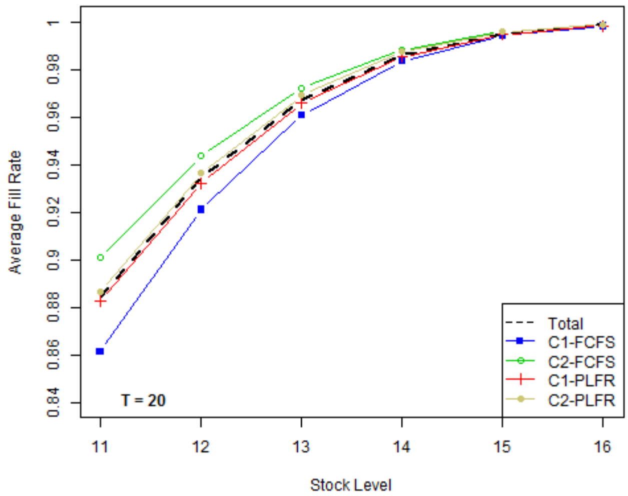
In this subsection, we explore the case where customers have different mean demand but the same variance of demand. If the supplier serves two customers and the overall demand distribution is $\mathrm{Normal}(12{,}2)$, the demand of customer 1 is $\mathrm{Normal}(5, \sqrt{2})$ and the demand of customer 2 is Normal $(7, \sqrt{2})$. Figure 3.21 displays the expected FR value for each customer for different values of *S* and *T*. Figure 3.22 displays the probability of exceeding a target FR of 0.95 for different values of S and T. The results are similar to the results in subsection 3.7.1, meaning that the customer with the larger demand will have a higher expected FR value (i.e., $\min_i\{\bar{\alpha}^{C_i}(S,T)\}$ belongs to the smaller customer) and will similarly have a higher opportunity of exceeding the target FR than the customer with the smaller demand. Moreover, the difference between the average FR of the larger customer and the smaller customer is higher for shorter values of *T*. Conversely, the difference between the probability of exceeding the FR between the customer with larger demand and the customer with smaller demand is higher for longer *T*. This finding is due to the fact that when satisfying the demand of the smaller customer first (either with the FCFS or PLFR policy) due to its lower mean demand, it is still highly likely that there is sufficient stock available for the larger customer. The results of another experiment when two customers have the same mean demand but the variance of the demand is higher for customer 2 is presented in Figure 3.23. In this situation, the average FR is the same for both customers; however, the probability of exceeding the target FR is higher for the customer with larger variance.



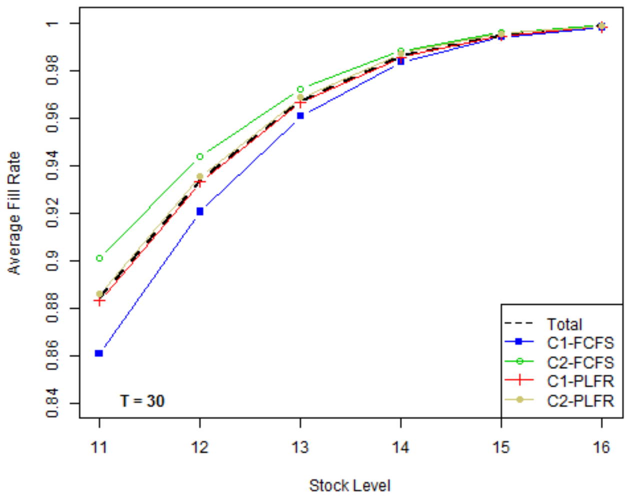
a. FR when T=5 and S varies from 11 to 16



b. FR when T=10 and S varies from 11 to 16

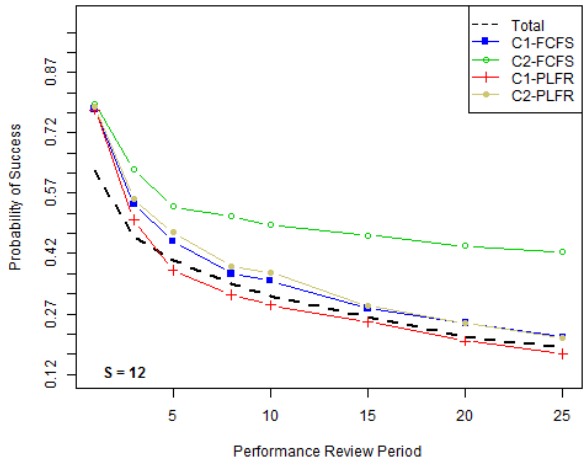


c. FR when T=20 and S varies from 11 to 16

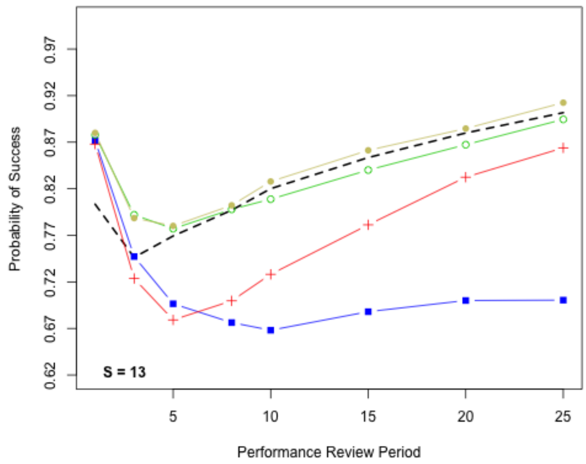


d. FR when T=30 and S varies from 11 to 16

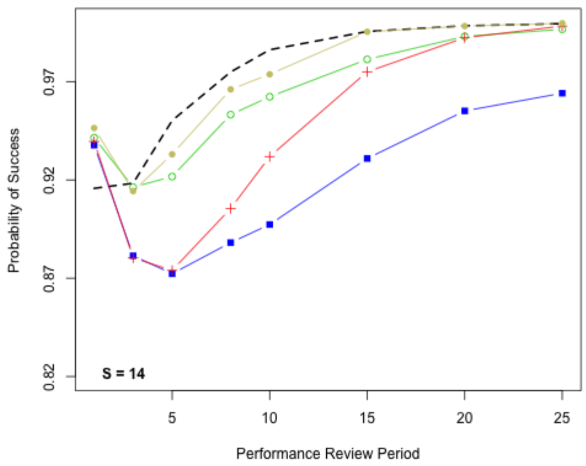
Figure 3.21: The average FR for customers 1 and 2 (C2 is the larger customer). The demand of customer 1 is Normal$(5, \sqrt{2})$ w and the demand of customer 2 is Normal (7, $\sqrt{2}$). The dashed line shows the average FR for a single customer with a Normal$(12, 2)$ demand distribution.



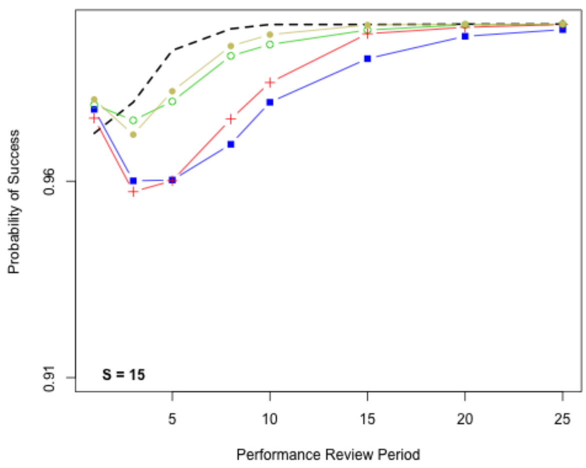
a. PS when S=12 and T varies from 1 to 25



b. PS when S=13 and T varies from 1 to 25



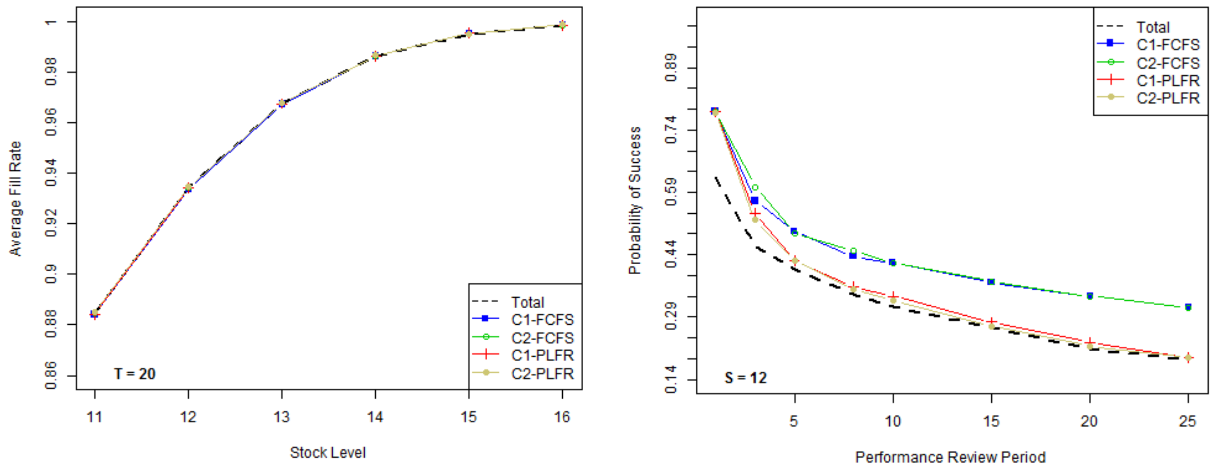
c. PS when S=14 and T varies from 1 to 25



d. PS when S=15 and T varies from 1 to 25

Figure 3.22: The probability of exceeding the target FR of 0.95 for customers 1 and 2 (C2 is the larger customer). The demand of customer 1 is Normal$(5, \sqrt{2})$ w and the demand of customer 2 is Normal $(7, \sqrt{2})$. The dashed line shows the average FR for the system with a single customer with a Normal$(12, 2)$ demand distribution.

The following plots on Figure 3.23 present the case that both customers have the same average of demand but different variance of demand. When both customers have the same average of demand, the average FR for both customers are the same. However, the customer with higher variance of demand has a higher probability of overreaching the target FR.



a. FR when T=20 and S varies from 11 to 16

b. PS when S=12 and T varies from 1 to 25

Figure 3.23: The average FR for customers 1 and 2 (customer 2 has larger variance of demand). The demand distribution for customers 1 and 2 are Normal$(6, 1)$ and Normal$6, \sqrt{3})$ respectively. The dashed line shows the average FR for the system with a single customer with demand distribution of Normal$(12, 2)$.

The final research related question of this chapter was: '*who receives the better service (higher average FR and higher probability of exceeding the target FR): the customer with higher demand or the customer with lower demand?'* The answer is the customer with higher demand. Even if the lower demand customer was served first, the probability was that there would be sufficient stock to fill a larger order.

3.8 Chapter Summary

In this chapter, we examined the behaviour of FR over different performance review periods when there are multiple customers in the SC. We studied the FR measure, which is the most common service level measure. The findings provide an opportunity for the suppliers in design and negotiation of SLA. Two cases were studied and designed in our experiments: one customer and multiple customers. When the supplier faces one customer, the demand distribution for the single-customer case was similar to the aggregated demand distribution in the multiple-customer case. In the multiple-customer case we considered two different policies for fulfilling customers' demands: a FCFS policy and a second policy (called PLFR), based on prioritising customers due to current measured FR performance.

Our findings and insights were entirely consistent when using either service policy. First, in both the single and multiple-customer cases, it was shown that when the performance review period duration is increased, the average FR decreases. Second, in both the single and multiple-customer cases, we proved that the probability of exceeding a target FR is highly dependent on the length of the performance review period T. When T is short (i.e., less than a threshold value θ), the supplier requires less stock in the multiple-customer case. Conversely, when T is longer than θ, the supplier is required to stock a higher base stock level in the single-customer case. The value of this threshold θ decreases when a supplier chooses to achieve a higher probability of exceeding the target FR.

In this chapter, we studied the multiple-customer cases with correlated demand. From our results, we can see that when the correlation between customers' demands is negative, the supplier is required to maintain a higher base stock level than in the case of independent demands; the opposite holds when the correlation between customers' demands is positive. Finally, we examined multiple customers when the demand of one of the customers is larger than the other. We observed that the customer with the higher average demand has a higher FR and higher probability of exceeding the target FR.

Chapter 4: The Impact of Lead Time on the Finite Horizon FR in Single and Multiple-Customer Cases

4.1 Introduction

In the previous chapter, the supplier's FR for inventory was measured over varying review period lengths for single and multiple customers with zero lead time. The objective of this chapter is to examine the effect of lead times on the finite horizon FR, as existing studies assume a zero lead time for the supplier (Abbasi et al. (2017); Arıkan, Fichtinger & Ries 2014; Thomas 2005). This chapter sets research questions to guide our analyses. We study the effect of lead times on the distribution of the FR and base-stock level over finite horizons, with various review period lengths for both single and multiple customers.

In response to growing global industrialisation, there is an ongoing challenge in the prevailing economic environment to deliver products to customers in a short time (Schwartz & Rivera 2010). On the one hand, holding inventory imposes costs on the suppliers, while it increases customer satisfaction by avoiding any delays to respond to demand. A stock-out may not only cancel the order, it will also affect the probability of future customer demand (Anderson, Gavan & Duncan 2006). Gruen (2002) showed that 45% of customers facing a stock-out will buy products from another store. Intense competition between companies means additional pressure to provide and deliver products to customers on time, with many firms placing greater focus on improving their level of service to provide greater customer satisfaction (Vigoroso 2005). Seller and customer connection is the main point in assisting the flow of products. In an SC the delay between two successive members (i.e., ordering delay, shipment, or information flow) will result in an interruption in product delivery (Heydari 2014). Improper lead time planning can cause either large inventories or a low customer service levels (Louly & Dolgui 2013). Sandvig and Allaire (1998) found that a lack of customer service plays a pivotal role in a company losing its customer base. Currently, the increase in global trade with offshore migration of US manufacturing productions has affected SCs with longer and added indefinite lead times (Blackburn 2012).

4.2 Related Research Questions in this Chapter:

The following questions serve to accomplish the research objective:

Q4.1 How does lead time impact base stock levels for single and multiple customers cases over various review period lengths?

Q4.2 How does lead time affect the average FR for single and multiple customers?

Q4.3 How does lead time affect the probability of exceeding the target FR for single and multiple customers?

Q4.4 How does lead time together with the correlation of demand affect the probability of overreaching the target FR for each of multiple customers?

4.3 Model and Notations

The notations used are as follows:

L: lead time;

T: review period for which the FR is calculated;

S: base stock level;

n: number of customers;

C_i: customer index or reference to customer i (i =*1, 2,…,n*);

$\alpha_t^{C_i}(S, L, T)$: FR for customer i at period *t* when the review period is T, base stock is S and lead time of L. $\alpha_T^{C_i}(S, L, T)$ is a random variable. The customer index is deleted for identical customers;

X_{it}: demand that is filled for customer i at time t (i is a discrete time (e.g., a day or week);

Y_{it}: demand of customer i at time t;

$\bar{\alpha}^{C_i}(S, L, T)$: average FR for customer i when the base stock is S, lead time is L and review period is T. The customer index is deleted for identical customers;

$P_{\pi}^{C_i}(S, L, T)$: probability of exceeding the FR of π for customer i when base stock level is S, lead time is L and review period is T. The customer index is deleted for identical customers;

Demand Y_{it} per period is a non-negative random variable, and is *i.i.d.* from period to period. Demand arrives from the start of time period *t*, but is not fulfilled until the end of the period. Inventory is reviewed at the end of every period *t*, after demand is satisfied, and an order is placed to increase inventory (in-hand and receivable inventory), the order arrives after L (lead time).

A stationary stocking policy does not need to be optimal for a supplier who is looking to meet or overestimate the target FR over a finite horizon. The policies are intended to establish good practice; thus, this simulation is limited to order-up-to policies. The FR at period *t* is defined as:

$$\alpha_t^{C_i}(S, \mathrm{L}, T) \equiv \frac{X_{i(t-T+1)} + \cdots + X_{i(t)}}{Y_{i(t-T+1)} + \cdots + Y_{i(\mathrm{t})}} \tag{4.1}$$

the average FR for customer i, as a function of T and S, is defined as:

$$\bar{\alpha}^{C_i}(S, \mathrm{L}, T) = \mathrm{E}\left(\frac{X_{i(t-T+1)} + \cdots + X_{i(t)}}{Y_{i(t-T+1)} + \cdots + Y_{i(t)}}\right) \tag{4.2}$$

the probability of exceeding the target fill rate (π) is:

$$P_{\pi}^{C_i}(S, \mathrm{L}, T) = \Pr(\alpha_t^{C_i}(S, \mathrm{L}, T) \geq \pi) \tag{4.3}$$

For example, assume there are three customers, with a base stock level of 10, the lead time of 4, the length of the review period being 3 (e.g., three days) and in the period *t* (e.g., the t^{th} day), the demand and filled demand for customer i are Y_{it} and X_{it}, respectively. The average FRs for the three customers are:

$$\bar{\alpha}^{C_1}(10,4,3) = \mathrm{E}(\frac{X_{11} + X_{12} + X_{13}}{Y_{11} + Y_{12} + Y_{13}}), \bar{\alpha}^{C_2}(10,4,3) = \mathrm{E}(\frac{X_{21} + X_{22} + X_{23}}{Y_{21} + Y_{22} + Y_{23}}),$$

$$\bar{\alpha}^{C_3}(10,4,3) = \mathrm{E}(\frac{X_{31} + X_{32} + X_{33}}{Y_{31} + Y_{32} + Y_{33}})$$

Conversely, if all demand came from a single customer, the average FR would be:

$$\bar{\alpha}(10,4,3) = \mathrm{E}(\frac{X_{11}+X_{12}+X_{13}+X_{21}+X_{22}+X_{23}+X_{31}+X_{32}+X_{33}}{Y_{11}+Y_{12}+Y_{13}+Y_{21}+Y_{22}+Y_{23}+Y_{31}+Y_{32}+Y_{33}}) \tag{4.4}$$

The next section includes an analysis of the relationship between the average FRs and lead time, and between the probabilities of exceeding the target FR and lead time, considering various review period lengths, and one or more customers in the system.

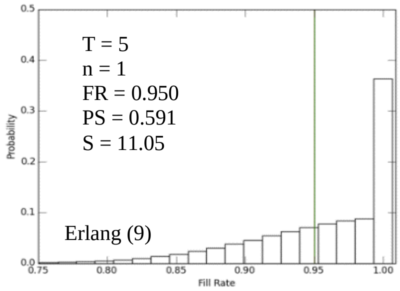
4.4 Analysis of FRs, Lead Times and Exceeding FR

Similar to Thomas (2005) and Abbasi et al. (2017), we developed a simulation model to study the impact of lead time on FR behaviour. The simulation model is a discrete event simulation and was run for an extended time to collect the data for the designed scenarios. The scenarios are different in terms of the lead time, the number of customers, customers' demand, the correlation between customers' demands and the length of the review period in the SLA. We assumed an *Erlang* distribution for the demand because it is flexible and covers various shapes (different skewness and kurtosis). Additionally, the *Erlang* distribution has been used by Thomas (2005) and Abbasi et al. (2017) to study the FR distribution. It is assumed that a base stock policy is used by the supplier, and an order from the supplier to replenish inventory is received after L periods (lead time) from the point that the order is placed.

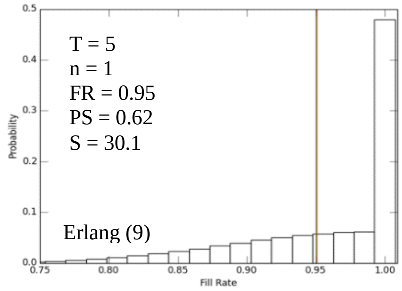
All experiments were performed on an Intel Core i7 CPU with a 2.2 GHz processor and 16 GB RAM. The codes were written in Python version (3.4). To solve these models we applied other packages, such as Numpy, Sicipy, Matplotlib and Anaconda. It takes a long time to reach results when solving simulation algorithm codes in Figures 4.1 - 4.10 (in some cases 2–5 days, especially when we increase the performance review period lengths), due to many loops and the desired accuracy of the results.

4.4.1 FR distribution

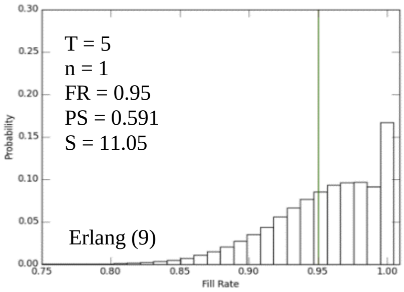
First, we look at the distribution of FR for various lead times and numbers of independent customers in the SC. In each case, we determined *S* (the base stock level) such that the long-run average FR becomes 0.95. Additionally, we assumed a target FR of 0.95, which means that achieving an FR above 0.95 in each period is considered a '*success*'. Figure 4.1 shows the distribution of the FR for different lead times and different review period durations when there is one customer. There are two vertical lines in each plot; the green dashed line indicates the threshold of a 0.95 long-run FR, while the solid red line shows the average FR. The demand distribution is the same in all subplots in Figure 4.1, i.e. Erlang (9). In Figure 4.1.a, S is set to 11.05 to assure a long-run FR of 0.95. Comparing the subplots in Figure 4.1 displays that to achieve the same long-run FR when the lead time increases in all cases (different review periods), a higher base stock level is required. However, when lead time increases, the probability of exceeding the target FR also increases. Figure 4.1 also shows that by increasing the review period duration, the probability of exceeding the target FR decreases (thus confirming the existing literature). Figure 4.2 presents the distribution of FR, similar to Figure 4.1, however in Figure 4.2 there are three independent customers each with an Erlang (3) demand distribution (i.e. the overall distribution is Erlang (9), similar to the one customer cases).



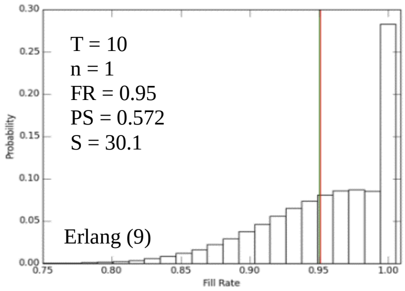
a. One customer with a demand distribution of Erlang (9) and T = 5, L = 1.



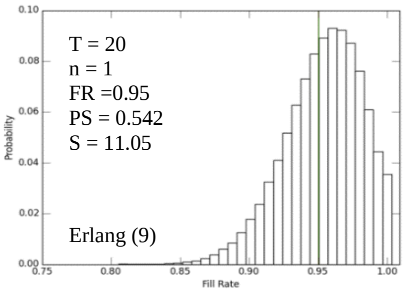
b. One customer with a demand distribution of Erlang (9) and T = 5, L = 3.



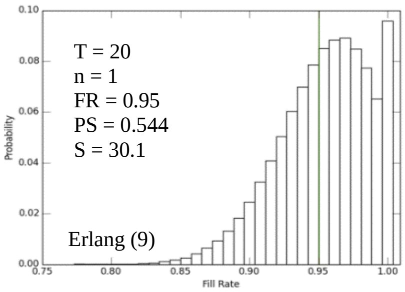
c. One customer with demand distribution of Erlang (9) and T = 10, L = 1.



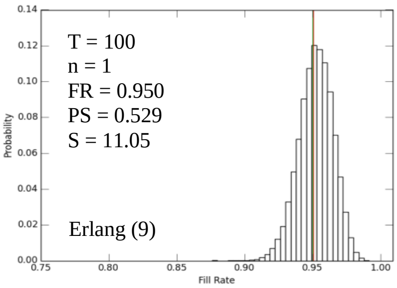
d. One customer with a demand distribution of Erlang (9) and T = 10, L = 3.



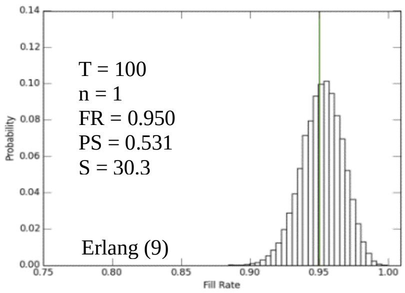
e. One customer with demand distribution of Erlang (9) and T = 20, L = 1.



f. One customer with demand distribution of Erlang (9) and T = 20, L = 3.

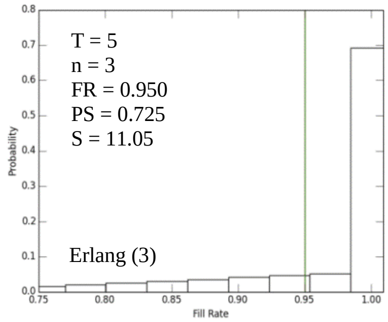


g. One customer with demand distribution of Erlang (9) and T = 100, L = 1.

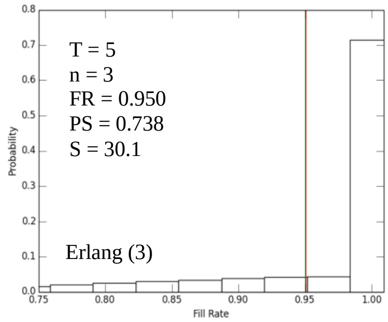


h. One customer with demand distribution of Erlang (9) and T = 100, L = 3.

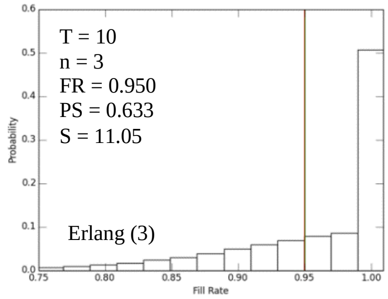
Figure 4.1: The distribution of FR for various lead times and review period durations when the long-run FR is fixed at 0.95 in all the plots. The demand distribution is Erlang (9). (T is review period, N is the number of customers, FR is the long-run review period, PS is the probability of success and S is the base stock level).



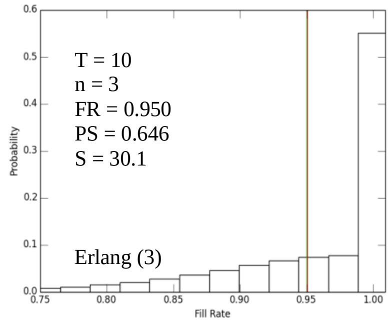
a. Three customers with a demand distribution of Erlang (3) and T = 5, L = 1



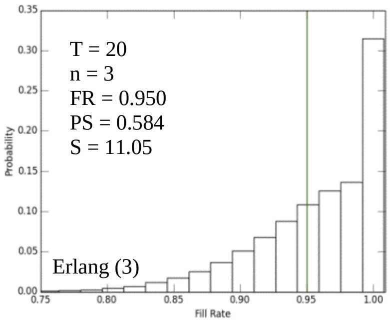
b. Three customers with a demand distribution of Erlang (3) and T = 5, L = 3.



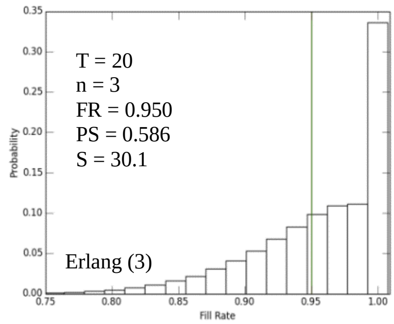
c. Three customers with demand distribution of Erlang (3) and T = 10, L = 1.



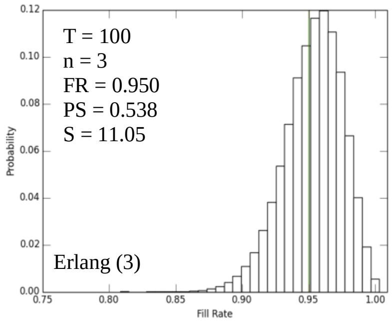
d. Three customers with demand distribution of Erlang (3) and T = 10, L = 3.



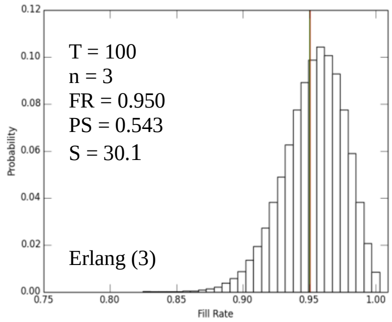
e. Three customers with demand distribution of Erlang (3) and T = 20, L = 1.



f. Three customers with demand distribution of Erlang (3) and T = 20, L = 3.



g. Three customers with demand distribution of Erlang (3) and T = 100, L = 1



h. Three customers with demand distribution of Erlang (3) and T = 100, L = 3.

Figure 4.2: The distribution of FR for various lead times and review period durations when the long-run FR is fixed at 0.95 in all the plots. The demand distribution is Erlang (3) for each customer and there are three independent customers.

Therefore, Q4.1 (*What is the impact on base stock levels for single and multiple customers' cases over various review lengths and various lead times?*) is answered. Increasing the lead time requires a higher base stock level, however the probability of exceeding the target FR also increases.

4.4.2 The impact of lead time on the average FR and PS

Here, we look at the impact of various base stock levels and lead times on the average FR and the probability of exceeding the target FR. Figure 4.3 shows the changes in the average FR for cases with different lead times and base stock levels. As expected, by increasing the base stock level the average FR increases. Additionally, for a given base stock level, increasing the lead time reduces the average FR for all of the review period durations. The slopes of the lines in Figure 4.3 are almost constant across all cases.

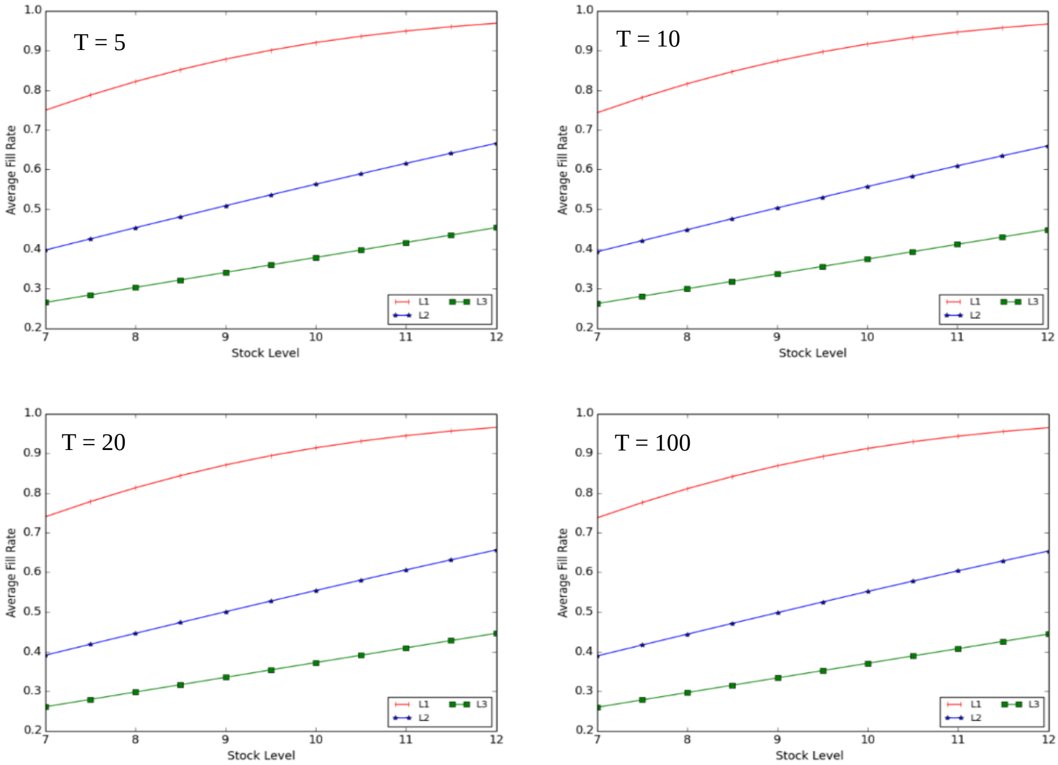


Figure 4.3: Changes in the average FR by altering the stock level and lead time. The demand distribution is $Erlang$ (9) in all cases.

Figures 4.4 and 4.5 show how the different lead times ($L = 1, 2, 3, \ldots, 10$) over different finite periods can affect the PS when the base stock level S is determined to achieve a 0.95 long-run FR. Figure 4.4 represents a case of one customer with an Erlang (9) demand distribution. Table 4.1 illustrates the details, including the base stock levels required in each case. Figure 4.5 and Table 4.2 present the results when there are three customers each with an Erlang (3) demand distribution. Figures 4.4 and 4.5 show that by increasing the lead time, the probability of exceeding the target FR increases, and that this increase is higher for shorter review period durations. It can be seen in Figures 4.4 and 4.5 that the line for T = 5 is both higher and steeper than the other lines for T= 10 and T = 20.

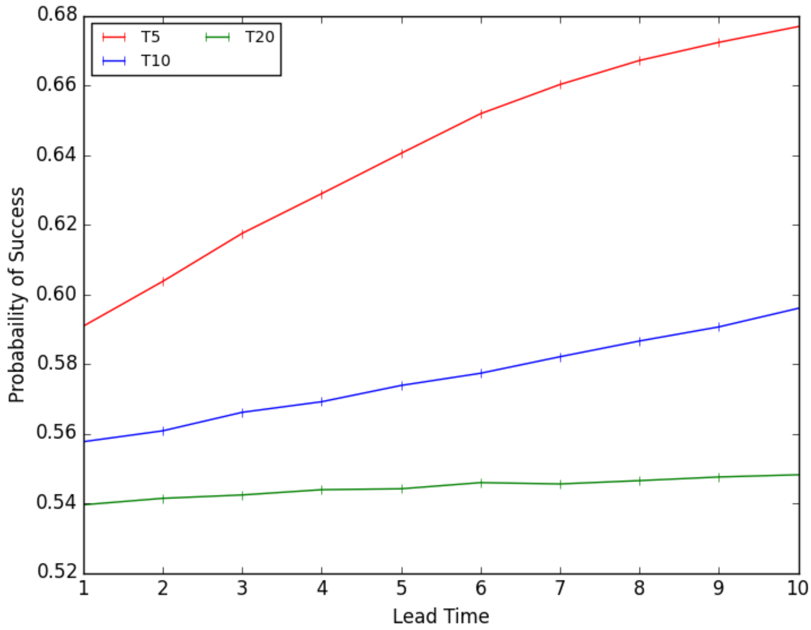


Figure 4.4: The PS when there is one customer with demand distribution Erlang(9) and the long-run FR is set to be 0.95 for different review horizon lengths and different lead times.

Table 4.1 provides greater detail for Figure 4.4. The mean of FR to be achieved is 0.95 for all plots.

Table 4.1: Detail of probability of success with one customer, different review periods and lead times Erlang (9)

Lead time (L)	Stock (S)	Variance FR, T=5	Probability (PS), T=5	Variance FR, T=10	Probability (PS), T =10	Variance FR, T=20	Probability (PS), T =20
1	11.05	0.0538	0.591	0.0390	0.557	0.0279	0.539
2	20.79	0.0596	0.605	0.0433	0.561	0.0313	0.541
3	30.10	0.0633	0.617	0.0466	0.564	0.0333	0.541
4	39.10	0.0665	0.628	0.0485	0.570	0.0352	0.544
5	48.05	0.0690	0.641	0.0502	0.574	0.0363	0.544
6	57.10	0.0716	0.652	0.0522	0.578	0.0376	0.546
7	65.70	0.0740	0.660	0.0534	0.581	0.0389	0.546
8	74.20	0.0752	0.667	0.0547	0.585	0.0394	0.548
9	82.95	0.0766	0.673	0.0561	0.591	0.0405	0.548
10	92.03	0.0775	0.676	0.0579	0.595	0.0412	0.548

Figure 4.5 continues the analysis, as the PS when there are three independent (and identical) customers, each with demand distribution Erlang (3). The long-run FR is set to be 0.95 for different review horizon lengths and different lead times.

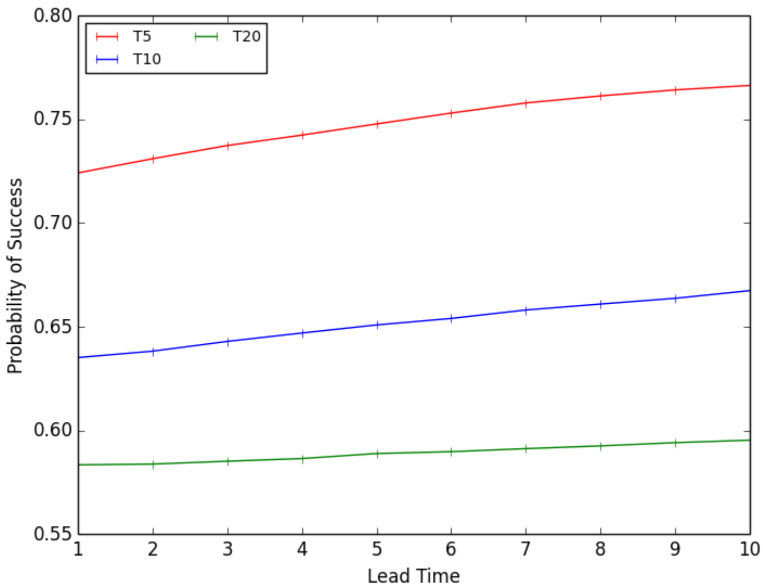


Figure 4.5: The PS when there are three independent (and identical) customers, each with demand distribution Erlang (3) and the long-run FR is set to be 0.95 for different review horizon lengths and different lead times.

Table 4.2 enumerates the data from Figure 4.5, again with FR target of 0.95

Table 4.2: Detail of probability of success with three customers, different review periods and lead times Erlang (9)

Lead time (L)	Stock (S)	Variance FR, T=5	Probability (PS), T=5	Variance FR, T=10	Probability (PS), T =10	Variance FR, T=20	Probability (PS), T =20
1	11.05	0.0932	0.725	0.0674	0.635	0.0484	0.583
2	20.79	0.0949	0.731	0.0693	0.638	0.0499	0.583
3	30.10	0.0964	0.737	0.0706	0.642	0.0510	0.585
4	39.10	0.0980	0.743	0.0718	0.647	0.0517	0.587
5	48.05	0.0993	0.747	0.0730	0.651	0.0525	0.588
6	57.10	0.1014	0.754	0.0738	0.654	0.0534	0.590
7	65.70	0.1030	0.758	0.0749	0.658	0.0539	0.591
8	74.20	0.1042	0.761	0.0759	0.661	0.0547	0.592
9	82.95	0.1050	0.764	0.0768	0.663	0.0550	0.594
10	92.03	0.1056	0.766	0.0768	0.667	0.0553	0.596

79

4.4.3 Summary of Lead Times, FR and PS

Figure 4.3 shows the analysis for the estimated FR: that increased stock levels result in average FR increases, although the review period duration shows an inverse relationship with the average FR. The lead times, however, do not have a significant effect on the average FR. Figure 4.4 and Table 4.1 show the results of the probability analysis, that the target of 0.95 FR would be exceeded. As expected, the probability is higher for shorter review period durations. Furthermore, extended lead times affect shorter review periods to a greater extent. For three independent customers (Figure 4.5, Table 4.2), the results for one customer were duplicated, but to a less pronounced extent. That is, the probability of exceeding the target FR was not as high. Hence it can be seen that review period durations have more effect than lead times on probability of success.

4.4.4 Achieving a defined probability of exceeding the target FR

Analysing the probability of exceeding the target FR for the base stock level is critical for a supplier, through association with bonuses or penalties. Similar to Abbasi et al. (2017), the following optimisation problem is numerically solved to find the required S for each case with different lead times, review periods and numbers of customers.

$$S = Inf\{S \; where \; \Pr(\alpha_t^{C_i}(S, \mathrm{L}, T) \geq \gamma \text{ and } S > 0\}$$

The PS is denoted by $P_\pi^{C_i}(S, \mathrm{L}, T)$, where T, L and π are parameters in this function, π being the target FR. The given probability of exceeding the target FR as a parameter in the model is denoted by γ. This model has been applied with different parameter values, lead times and numbers of customers. Figure 4.6 presents the results for various γ for different cases where the overall distribution of demand is *Erlang (9)*. In the case of multiple customers, each customer is assumed to be identical in terms of demand distribution. Figure 4.6 shows that for any lead time, when T is smaller than the threshold value (θ), having three customers is more efficient because the supplier can achieve the given PS with a smaller base stock level (S). Conversely, when T is larger than θ, having one large customer is more desirable. The results indicate that θ can be very large when the given PS is low (e.g., 0.6), and it decreases when the given PS is high (e.g., 0.9 or 0.95).

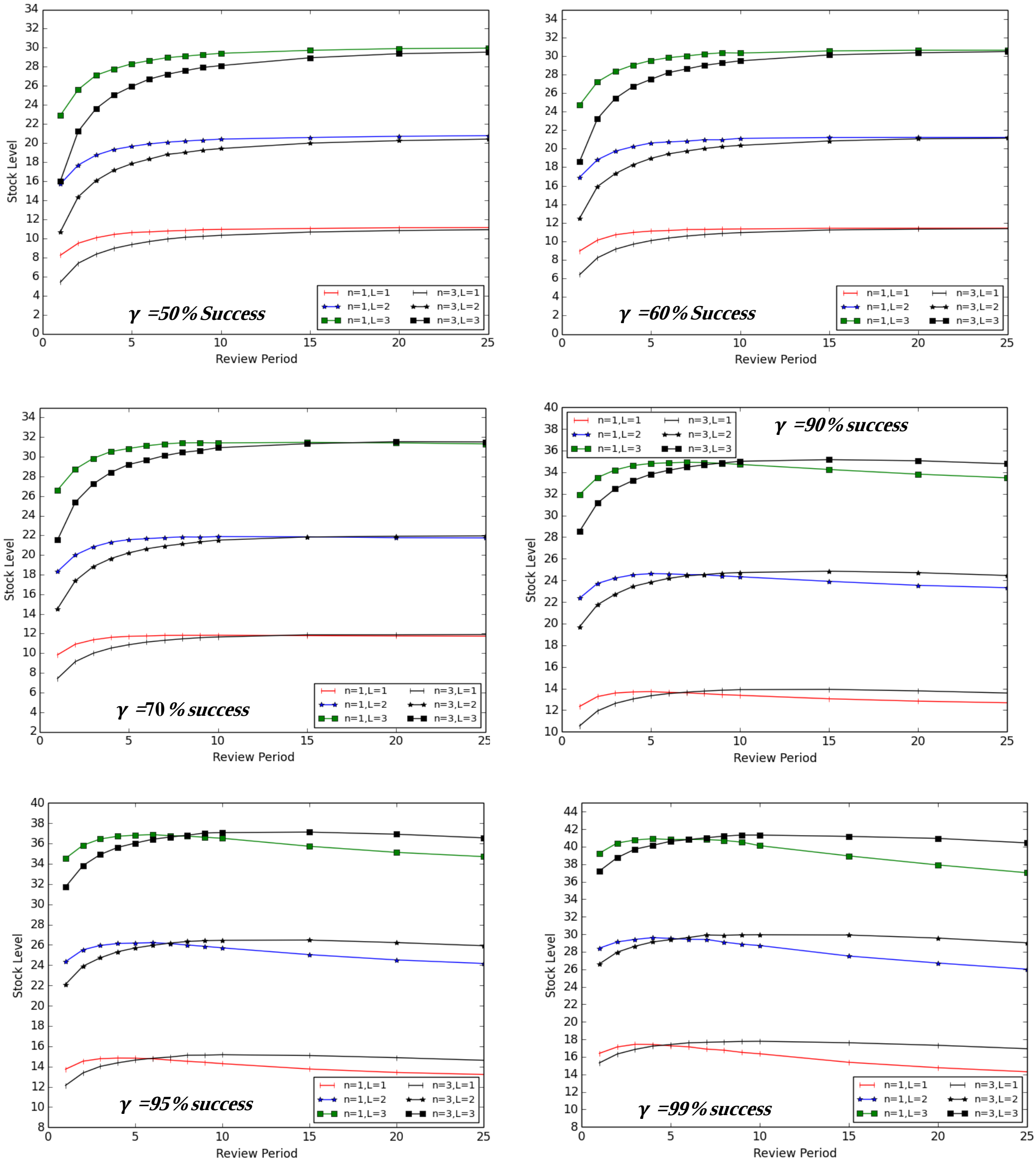


Figure 4.6: The required base stock to achieve a defined probability of success (γ) in various cases. The target FR is 0.95.

In the γ = 50% case and lead time, L = 1 and T = 5, when the supplier deals with one customer, 10.5 stock units are needed to have an even chance of meeting the FR target; and for three customers, 9.3 units are necessary. Similarly, when the lead time is L=3 for one customer, the supplier needs to stock more than 27.8 units to achieve FR probability of 50%, and for three customers, 25 units. Thus, as lead times grow, more stock is necessary in all scenarios for the supplier to achieve the target FR of 0.95.

4.4.5 Customers with correlated demands

The final question of this chapter is Q4.4: *How does lead time together with correlation of demand affect the probability of overreaching the target FR for each of multiple customers?*

Here, we investigate cases where customers' demands are correlated and lead times are non-zero. We consider two situations: first, when under various correlation settings the marginal demand distribution of each customer remines unchanged; second, when for various correlations, the marginal demand distribution of each customer change but the total variance is fixed in all cases.

4.4.5.1 Fixed marginal demand distribution (in both correlated and independent demand cases)

Here, we study the impact of positive and negative correlation of demands involving various fixed lead times over a finite review horizon for both one customer and a set of multiple customers. This study assesses the effect of positive and negative correlation of demand on the base stock level needed to achieve a given probability of exceeding the target FR of 0.95. We examine the case of having one customer with an Erlang (6) demand distribution, and the case of having two customers, each with an Erlang (3) demand distribution, with various correlation coefficients and different lead times and review period lengths.

Figure 4.7 shows that for any fixed lead time, in the case of negative correlation, it is preferable for the supplier to deal with multiple customers with smaller demands than one customer. The first column shows positive correlation cases and the second column, negative correlation. By positively increasing the correlation coefficient, ρ, the threshold θ decreases (θ is the review period 'cross-over point' after which the target FR can be achieved with a lower base stock level in the one-customer case). Therefore, in the case of high positive correlation, for any lead time, having one large customer is always preferable to multiple smaller customers. The results show that for different lead times and target success rates, negative correlation in demand requires a lower base stock level. Thus, a lower stock level is more efficient for the supplier.

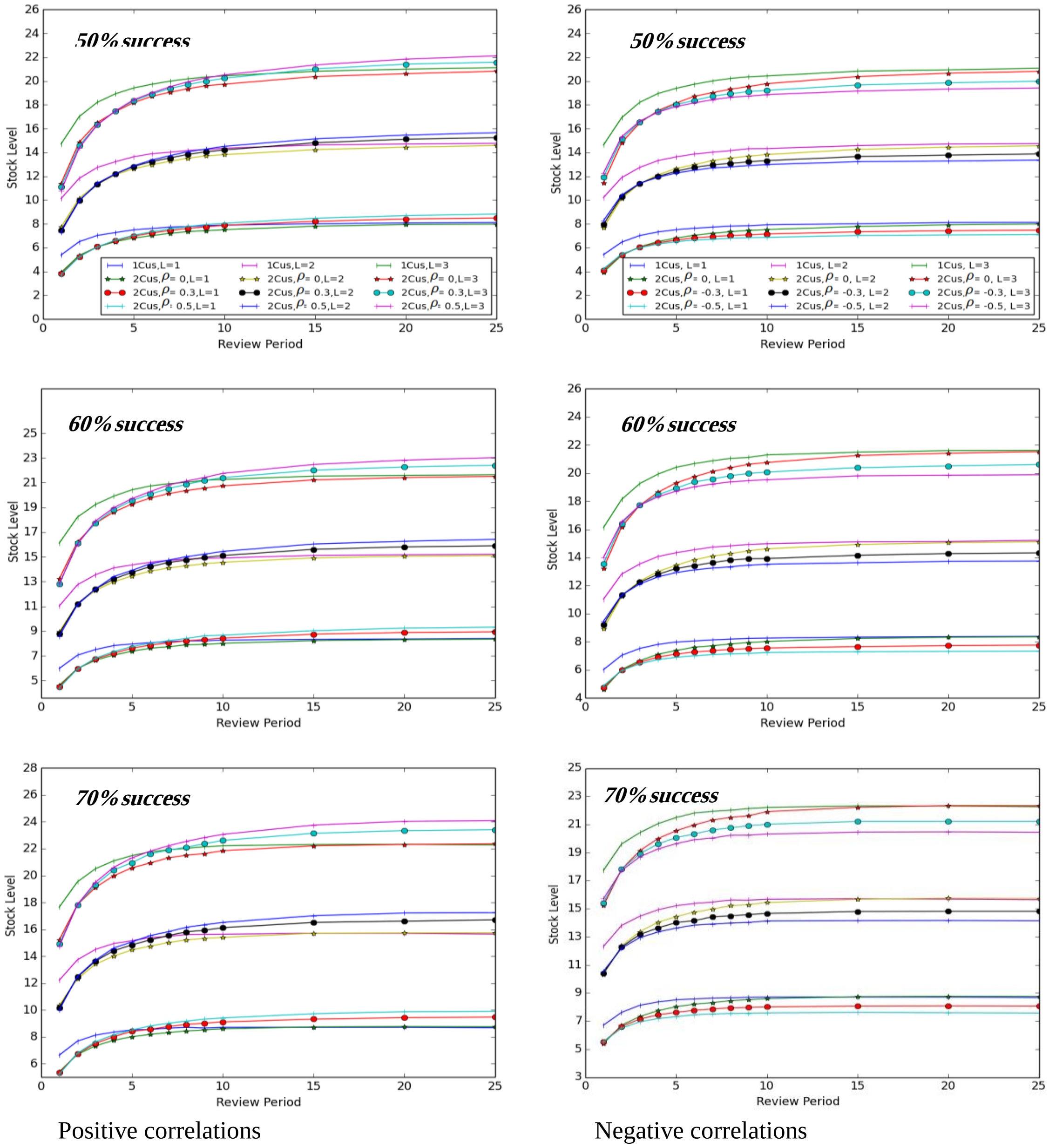
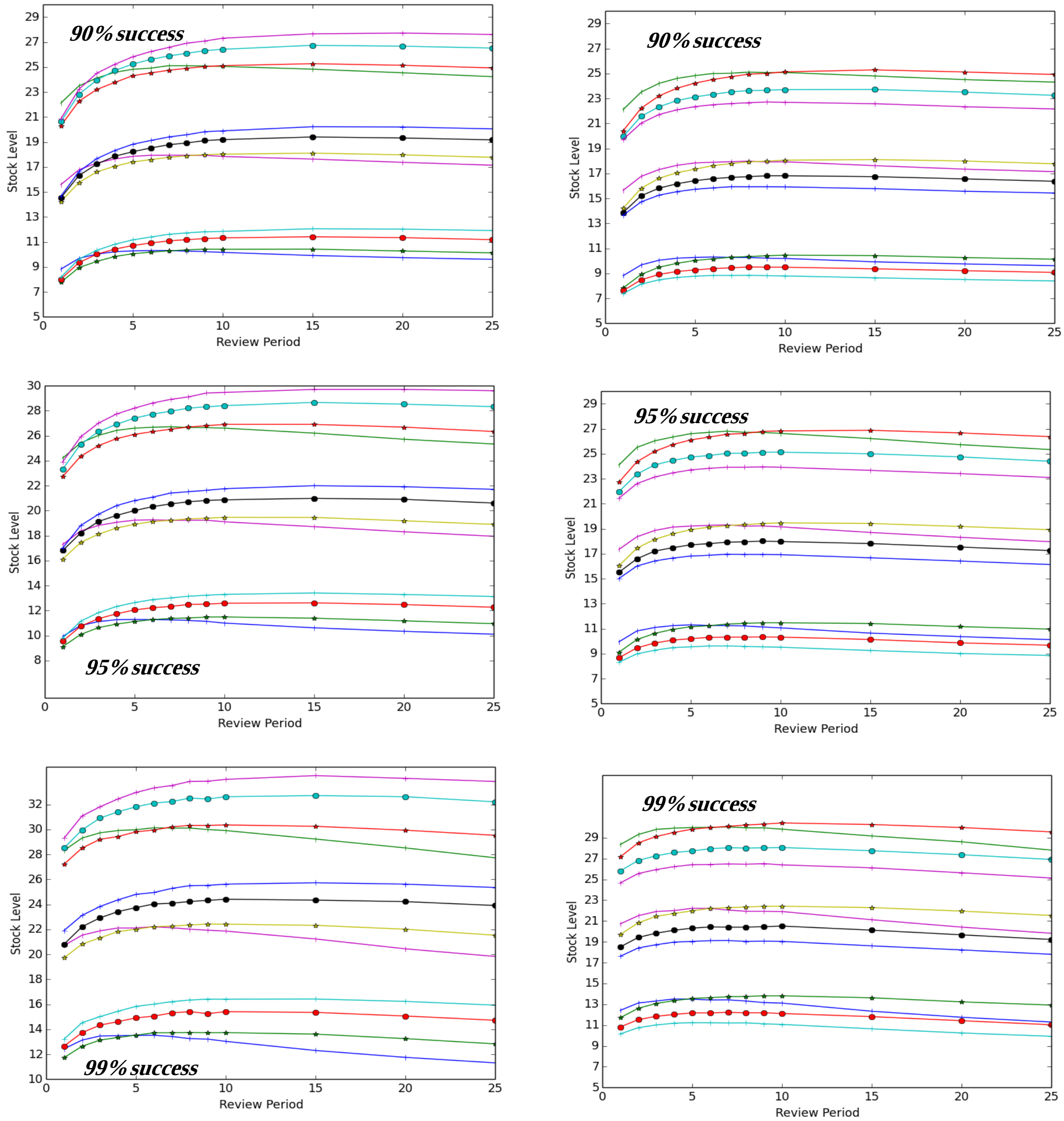


Figure 4.7: Comparison between positive and negative correlation demand, involving different lead times for one customer and two customers (with demand of Erlang(6)), two independent customers (each with demand of Erlang(3)), in terms of the stock levels needed to meet a 95% target FR with various probabilities of success (50%, 60%, 70%, 90%, 95% and 99%). The demand distribution of each customer is Erlang (3) in both correlated and independent demands cases.



Positive correlations Negative correlations

Figure 4.7 (cont.): Correlation of customer demand, fixed marginal demand.

4.4.5.2 Constant total variance

Next, we investigate the effects of various lead times over different review horizons for one customer with an Erlang (6) demand, two independent customers (each with an Erlang (3) demand), and two customers with correlated demands (each with $\Gamma(3(1+\rho), 1/(1+\rho))$). We consider the base stock level required to meet a 95% target FR with several probabilities of success. The total variance of demand is fixed in all scenarios. Figure 4.8 shows that the lead time and the correlation of demand have an effect on the performance of the SC. From these results, we can remark that by increasing the lead time, the supplier needs to stock more units to achieve customer satisfaction.

Figure 4.8 also demonstrates that in cases of constant variance, having multiple customers with either negative or positive correlation favours the supplier for short review periods, but there is a threshold for the review period beyond which the performance in the correlated cases become worse than the one customer case.

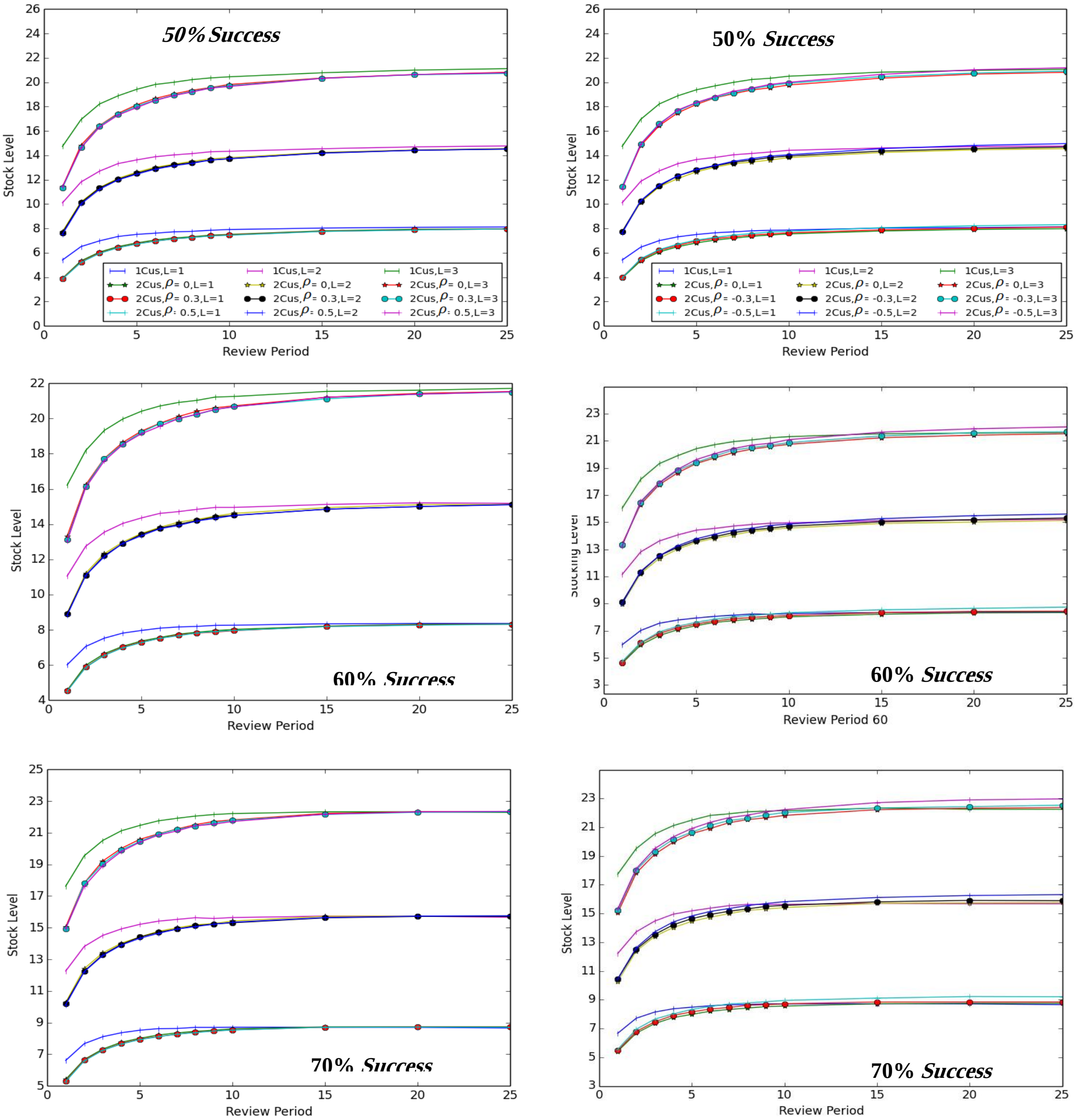


Figure 4.8: Comparison between positive and negative correlation demand, involving different lead times for one and two customers (with demand of Erlang(6)), two independent customers (each with demand of Erlang(3)) and two customers with correlated demands (each with Gamma(3(1+ ρ),1/(1+ ρ))), in terms of the stock levels needed to meet a 95% target FR. The figures in the left column are for positive correlation cases, and those in the right column are for negative correlation cases.

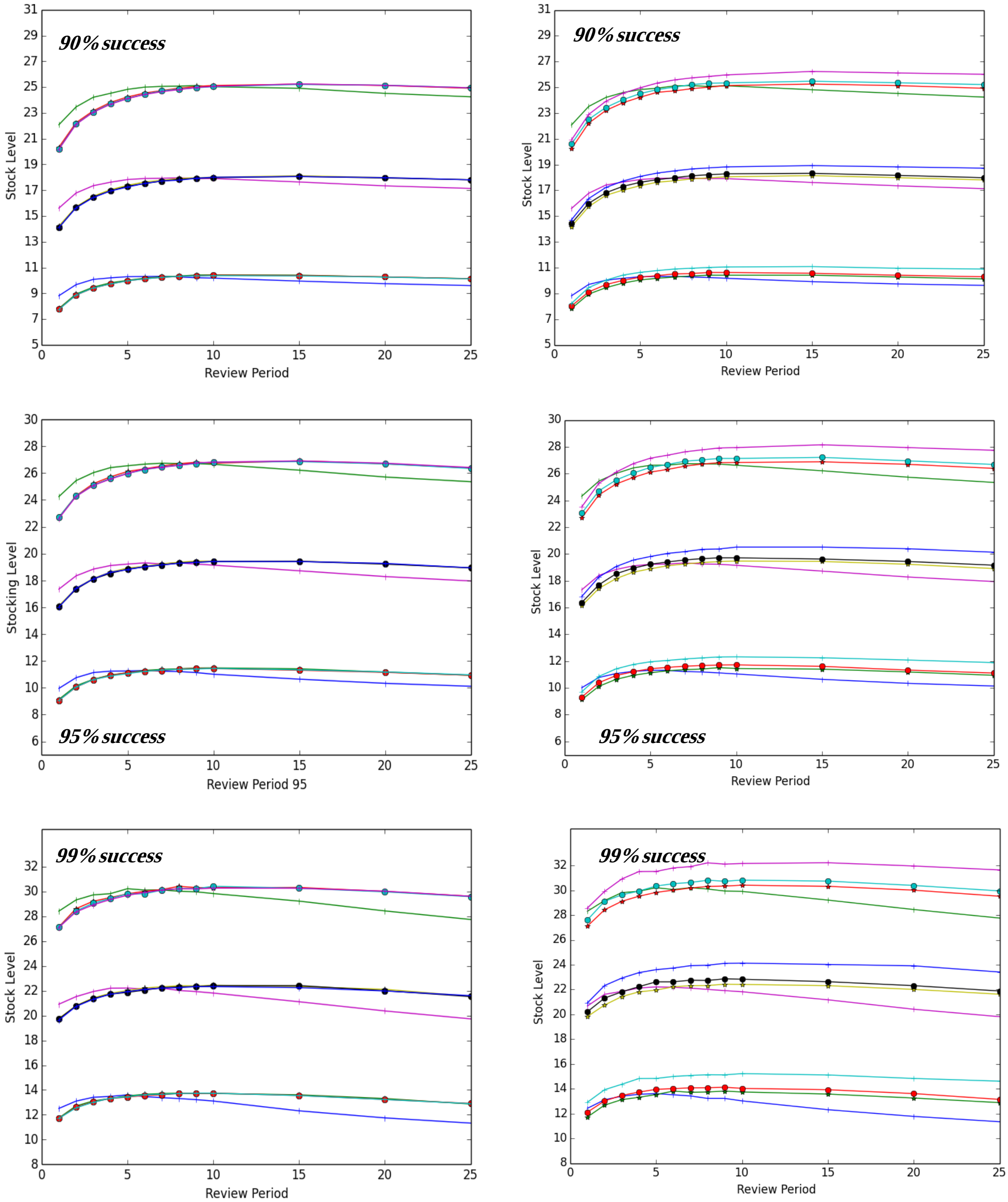


Figure 4.8 (cont.): Correlation of customer demand, constant variance of demand.

4.4.5.3 Service level received by non-identical customers dealing with a supplier

Here, we examine the case of a supplier with multiple customers with lead times of $L = 1$ and $L = 3$, and with the supplier's base stock level set to achieve an average FR for overall demand across all customers. It is supposed that the supplier has two customers and an overall Erlang (6) demand distribution, where customer 1 has an Erlang (2) demand and customer 2 has an Erlang (4) demand.

Figure 4.9 presents the expected FR value for two cases involving different lead times for each customer for different base stock levels (S) and review periods (T). This Figure shows that as long as the stock level increases for various lead times, the customer with higher demand will have a higher expected FR value. Figure 4.10 verifies that the customer with the larger demand has a greater chance of exceeding the target FR than the customer with the smaller demand for different lead times.

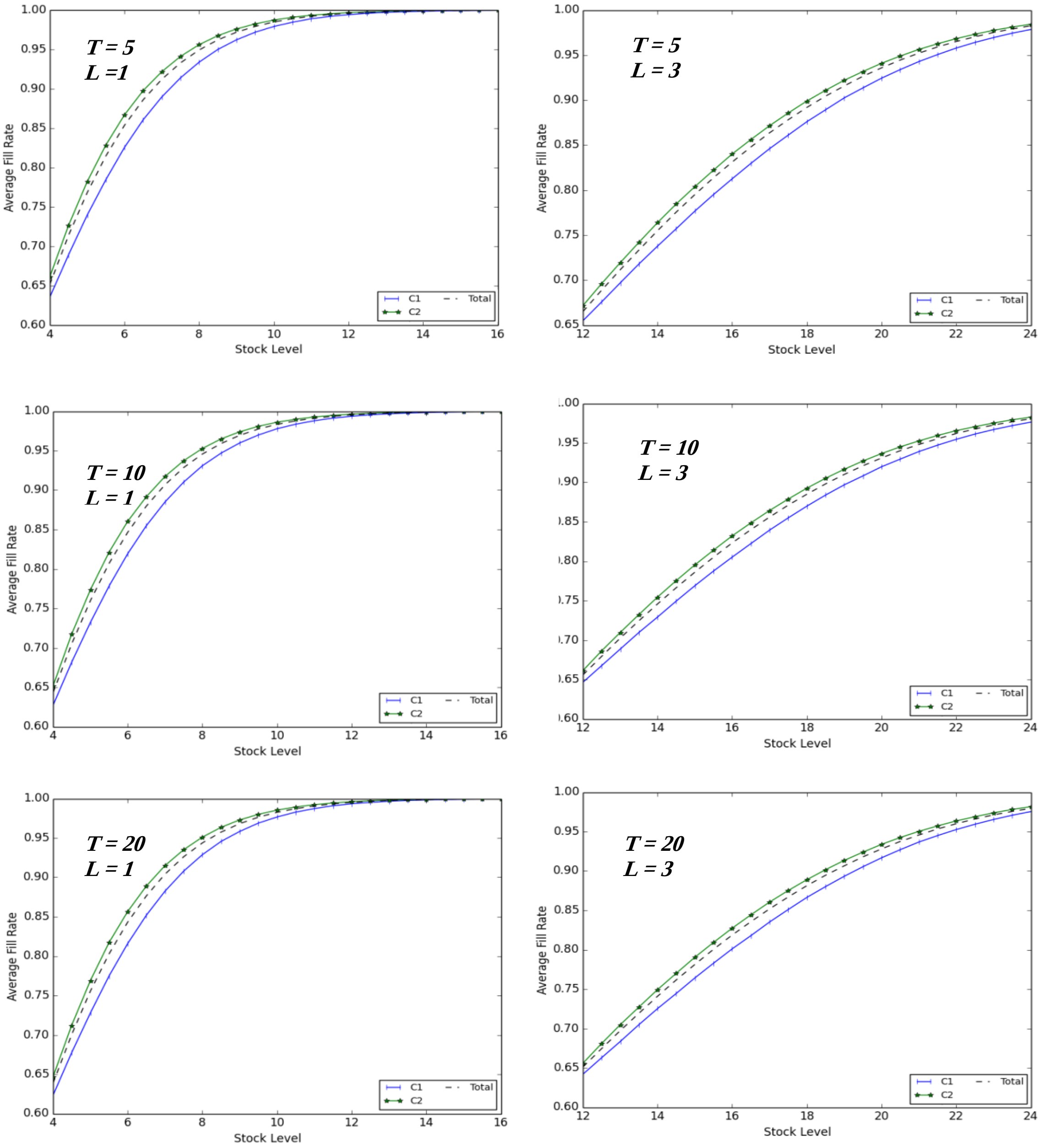


Figure 4.9: The average FR for customers 1 and 2 with different lead times (L = 1, and 3). The demand distribution for customers 1 and 2 are $Erlang\ (2)$ and $Erlang\ (4)$, respectively. The base stock level (x-axis) changes from 2 to 12 when L = 1, and from 10 to 24 when L = 3; three review periods (T = 5, 10 and 20) are considered. The dashed line shows the average FR for the system with one customer with an $Erlang\ (6)$ demand distribution.

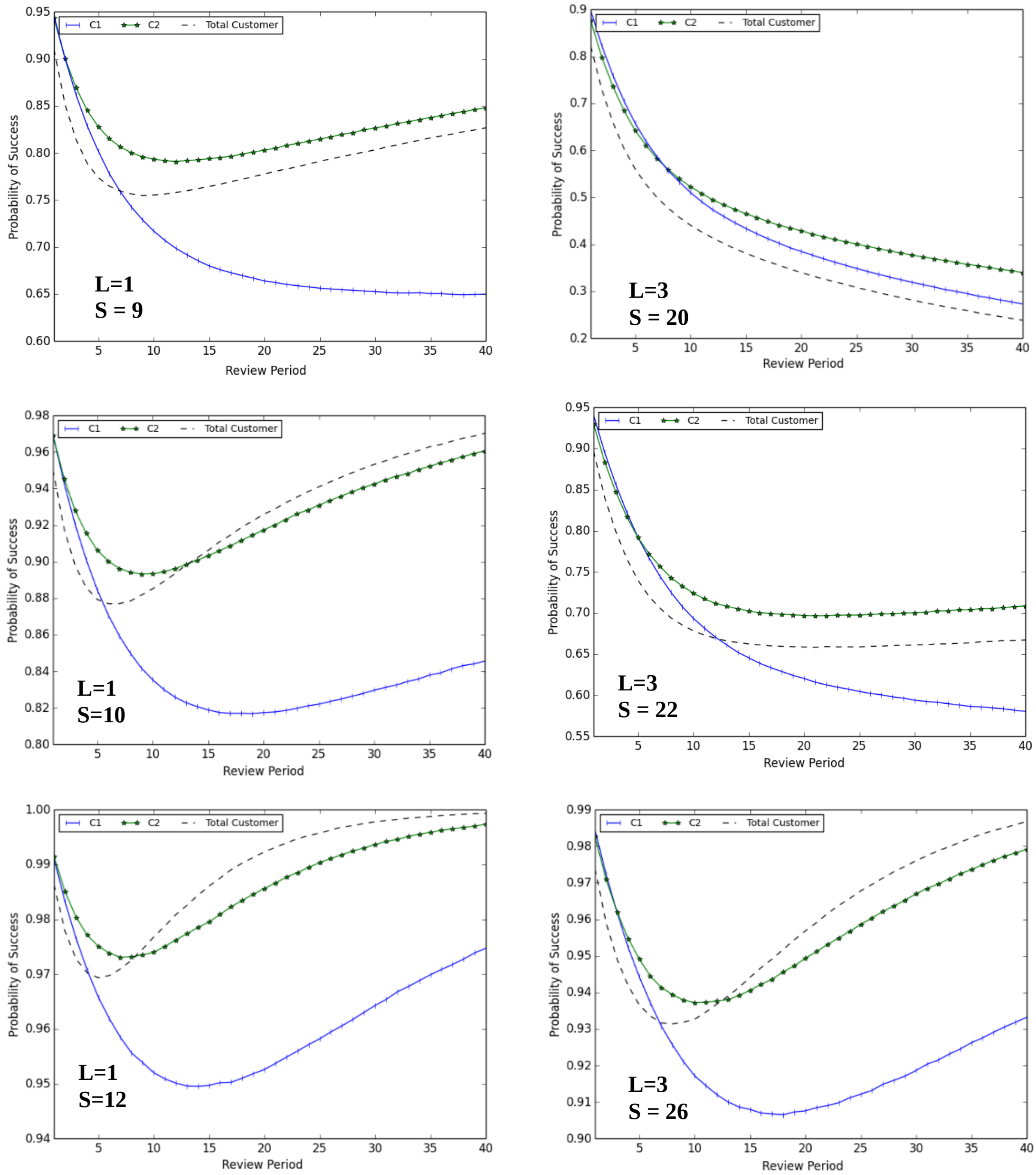


Figure 4.10: The probability of exceeding the target FR of 0.95 for customers 1 and 2. The demand distribution of customers 1 and 2 and are $Erlang\ (2)$ and $Erlang\ (4)$, respectively. The review period T (x-axis) changes from 1 to 40, and three base stock levels (S = 9, 10 and 12) are considered for L = 1, and when L = 3, base stock levels S = 20, 22 and 26 are considered. The dashed line shows the average FR for the system with one customer with an $Erlang\ (6)$ demand distribution.

Question 4.4: *How does lead time affect correlation of demand and the probability of overreaching the target FR of each of multiple customers?*

This was answered in several parts. First, with negativity correlated demands, given any fixed lead time, the supplier should select multiple customers with smaller demands, rather than one customer. With high positive correlation for any lead time, one large customer is preferable. Next, with constant variance, multiple correlated customers proved greater productivity. However, once replenishment periods reached a threshold, one customer was preferable. The following analysis showed that as stock increases with lead times, the larger customer proved a higher expected FR value for the supplier. This was subsequently verified.

4.5 Chapter Summary

Four categories of variables with multiple scenarios were assessed in this chapter. Given one large customer and multiple small customers, the finding was that as lead times increase, the supplier needs more stock units to reach a defined service level. When the expected FR is fixed, longer lead times produce a higher probability of exceeding the target FR. For any lead time, it is more productive for a supplier to have multiple independent smaller customers than a single larger customer if the review period is less than a given threshold (θ). The opposite holds true when the review period is greater than θ.

When customer demands are correlated with variance, positive increase in the correlation coefficient results in a decrease of the threshold θ. Therefore, a scenario of high When customer demands are correlated, positive increase in the correlation coefficient results in a decrease of the threshold θ. Therefore, in a scenario of high positive correlation for any lead time, having one large customer for a supplier is always preferable to multiple smaller customers. Moreover, the results showed that negative correlation in demand requires a lower base stock and is profitable for the supplier. Analysing the impact of correlation with constant variance showed that having multiple customers with either negative or positive correlation is better for the supplier for short review periods. Conversely, one larger customer is preferable beyond a threshold review period duration. It was also observed that with multiple customers, a larger customer receives a better service level and has a higher probability of exceeding the target FR than smaller customers, regardless of the lead time.

Chapter 5: FR: Measurement of Ensuring Reliability in a Supply Chain

5.1 Introduction

This chapter considers the consistency and asymptotic normality of the fill rate. It extends the range of this measure to determine an optimal stock level, and evaluates the reliability of different SCs. Up to this point the analysis has not taken into consideration measurement of FR variability, as the scale was taken as one. This chapter focuses on variability; consistency and asymptotic normality in FR for optimal stock levels and in evaluating the reliability of SCs. Scenarios include single and multiple customers. This chapter first explains renewal reward theory as a measure of SC reliability. This is followed by modelling a single customer in the FR, which is then generalised to multiple independent customers. Finally, numerical examples validate the equations for the FR measure, to complete the objective.

5.2 Introduction to Renewal Reward Processes

The renewal reward model is one of the most powerful tools to evaluate applied probability models such as inventory management, queueing and reliability applications, and others. Many stochastic processes are regenerative, which means they renew themselves from time to time. Thus, the conduct of the process after the regeneration time is a probabilistic copy of conduct of the process beginning from time zero. The time interval between two regeneration periods is called a cycle. Whereas, the series of cycles creates a so-called renewal process (Tijms, H.C., 2003).

Renewal theory is a generalization of a Poisson process, where time intervals between successive events are independent and identically distributed exponential random variables. A generalisation of a Poisson process can be established, where times between successive events are independent and identically and arbitrarily distributed. Mid-century Cox (1983, cited in Settanni et al. 2016, p. 13) ‘modelled recurrence data as a Poisson process’. Renewal theory can be either discrete time or continuous. Kim (2016, p.1) explained that renewals are randomly occurring events that can occur as bulk-renewal or single occurrences. Asymptotic results from renewal processes are used to study long-term events, and recent advances in data management have renewed researcher interest in these ‘lengthy and complex expressions’. In 2016, ‘connection between the asymptotic results in continuous and discrete-time bulk-renewal processes had not been determined’ (Kim, 2016, p. 1). The formal definition for this study is:

Let X_1, X_2, X_3, $X_4, X_5, \ldots$ be a positive sequence independent identically distributed random variables such that:

$$0 < E[X_i] < \infty$$

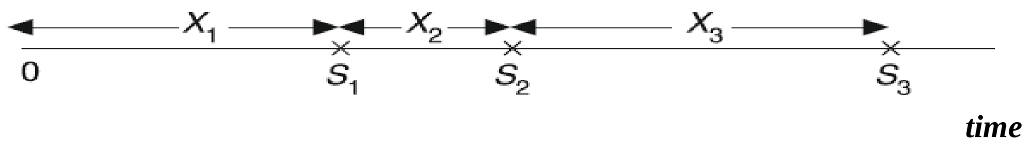


Figure 5.1: Renewal reward process and interval times

Define for each $n > 0$:

$$S_0 = 0$$

$$S_n = \sum_{1}^{n} X_i$$

Where, $n \geq 1$

It is, $S_1 = X_1$ which is the time of the first renewal; $S_2 = X_1 + X_2$ means the time till the renewal plus the time between the first and the second renewal, S_2 is the second time of renewal. In overall, S_n denotes the times of nth renewal (see Figure 5.1) (Nebres, 2011).

$$[S_n, S_{n+1}]$$

Then (N_t), which the random variable is given by:

$$N(t) = \sum_{n=1}^{\infty} 1_{\{S_i \leq t\}} = sup\{n : S_n \leq t\}$$

5.3 Modelling FR with a Single Customer

5.3.1 Model, assumptions and notation through Chapter 5

The following notations are used in this chapter, in the case of a single customer:

T : duration of the review period the FR is calculated for;

S : base stock level;

X_i : number of units demanded by a customer during period i;

$f_{(x)}(X)$: probability density function (pdf) of X_i.;

Y_i: number of units of fulfilled demand during period i, i.e., Y_i = min (S, X_i);

$\alpha_{(T)}(S)$: the FR for review period of length T;

$P_{(T)}(S, \pi)$: the probability that the FR exceeds threshold π, i.e., $P_{(T)}(S, \pi)$ = Pr($(\alpha_{(T)}(S) \geq \pi)$.

In this model, during each equally spaced time interval i = 1, 2, …, a single customer demands items from an inventory of similar products. The inventory system is an order-up-to inventory type, with a base stock level S, which is replenished instantly, and the lead time is equal to zero. All unmet demand is lost, so no backlogging occurs.

The model also assumes that demands X_i are independent and identically distributed each with pdf $f_{(x)}(X)$. Due to this assumption, the system is stationary and events occurring during a review period of length T are stochastically identical to events in other review periods of the same length. The FR is defined as:

$$\alpha_{(T)}(S) = \frac{Y_1 + Y_2 + \cdots + Y_T}{X_1 + X_2 + \cdots + X_T} \tag{5.1}$$

This reflects the proportion of demands that are filled from the base stock during a review period. Throughout this chapter we will make the following assumptions:

- The variance of X, i.e., $Var\ (X) = E(X^2) - E^2(X) < \infty$.
- $\Pr(X > 0) > 0$; this will imply by the Glivenko– Cantelli Lemma that:

$$\Pr(X_i > 0 \text{ infinitely often}) = 1$$

Therefore, the denominator of (5.1) is never equal to zero with probability 1, as the renewal process $S_n = \sum_1^n X_i$ is non–terminating.

5.3.2 Folk theorem

The main result of (5.2), and specifically (5.3) below, is well known, so we refer to it as a Folk Theorem. Nevertheless, based on our knowledge, it does not appear to have been proven, so we do so here.

Theorem 1 With probability 1,

$$\alpha_{(T)}(S) \rightarrow \frac{E\ (Y)}{E\ (X)} \tag{5.2}$$

and, in addition

$$E(\alpha_{(T)}(S)) \rightarrow \frac{E\ (Y)}{E\ (X)} \tag{5.3}$$

as $T \to \infty$ Here:

$$E(Y) = \int_0^{\infty} \min(S, x) f_x(x)\, dx$$

$$= \int_0^{S} x f_x(x)\, dx + S \Pr(X \geq S).$$

Proof:

In the proof, all convergence will be expected to occur as $T \to \infty$ and with probability of 1. Define $N(t) = sup\{n : Sn \leq t\}$ where the renewal process:

$$S_n = \sum_1^n X_i$$

We first note that:

$$\alpha_{N(T)}(S) = \frac{Y_1 + Y_2 + \cdots + Y_{N(T)}}{T} / \frac{S_T}{T} \tag{5.4}$$

where $S(t) = S_{N(t)}$. Using the *Renewal Reward Theorem 3.16* in Ross (1970):

$$\frac{Y_1 + Y_2 + \cdots + Y_{N(T)}}{T} \to \frac{E(Y)}{E(X)} \tag{5.5}$$

By the Strong Law of Large Numbers $\frac{S(T)}{N(T)} \to E(X)$, and by the Renewal Theorem:

$\frac{N(T)}{T} \to \frac{1}{E(X)}$; hence

$$\frac{S(T)}{T} = \frac{S(T)}{N(T)} / \frac{T}{N(T)} \to 1. \tag{5.6}$$

The result of equation (5.2) follows from (5.4), (5.5), (5.6) and observes that as the renewal process is non–terminating, $lim_T\, \alpha_{(T)}\, S = lim_T\, \alpha_{N(T)}(S)$. Finally, equation (5.3) can be proven that in exactly the same way using Theorem 3.16 in Ross (1970). A stronger result than (5.33) was shown to be valid by Banerjee (2005) and Chen (2003).

They specifically proved that expected FR is actually non-decreasing in T while converging to $\frac{E\,(Y)}{E\,(X)}$.

5.3.3 Consistency and asymptotic normality of the FR

As far as we know, under the FR SLA performance, no previous works or research has discussed the consistency and asymptotic normality of the FR. However, Thomas (2005) and Abbasi et al. (2017) discussed the FR SLA measure and presented their works using a Monte Carlo simulation. So, the consistency of the FR and its asymptotic normality as T increases over different review period lengths are clearly obvious in graphical displays. By considering the first-order approximation (i.e., ignoring terms with order greater than one), we can display the dependence of the variance on T specifically and prove asymptotic normality. However, first we collect some results concerning the variance of Y and covariance between X and Y:

$$Var(Y\,) \;=\int_0^S \mathrm{x}^2\, f_X\,(x)\,dx + S^2\,\Pr(\,X\;\geq S) - \;E^2(Y) \tag{5.7}$$

$$cov(X,Y\,) \;=\; \int_0^S \mathrm{x}^2\, f_X\,(x)\,dx + S\,\int_S^\infty x\,f_X\,(x)\,dx - E(X)E(Y) \tag{5.8}$$

We note the formula for $E(Y\,)$ from the previous:

$$\text{Here, } E(Y\,) \;=\; \int_0^\infty \min\,(S,X_i)\,f_x\,(x)\;dx$$

$$=\int_0^S \mathrm{x}\,f_x\,(x)\;dx + S\,\Pr(X\;\geq S)$$

Theorem 2 For the first-order approximation, the mean and variance of:

$\alpha_{\mathrm{T}}(S)$ are:

$$E(\alpha_T(S)) = \frac{E\,(Y)}{E\,(X)} \tag{5.9}$$

$$Var\,(\alpha_{(T)}(S)) = \frac{E^2(Y)}{T\,E^2(X)}\,\left(\frac{Var\,(Y)}{E^2(X)} - \frac{2\,cov(X,Y)}{E(X)E(Y)} + \frac{Var\,(X)}{E^2(X)}\right. \tag{5.10}$$

respectively.

Proof

The first-order approximation to the mean and variance of the ratio random variable $\frac{Z}{W}$ where Z and W are correlated are given by Kendall and Stuart (1969):

$$E\left(\frac{Z}{W}\right) = \frac{E(Z)}{E(W)} \tag{5.11}$$

$$Var\left(\frac{Z}{W}\right) = \frac{E^2(Z)}{E^2(W)}\left(\frac{Var\,(Z)}{E^2(Z)} - \frac{2\,cov(Z,W)}{E(Z)E(W)} + \frac{Var\,(W)}{E^2(W)}\right) \tag{5.12}$$

respectively. Applying (5.11) and (5.12) to:

$Z = Y_1 + Y_2 + \cdots + Y_T$, and W = $X_1 + X_2 + \cdots + X_T$ proves the results.

We note from equation (5.10) and from using one of the assumptions under sub-section (5.3.1), that:

$$Var\left(\alpha_T(S)\right) \to 0 \text{ as } T \to \infty;$$

Hence, the following corollary is an immediate consequence of applying Chebyshev Inequality.

Corollary 1

The fill rate converges in probability (i.e., it is a consistent estimator, of the ratio $\frac{E(Y)}{E(Z)}$) as $T \to \infty$.

Theorem 3

In the next result $\xrightarrow{d}$ represents convergence in distribution and $N(.,.)$ is the univariate normal distribution.

The following asymptotic normal convergence property of the FR SLAs holds:

$$\sqrt{T}\,(\alpha_{(T)}(S) - \frac{E(Y)}{E(X)}) \xrightarrow{d} \mathrm{N}\,(0, \sigma^2) \tag{5.13}$$

Where $\sigma^2 = T\,Var\left(\alpha_{(T)}(S)\right)$ $and\ Var\left(\alpha_{(T)}(S)\right)$ is the first-order approximation of the variance given by (5.10).

Proof

Applying the multivariate Central Limit Theorem (c.f. Theorem 5.4.4 in Lehman & Weibel [1999]), we obtain:

$$\sqrt{T}\ (\bar{X} - E\,(X), \bar{Y} - E\,(Y)) \xrightarrow{d} N(0, \Sigma\ ) \tag{5.14}$$

Where

$$\bar{X} = \frac{1}{T}\sum_{i=1}^{T} X_i\,, \bar{Y} = \frac{1}{T}\sum_{i=1}^{T} Y_i$$

and Σ is the two-by-two variance–covariance matrix of the bivariate random variable $(X, Y\ )$. That is, with diagonal elements:

$$a_{11} = V\,ar(X),\ a_{22} = V\,ar(Y), \text{ and } a_{12} = a_{21} = cov\,(X, Y).$$

We also note that $\alpha_{\mathrm{T}}(S) = \frac{\bar{Y}}{\bar{X}}$.

By applying the Delta Method (c.f. Theorem 5.4.6. in Lehman [1999]) to the function:

$f(x, y) = \frac{y}{x}$, we further obtain:

$$\sqrt{T}\,(\alpha_{(T)}(S) - \frac{E(Y)}{E(X)}\,) \xrightarrow{d} N\,(0, \sigma^2\,) \tag{5.15}$$

Where:

$$\sigma^2 = a_{11}\left(\frac{\partial f}{\partial x}\right)^2 + 2\,a_{12}\,\frac{\partial f}{\partial x}\,\frac{\partial f}{\partial y} + a_{22}\left(\frac{\partial f}{\partial y}\right)^2$$

$$= Var\ (X) \left(\frac{\partial f}{\partial x}\right)^2 + 2\ cov\ (X, Y)\ \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} + Var\ (Y) \left(\frac{\partial f}{\partial y}\right)^2$$

and all partial derivatives are evaluated at the point $\eta = (\ E(X), E(Y))$. Since:

$$\frac{\partial f}{\partial x} = -\ \frac{y}{x^2} \text{ and } \frac{\partial f}{\partial x} = \ \frac{1}{x}$$

$$\sigma^2 = var(x) \frac{E^2(Y)}{E^4(X)} - 2\ cov\ (X.Y) \frac{E(Y)}{E^3(X)} + \frac{Var\ (Y)}{E^2(X)}$$

$$= \frac{E^2(Y)}{E^2(X)} \left(\frac{Var\ (X)}{E^2(X)} - \frac{2\ cov(X,Y)}{E(Y)E(X)} + \frac{Var\ (Y)}{E^2(X)} \right)$$

$$= T\ Var\ \left(\alpha_T(S)\right) \tag{5.16}$$

On referring to equation (5.10), this completes the proof of the theorem.

5.3.4 Reliability of the Fill Rate

In terms of meeting customers' demands, the higher the FR, the more reliable the performance of the SC. As was displayed by Banerjee (2005) and Chen (2003), the average FR decreases when the performance review period is increased in a base stock inventory system, which means average FR is a decreasing function of the number of replenishment periods in T. Nevertheless, this does not indicate that the reliability in performance as measured by the FR is decreasing with T, since the real distribution of the FR can change quite frequently within each period as it is dependent on the type of distribution of the demand. This was presented in several simulation study examples by Abbasi et al. (2017) and Thomas (2005).

The perfect measure of reliability is that the FR exceeds a target π with a probability of at least γ where $0 < \gamma < 1$, that is:

$$P_T\ (S, \pi) \ \geq\ \gamma. \tag{5.17}$$

From Theorem 3, for large T, applying the asymptotic property of the FR, the probability$P_T\ (S, \pi)$ can be approximated by $1 - \Phi_Z\left(\frac{\sqrt{T}}{\sigma}(\pi - \frac{E(Y)}{E(X)}\right)$ where $\Phi_Z(.)$ is the cumulative distribution of the standard normal random variable. Applying the same large sample approximation, the target π that fulfils the reliability criterion (5.17) satisfies the following inequality:

$$\pi \leq \frac{E(Y)}{E(X)} + \frac{\sigma}{\sqrt{T}} \Phi_Z{}^{-1}(1 - \gamma) \qquad (5.18)$$

where $\Phi_Z{}^{-1}(.)$ is the inverse of $\Phi_Z(.)$.

In previous works, the base stock level *S* has been considered fixed. However, it is a key determinant in realising a reliable supply and satisfying customers' demands. As we have realised, base stock level *S*, demand, and the length of a review period T all effect the distribution of the FR SLA measure. In practice, customers and supply managers would discuss suitable base stock levels and those who fail to reach agreed service levels, and penalties are imposed on suppliers who fail to reach these levels. A sensible objective for both is to discover the minimum stock level *S > 0*, which will satisfy the condition in Equation (5.19). More specifically, the SC manager would prefer to solve the following optimisation problem by:

$$min\ S$$

$$\text{Such that} \qquad P_T\ (S, \pi) \geq \gamma, S > 0. \qquad (5.19)$$

First, the following Lemma should be presented:

Lemma 1 For each fixed T, the fill rate $\alpha_{(T)}\ (S)$ is a stochastically increasing function of S, i.e., $Pr(\alpha_T\ (S_1) \leq x) \leq Pr(\alpha_T\ (S_2) \leq x)$ if $S_1 > S_2$ for every $x > 0$.

Proof.

If $S_1 > S_2$ then the following strict inequality must hold for any random demand X:

$$min\{S_1, X\} > min\{S_2, X\}.$$

Hence $\alpha_T\ (S_1) > \ \alpha_T\ (S_2)$ with probability one which implies that:

$Pr(\alpha_T\ (S_1)\ \leq\ x) \leq Pr(\alpha_T\ (S_2)\ \leq\ x)$ for every $x\ >\ 0$.

Theorem 4

The solution S^* to the problem (5.19) is unique and is given by:

$$S^* = \inf\ \{ S :\ P_T\ (S, \pi) = \gamma \} \tag{5.20}$$

Proof.

From ***Lemma 1***, $\alpha_T\ (S)$ is a stochastically increasing function of S, hence for fixed target π, $P_T\ (S, \pi)$ is non-decreasing function of S. Therefore, over the region:

$S\ >\ 0$ the solution to (5.10) is given and presented by (5.20).

The next corollary is an immediate consequence of Theorem 4.

Corollary 2

If $P_T\ (S, \pi)$ is a strictly increasing function of S for each fixed π, then the unique solution S^* to problem (19) satisfies:

$$PT\ (S, \pi) =\ \gamma. \tag{5.21}$$

5.4 Fill Rate with Multiple Customers

In Section 5.2, we were looking for the model for the FR in an SLA with one customer. Now, we will extend our analysis to more than one customer. We note that similar results discussed in the previous section will apply to each customer *independently*. In this section, we will study the results that stratify to *all* customers' cases, which can be straightforward generalisations of the results in Section 5.2. For an example of these generalisations, we will examine the case of two customers $C_1$and C_2 who demand items from a supplier with base stock level S. Most of the results in Section 5.2 are applicable to this case, provided we assume that demands from each customer are independent and identically distributed, and also independent between customers.

The aggregated FR is now defined as:

$$\alpha_T^{(2)}(S) = \frac{Y_1{}^{(1)} + \cdots + Y_T{}^{(1)} + Y_1{}^{(2)} + \cdots + Y_T{}^{(2)}}{X_1{}^{(1)} + \cdots + X_T{}^{(1)} + X_1{}^{(2)} + \cdots + X_T{}^{(2)}}$$

Where $Y_i{}^{(1)}, Y_i{}^{(2)}, X_i{}^{(1)}$ and $X_i{}^{(2)}$ have the obvious interpretations of customer specific demands. Let the probability density function of $C_1{}'$s demand and of $C_2{}'$s demand be denoted by $f_1(x)$ and f_2(x), respectively. The mean and variance of X_j, j = 1, 2 will be denoted by $E(X_j)$ and $Var(X_j)$, respectively. Similarly, the mean and variance of Y_j, j = 1,2 will be denoted by $E(Y_j)$ and $Var(Y_j)$, respectively. The Folk Theorem presented in section 5.3.2 generalises to:

Theorem 5

With probability 1,

$$\alpha_T^{(2)}(S) \rightarrow \frac{E(Y_1) + E(Y_2)}{E(X_1) + E(X_2)}$$

and, in addition $E\left(\alpha_T^{(2)}(S)\right) \rightarrow \frac{E(Y_1) + E(Y_2)}{E(X_1) + E(X_2)}$

as T → ∞. Here $E(Y_j) = \int_0^{\infty} \min(S, x)\, f_j(x)\, dx$

$$= \int_0^S x\, f_j(x)\, dx + S \Pr(X_j \geq S), for\ j = 1, 2.$$

Now we apply between a pair of sums of two independent random variables, the standard results for the variance and covariance of two independent random variables and covariance: the first-order approximation to the variance of $\alpha^{(2)}{}_{(T)}(S)$, indicated by var $(\alpha^{(2)}{}_{(T)}(S))$, can simply be deducible from Equation (5.10) with apparent changes.

If we let $\sigma_{(2)}^2 = T\, Var\, (\alpha_T^{(2)}(S))$, then the asymptotic normality theorem for the FR is now given by the following result:

Theorem 6

The following asymptotic normal convergence property holds for the joint FR:

$$\sqrt{T}\,(\alpha_T^{(2)}(S) - \frac{E(Y_1)+E(Y_2)}{E(X_1)+E(X_2)}) \xrightarrow{d} \mathrm{N}\,(0, \sigma_{(2)}^2) \tag{5.22}$$

Next, we introduce a similar measure of reliability to Equation 5.17 (i.e. the FR exceeds a target π with probability of at least γ where $0 < \gamma < 1$) but we will invoke this under two customers:

$$P_T^{(2)}\,(S, \pi) \geq \gamma \tag{5.23}$$

Where $P_T^{(2)}\,(S,\pi) = Pr\left(\alpha_T^{(2)}(S) > \pi\right)$. For large T and using the asymptotic result of Theorem 6 **,** $P_T^{(2)}\,(S,\pi)$ and it can be approximated by:

$$1 - \Phi_Z\left(\frac{\sqrt{T}}{\sigma_{(2)}}(\pi - \frac{E(Y_1)+E(Y_2)}{E(X_1)+E(X_2)}\right).$$

Therefore, similar to (5.18), the target π that satisfies the reliability criterion (5.23) must also be satisfied by the next inequality:

$$\pi \leq \frac{E(Y_1)+E(Y_2)}{E(X_1)+E(X_2)} + \frac{\sigma_{(2)}}{\sqrt{T}}\,{\Phi_Z}^{-1}(1-\gamma) \tag{5.24}$$

Finally, the base stock level *S* for two customer-cases under the optimisation problem in (5.19) has a unique solution specified by the following theorem.

Theorem 7

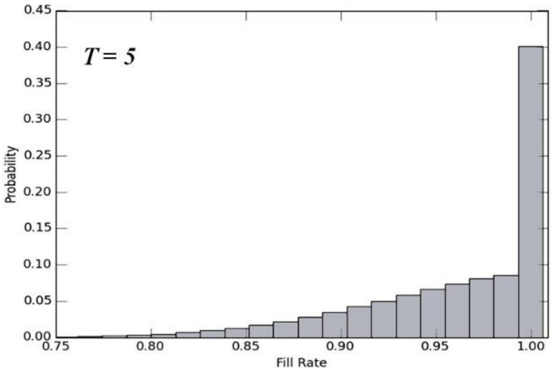
The solution S^* to the problem (5.19) with two customers is given by:

$$S^* = \inf\{\, S : P_T^{(2)}\,(S,\pi) = \gamma\} \tag{5.25}$$

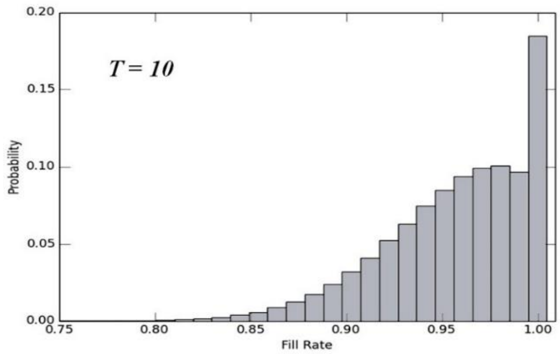
5.5 Numerical Examples Under the FR SLAs Measure

In this section, we conduct some simulation examples and studies, assuming that the supplier uses a static, periodic review, base stock policy with backorders. We analyse the impact of performance review period length on the average FR and the probability of exceeding the target FR for a supplier with single and multiple customers. The simulation was run for the one and two-customer cases, with customer demand *Erlang* distribution. We compare the results over several performance review phase durations.

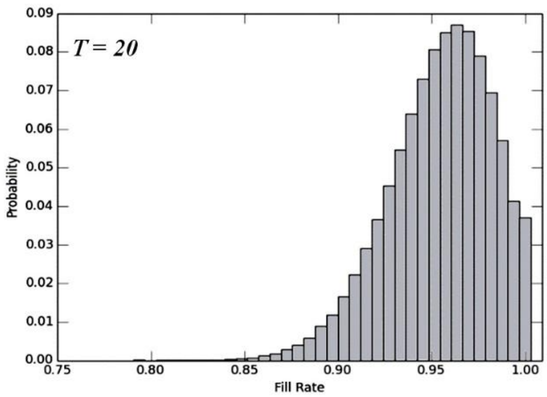
In example 1, for two cases (one and two customers), we study the sampling distributions for the FR SLA measure when we increase the size of the review periods length (i.e., T = 5, 10, 20, 100). So, in example 1, eight scenarios in total were considered. For each scenario, the simulation was run for 18,445 performance review periods. We consider that for each customer, the demand is Erlang distributed with mean $E(X) = 9$. The long-run FR is set at $\frac{E\,(Y)}{E\,(X)} = 0.95$. Base stock level ($S$) is calculated so as to achieve a long-run FR of 0.95 (in the case of multiple customers, the long-run FR is for aggregated demand), resulting in a base stock level of $S = 24$. In Figures 5.2 and 5.3 the distributions are captured. When we increase T it leads to a change in shape, and that confirms the consistency and asymptotic normality shown in Theorem 3.



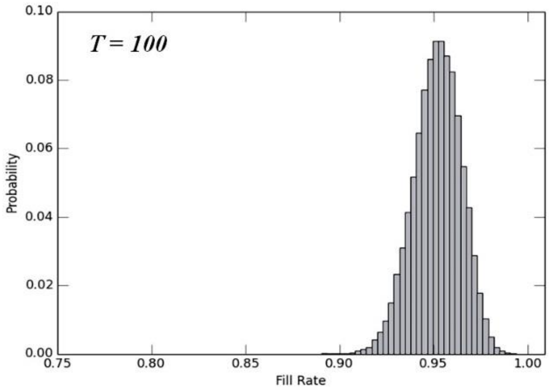
a. One customer with demand distribution of Erlang(9) and T = 5, and scale = 1



b. One customer with demand distribution of Erlang(9) and T = 10, and scale = 1

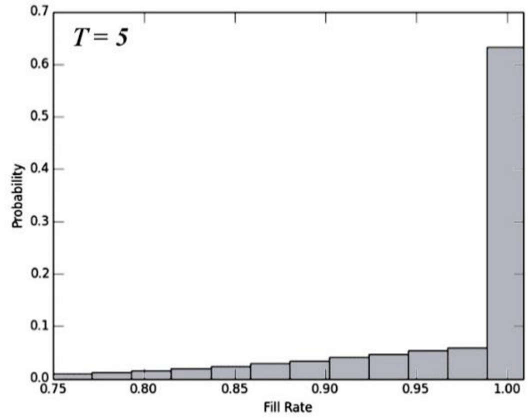


c. One customer with demand distribution of Erlang(9) and T = 20, and scale = 1

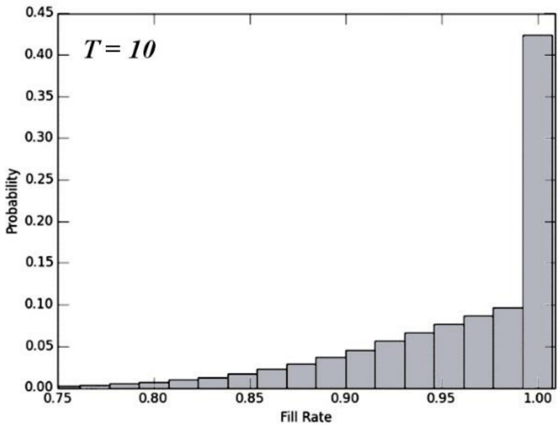


d. One customer with demand distribution of Erlang(9) and T = 100, and scale = 1

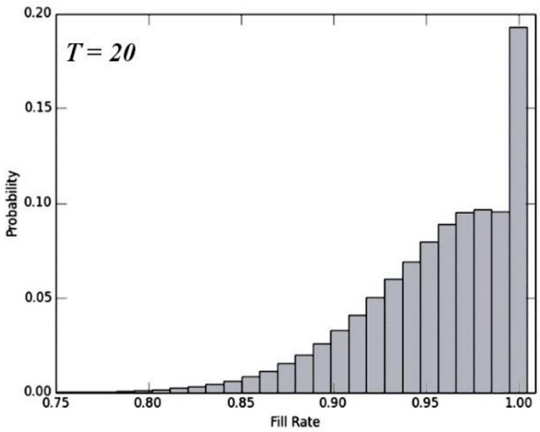
Figure 5.2: The sampling distribution of the FRs as T increases with a single customer.



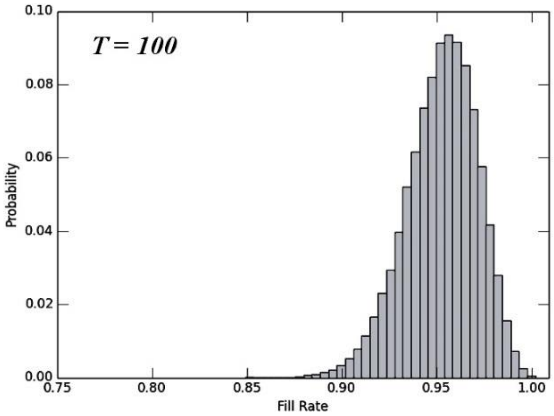
a. Two customers with total demand distribution of Erlang(9) and T = 5, and scale = 1



b. Two customers with total demand distribution of Erlang(9) and T = 10, and scale = 1



c. Two customers with total demand distribution of Erlang(9) and T = 20, and scale = 1



d. Two customers with total demand distribution of Erlang(9) and T = 100, and scale = 1

Figure 5.3: The sampling distribution of the FRs as T increases with multiple customers.

In example 2, for single and two-customer cases for each of the various scenarios, we found the required base stock level for realising a specified PS (i.e., exceeding and meeting the target FR level π with probability at least γ, c.f. equations 5.20 and 5.25). We let π = 0.95 and set $\gamma = 0.5, 0.6, 0.7, 0.9, 0.95$ and 0.99. In each case we numerically searched for the lowest base stock level (S) that can satisfy the given PS.

In Figure 5.4, the overall distribution of demand for one customer was *Erlang* with shape = 1, and scale = 8, we compare that with two customers with aggregate demand *Erlang* distributed with shape = 2, and scale = 16. It can be seen in the graphs that the minimum stock level required to realise an agreed confidence level γ is greater for the two-customer case than the one-customer case, and that this difference is greater as the review period duration increases.

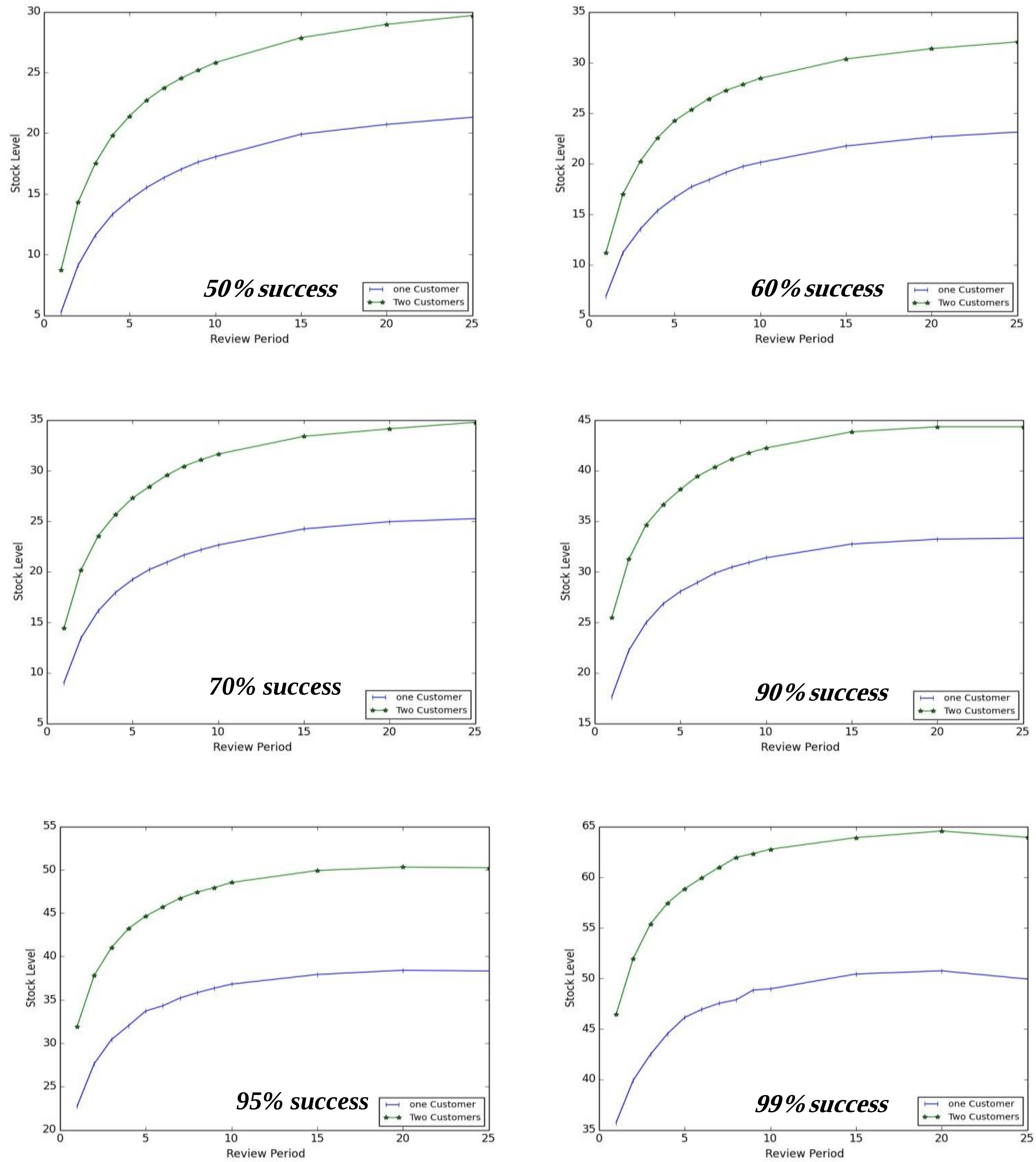


Figure 5.4: Comparison between one and two customers in terms of base stock levels required to meet target level 0.95 with various probabilities (50%, 60%, 70%, 90%, 95% and 99%).

In example 3 we apply (5.18) to find the upper bounds for the target π for several combinations of confidence level γ and different period length T = 30, 40, 50 …, 100. By applying formula (5.10), the asymptotic variance of the FR was calculated and considered to equal $\sigma^2 = 31.66$. The calculated upper bounds are tabulated (Table 5.1). An entry with an asterisk indicates that (5.18) gave a negative value, so the confidence level γ is lower than the acceptable range using the normal approximation. From Table 5.1 it can be seen that the upper bounds of these target values increase with the number of periods (except when $\gamma = 0.5$), and decrease as the confidence levels increase. This is also obvious in Figure 5.2, which displays sampling distributions that are heavily skewed to the left and move to the right when increasing the period lengths T, finally realising symmetry with large T centred at the long-run SLA FR.

Table 5.1: Upper bounds for target values for combinations of γ and T values

	T							
γ	30	40	50	60	70	80	90	100
0.50	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
0.60	0.69	0.72	0.75	0.77	0.78	0.79	0.8	0.81
0.70	0.41	0.48	0.53	0.57	0.6	0.62	0.64	0.65
0.80	0.09	0.2	0.28	0.34	0.38	0.42	0.45	0.48
0.90	*	*	*	0.02	0.09	0.14	0.19	0.23
0.95	*	*	*	*	*	*	*	0.02

5.6 Chapter Summary

In inventory management systems, the FR is an important performance measure commonly used as an indicator of reliable and efficient service by suppliers to customers. In this chapter, we discussed several results related to the FR SLA measure. essentially associated with the measure’s variability, such as consistency and asymptotic normality. Variability is important when measuring the risk involved in applying the FR, to coordinate SLAs between suppliers and their customers. Moreover, we introduced the results for a single customer, then extended those to multiple customers. Specifically, we extended our analysis to a two-customer case because, based on the assumptions imposed on our model, further increasing the number of customers does not contribute any additional insights into how to improve performance in the current systems. This chapter demonstrated that the upper bounds of these target values usually increase with number of periods, and decrease as confidence levels increase, finally realising symmetry for the long-run contract FR. These are the optimal stock levels for the SC to retain reliability over time.

Chapter 6: SLA: Ready Rate Analysis with Lump-Sum and Linear Penalty Structures

6.1 Introduction

This chapter examines the impact on the supplier's costs for ready rate SLAs measure with single and multiple customers given base stock levels, lump-sum and linear penalties, variety ready rate thresholds, and different review periods. Specifically, we examine two different types of contracts the suppliers face. Via these types of contracts, we study several impact factors, such as the base stock level, two types of penalties (lump-sum and linear penalties) and the review period duration on the supplier's cost function, and the impact of several behavouirs of the ready rate faced by the supplier.

The ready rate is the proportion of replenishment periods in which stock is available. In a typical supplier–customer contract within an SC, the ready rate is periodically measured and a financial penalty is incurred if the set target is not met. In Chapter 3, the assumption was that the FR had zero lead time. In Chapter 4, the measure was an FR with a positive lead time. This chapter uses the ready rate as the SLA performance measure, defined as the long-run fraction of periods in which all customer demand is filled immediately from on-hand stock. The scenarios considered in this chapter assume either single or multiple customers. In the first scenario, the supplier has a single customer with a large demand D. We compare this to the second scenario, in which the supplier has more than one customer, whose combined demands are the same as the single-customer case. Separate contracts are held between the supplier and each customer, and a financial penalty is imposed if the supplier cannot fulfil the contract conditions.

6.1.1 Ready rate

The ready rate was described in Chapter 1. For the purposes of this thesis, Larsen and Thorstenson's (2014) description of the ready rate is adopted: the proportion of the time that items are available immediately to the customer. The supplier agrees to predefined service levels over a constant finite period, known as the performance review phase. The supplier's inventory performance (a random variable) is measured and evaluated over a finite time horizon. In practice, a supplier will fill up the inventory up to a base stock level at a specified time interval, the replenishment review period (e.g., days, or weekly), and then supplier performance is evaluated at a regular performance review phase (e.g., monthly). Such inventory policies minimise holding and shortage costs for suppliers in an integrated SC (Taleizadeh & Noori-daryan 2016). Suppliers' service to customers may fail as inventory becomes unavailable due to unexpected demand, transport or distribution issues, pricing or inclement weather (Spiegler, Naim & Syntetos 2016). Agreements between the supplier and customer can incur a flat fee or linear penalties for missing targets over a certain period, or a finite review horizon (Larsen & Thorstenson 2014). Penalties that a supplier may incur considered in this study are lump sum (a set penalty) and linear (where there is a scaled penalty based on amount of deviation from performance targets) (Giard & Sali 2013). These penalties are examined within a scenario of multiple customers, where each has a separate contract with a supplier employing a periodic review base-stock model with zero lead time.

6.2 Notations

The following notations are used in Chapter 6:

D_i : a random variable defined as the stationary demand of customer i where $i=1,\ldots,n$ and n is the number of customers. In the case of one customer, this is simply shown by D. In this study, the aggregated demand distribution for multiple-customer cases is the same as the distribution of the one-customer case. Therefore, D in multiple-customer cases refers to the aggregated demand (i.e., $D = D_1 + \cdots + D_n$);

$D_i(m)$: the demand of customer *i* in m periods, where *i=1,…,n* and *n* is the number of customers. Index *i* drops in the one-customer case;

$D_i[t,\tau]$: the demand of customer *i* in the interval $[t,\tau]$, where *i=1,…,n* and *n* is the number of customers. Index *i* drops in the one-customer case;

$F_{D_i}(.)$: the CDF of demand of customer *i*, where *i=1,…,n* and *n* is the number of customers. Index *i* drops in the one-customer case or in the case of referring to aggregated demand in multiple-customer cases;

$f_{D_i}(.)$: the probability density function of demand of customer *i*, where *i=1,…,n* and *n* is the number of customers. Index *i* drops in the one-customer case or in the case of referring to aggregate demand in multiple-customer cases;

S : the base stock (order-up-to) level;

h : the unit inventory holding cost per period;

X_t^i: a variable that indicates whether all demand is satisfied in period t for customer i, X_t^i is 0 or 1. Index *i* drops in the one-customer case;

R: the length of performance review phase (i.e., the number of replenishment periods in a performance review phase);

η_R^i: the cumulative inventory performance (i.e., number of periods where all demand is satisfied) during a performance review phase with length *R* for customer i, $\eta_R^i = \sum_{t=1}^{R} X_t^i$. Index *i* drops in the one-customer case;

A_R^i: the ready rate in a performance review phase for customer i, $A_R^i = \eta_R^i / R$. Index *i* drops in the one-customer case;

$\propto_i$=: the performance threshold for A_R^i (i.e., the target-ready rate for customer *i)*. Index *i* drops in the one-customer case;

K_i: the penalty paid by the supplier to the customer *i* for performance below threshold $\propto_i$. Index *i* drops in the one-customer case;

$V_0(S)$: the supplier's average cost under a static base stock S policy.

6.3 Distribution of the Ready Rate

Our model in this Chapter 6 is an SC consisting of a single supplier who serves (manages the demands for) either a single or multiple customers. In the first scenario the supplier has one customer with a large demand D. We compare this to the second scenario, in which the supplier has more than one customer, whose aggregated demand is the same as in the single-customer case.

To study the impact of R on the supplier's cost function, a variety of performance review phase lengths are considered, R ϵ {5, 10, 30, 50}. Similarly, two types of penalty (lump-sum and linear) are applied. We suppose that the supplier incurs a fixed unit inventory holding cost h for each period. Moreover, we assume that the supplier applies a periodic review base stock policy with a base stock level S. A customer order placed at the beginning of period t must be fulfilled within the same period, and any unmet demand is backordered. Let $X_t \in \{0, 1\}$ be the ready rate for period t, (where $1 \leq t \leq R$). X_t = 1 if all demand is filled in period *t*, conversely X_t = 0 if some demand is not filled in period *t* and requires backordering. When the supplier has a contract with a single customer, the probability that all customer demand in a single period is satisfied equals:

$$\Pr\{X_t = 1\} = Pr\{D[t, t+1) \leq S\} = \Pr\{D(1) \leq S\} = F(S) \tag{6.1}$$

When the demand is independent and identically distributed, and a stationary base stock policy is used (i.e., S is fixed for all *t*), the ready rate X_t for each period t has an identical distribution. So, the cumulative ready rate η_R has a binomial distribution:

$$\Pr\{\eta_R = i \mid S\} = \binom{R}{i} [F(S)]^i [1 - F(S)]^{R-i} \tag{6.2}$$

Where R is the number of trials and $F(S)$ is the probability of a success.

When the supplier has an agreement with two customers (buyers), then D_1 and D_2 are the respective random demands of customers 1 and 2 in each replenishment review period, with corresponding probability density function $f_{D_1}(.)$ and $f_{D_2}(.)$. The demand summation for both customers per replenishment review period is given by $D = D_1 + D_2$. Note that replenishment review periods here refer to equally spaced time intervals with R periods in a single performance review phase. Due to the stationarity of demand assumption, the events occurring in each replenishment review period are statistically identical to other replenishment review periods.

We assume that the customers are served in accordance with an FCFS policy, and that once a customer comes, their demand should be met in full. In the case of two customers, there is a 50% chance that customer 1 will come first. The sequence of the customer arrivals is defined by a binary variable (Y), such that Y = 1 when customer 1 arrives first, and Y = 0 otherwise. We assume P (Y = 0) = P (Y = 1) = 0.5. Therefore, the average ready rate for customer 1 of a supplier with two customers (buyers) is given by the following equation (recall that $D = D_1 + D_2$):

$$A_R^1 = P(D \leq S) + P(Y = 1, D_1 \leq S, D > S) \tag{6.3}$$

Similarly, the average ready rate for customer 2 is given by:

$$A_R^2 = P(D \leq S) + P(Y = 0, D_2 \leq S, D > S) \tag{6.4}$$

Here we define the following disjoint events and their probabilities:

ξ_1 = probability that both customer demands are fully satisfied within a period;

ξ_2 = probability that only customer 1's demand is satisfied within a period;

ξ_3 = probability that only customer 2's demand is satisfied within a period;

ξ_4 = probability that neither customer's demands are fully satisfied within a period.

A demand is fully satisfied if it is less than or equal to the base stock level S. To find the above probabilities, we partition the region $\{D_1 \geq 0\} \cup \{D_2 \geq 0\}$ as:

$E1 \quad = \{D \leq S\}$

E2 $\quad$ = {Y = 1, D1 ≤ S, D > S}

E3 $\quad$ *= {Y = 0, D2 ≤ S, D > S}*

E4 $\quad$ *= {Y = 1, D1 > S, D > S} U {Y = 0, D2 > S, D > S}*

Therefore:

$$\xi_1 = P(E1) = P(D \leq S) = \int_0^S \int_0^{S-x_1} f_{D_1}(x_1) f_{D_2}(x_2) dx_2 dx_1 \tag{6.5}$$

$$\xi_2 = P(E2) = 0.5 * P(D_1 \leq S \,|D > S\,) * P(D > S)$$

$$= 0.5 \int_0^S \int_{S-x_1}^{\infty} f_{D_1}(x_1) f_{D_2}(x_2) dx_2 dx_1 \tag{6.6}$$

$$\xi_3 = P(E3) = 0.5 * P(D_2 \leq S \,|D > S\,) * P(D > S)$$

$$= 0.5 \int_0^S \int_{S-x_2}^{\infty} f_{D_1}(x_1) f_{D_2}(x_2) dx_1 dx_2 \tag{6.7}$$

$$\xi_4 = P(E4) = 1 - \xi_1 - \xi_2 - \xi_3 \tag{6.8}$$

Since the four events are mutually exclusive and exhaustive,

$$\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1 \tag{6.9}$$

Now we can explain the following multinomial distribution concerning the number of possible events in a performance review phase that supports us to derive the ready rate:

$$\Pr\{ m, i, j, R-i-j-m \,|\mathrm{S}\} = \binom{R}{m\ i\ j\ R-i-j-m} \xi_1^m\, \xi_I^{\,i}\, \xi_3^j\, \xi_4^{R-i-j-m} \tag{6.10}$$

For instance, the probability that in exactly g time periods out of R time periods, the demand of customer 1 is satisfied (that means η_R^1= g) is given by:

$$\sum_{m=0}^{g} \sum_{j=0}^{R-g} \binom{R}{m\ g-m\ j\ R-g-j} \xi_1^m\, \xi_2^{g-m}\, \xi_3^j\, \xi_4^{R-g-j}$$

This expression calculates all possible combinations that in g periods out of R, the demand of customer 1 is fully satisfied.

6.4 The Supplier Cost Function Under a Lump-Sum Penalty SLA

First, we consider the supplier's cost function under the application of a fixed lump-sum penalty for two scenarios: when the supplier has one customer, and when the supplier has multiple customers. A lump-sum penalty SLA is one in which a supplier must pay a fixed penalty K to its customer whenever the ready rate for this customer is less than a specified threshold, that is $\mathrm{A_R} < \propto$ or, equivalently, $\eta_\mathrm{R} < \propto R$.

Therefore, the formula for the supplier's average cost with backlogging under a static base stock S policy and one customer is (Liang & Atkins 2013):

$$V_0(S) = hE\,[S - D]^+ + \frac{K}{R} \sum_{i=0}^{\lceil \propto R \rceil - 1} \Pr\{\eta_R = i \mid S\} \tag{6.11}$$

Where $\Pr\{\eta_R = i \mid S\}$ is given by (6.2).

In the second scenario, where the supplier has two customers, the supplier's average cost formula under the lump-sum penalty is obtained as:

$$V_0(S) = h\,E\,[S - D]^+$$

$$+ \frac{K_1}{R} \sum_{m=0}^{\lceil \propto_1 R\rceil - 1} \sum_{i=0}^{\lceil \propto_1 R\rceil - 1 - m} \sum_{j=\max(0, \lceil \propto_2 R\rceil - m)}^{R-i-m} \binom{R}{m\ i\ j\ R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m}$$

$$+ \frac{K_2}{R} \sum_{m=0}^{\lceil \propto_2 R\rceil - 1} \sum_{j=0}^{\lceil \propto_2 R\rceil - m - 1} \sum_{i=\max(0, \lceil \propto_1 R\rceil - m)}^{R-j-m} \binom{R}{m\ i\ j\ R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m}$$

$$+ \frac{(K_1 + K_2)}{R} \sum_{m=0}^{\min(\lceil \propto_1 R\rceil - 1, \lceil \propto_2 R\rceil - 1)} \sum_{i=0}^{\lceil \propto_1 R\rceil - 1 - m} \sum_{j=0}^{\min(\lceil \propto_2 R\rceil - m, R-m-i)} \binom{R}{m\ i\ j\ R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m} \tag{6.12}$$

The first term of Equation (6.12) computes the holding cost. The second term of Equation (6.12) calculates the penalty incurred when the target-ready rate of customer 1 is not achieved, but the target-ready rate of customer 2 is met. Thus, in the second term, the summation of the number of periods when only customer 1's demand is satisfied (with probability ξ_2), and those periods when both customers are satisfied (with probability ξ_1) is less than $\lceil \propto_1 R\rceil - 1$. Conversely, the summation of the number of periods when only customer 2's demand is satisfied (with probability ξ_3) and those periods when both customers are satisfied (with probability ξ_1) is greater than or equal to $\lceil \propto_2 R \rceil$. The third term of Equation (6.12) is similar to the second term but calculates the penalty incurred when the target-ready rate of customer 2 is not achieved. The last term calculates the penalty incurred when neither customer's demand is satisfied.

6.5 The Supplier Cost Function Under a Linear Penalty SLA

In the linear penalty SLA, the supplier is imposed to pay the customer a penalty proportional to the time that the supplier's inventory performance is insufficient to satisfy the SLA. This means that the penalty K is charged over $(\alpha R - \eta_R)^+$ periods of the review phase R.

The formula below shows the supplier's cost function under a linear penalty SLA, with a backlog assumption in the one-customer case (Liang & Atkins 2013):

$$V_0(S) = hE\,[S - D]^+ + \frac{K}{R} \sum_{i=0}^{\lceil \propto R \rceil - 1} \Pr\{\eta_R = i \mid S\} \tag{6.13}$$

where $\Pr\{\eta_R = i \mid S\}$ is given by (6.1).

$$\begin{aligned}
&V_0(S) \\
&= h\,E\,[\,S - D]^+ \\
&+ \frac{K_1}{R} \sum_{m=0}^{\lceil \propto_1 R \rceil - 1} \sum_{i=0}^{\lceil \propto_1 R \rceil - 1 - m} \sum_{j=\max(0, \lceil \propto_2 R \rceil - m)}^{R-i-m} (\lceil \propto_1 R \rceil - 1 - i - m) \binom{R}{m\; i\; j\; R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m} \\
&+ \frac{K_2}{R} \sum_{m=0}^{\lceil \propto_2 R \rceil - 1} \sum_{j=0}^{\lceil \propto_2 R \rceil - m - 1} \sum_{i=\max(0, \lceil \propto_1 R \rceil - m)}^{R-j-m} (\lceil \propto_2 R \rceil - 1 - j - m) \binom{R}{m\; i\; j\; R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m} \\
&+ \frac{K_1}{R} \sum_{m=0}^{\min(\lceil \propto_1 R \rceil - 1, \lceil \propto_2 R \rceil - 1)} \sum_{i=0}^{\lceil \propto_1 R \rceil - 1 - m} \sum_{j=0}^{\min(\lceil \propto_2 R \rceil - m, R-m-i)} (\lceil \propto_1 R \rceil - 1 - i - m) \binom{R}{m\; i\; j\; R-i-j-m} \xi_1^m\, \xi_2^i\, \xi_3^j\, \xi_4^{R-i-j-m} \\
&+ \frac{K_2}{R} \sum_{m=0}^{\min(\lceil \propto_1 R \rceil - 1, \lceil \propto_2 R \rceil - 1)} \sum_{i=0}^{\lceil \propto_1 R \rceil - 1 - m} \sum_{j=0}^{\min(\lceil \propto_2 R \rceil - m, R-m-i)} (\lceil \propto_2 R \rceil - 1 - j - m) \binom{R}{m\; i\; j\; R-i-j-m}
\end{aligned} \tag{6.14}$$

The first term of Equation (6.14) calculates the holding cost. The second term of Equation (6.14) calculates the penalty incurred when the target-ready rate of customer 1 is not reached but that of customer 2 is met. However, in the second term, the summation of the number of periods that only customer 1's demand is fulfilled (with probability ξ_2) and the number of periods that the demands of both customers are fulfilled (with perobability ξ_1) is less than $\lceil \propto_1 R \rceil - 1$. Conversely, the summation of the number of periods that only customer 2's demand is met (with perobability ξ_3) and the number of periods that the demands of both customers are satisfied (with perobability ξ_1) is greater than or equal to $\lceil \propto_2 R \rceil$. The third term of Equation (6.14) is like the second term, but calculates the penalty incurred when the target-ready rate of customer 2 is not achieved. The fourth and fifth terms calculate the penalty incurred when neither customer's demand is satisfied. In contrast to Equation (6.12), this requires a separate calculation for each customer, to account for differences in the amount each customer's realised ready rate deviates from the target-ready rate.

6.6 Numerical Examples

Numerical examples of the two types of penalty (fixed and linear) are used to answer the chapter's questions. The assumption is that the supplier applies a static, periodic review, base stock policy with backorders.[1] First, the multiple-customer case is run, then compared to the single customer scenario. Due to the number of variables in the first case, a Monte Carlo simulation approach with 14,000 replications is used to estimate the cost function and expected total cost (Abbasi et al. 2017; Thomas 2005).

[1] As there are backorders, the linear order cost is not included in the cost function, and the backorder cost is captured in the penalty term.

6.7 Related Research Question for Chapter 6

For each type of penalty, we address the following questions:

- (Q6.1) What base stock level is required to minimise the supplier's expected cost?
- (Q6.2) How is the supplier's expected cost affected by different performance review phase durations?
- (Q6.3) How does increasing demand affect the supplier's expected cost?
- (Q6.4) What is the impact of different target ready rate levels on expected total cost of the supplier for one and multiple-customer cases?

All experiments were performed on an Intel Core i7 CPU with a 2.2 GHz processor and 16 GB RAM. The codes were written in Python programming language (Python 3.4 version). To solve these models we applied other packages to solve methods, such as Numpy, Sicipy, Matplotlib and Anaconda. The computational time taken was around 15 minutes (for three and four-customer cases under two types of penalties). When solving the numerical formulas (6.4 and 6.6 for two customers), it took a long time (1–2 hours, especially when we increased the performance review period length) due to many loops and the multiple summations in both formulas.

6.7.1 Numerical study for lump-sum penalty SLA

In our numerical experiments, we considered that demand follows an *Erlang* distribution with shape parameter ζ and scale parameter 1, this was consistent with previous numerical studies of SLAs (Abbasi et al. 2017; Thomas 2005). In the case of multiple customer, the demand distribution of each customer is similarly *Erlang* with parameter ζ_i, where i indicates the customer number $i,\ldots,n$. When comparing the one-customer case and the several customer cases, the aggregated demand distribution is similar for all. In the next numerical examples, we examine the impact of increasing the number of customers and the effect of different performance review phase durations $R \in \{5, 10, 30, 50\}$. All numerical examples were examined under the ready rate with zero lead time.

Numerical Example 1

The first numerical example examines the case of when a supplier serves one customer and the demand distribution is Erlang with shape parameter $\zeta = 10$ and the scale parameter 1. Other parameters are set as $h = 1$, $R = 5, 10, 30, 50$, $K = 16 \times R$, and α $= 0.9$. The results from this single-customer case are compared with the multiple (two, three and four) customer cases, where the demand distribution for each customer is Erlang with the shape parameter $\zeta_i = \frac{10}{n}$, where n is the number of customers

and i indicates the customer number, and scale parameter is 1, $\propto_i = 0.9$ $K_i = 16 \times \frac{R}{n}$ for $i = 1, \ldots, n$.

Numerical Example 2

The next second numerical example is same as the previous numerical example, but we increase the shape parameter of the one customer case and aggregated demand to 20.

i.e., $\zeta = 20$ to study the effect of expected total cost for the supplier.

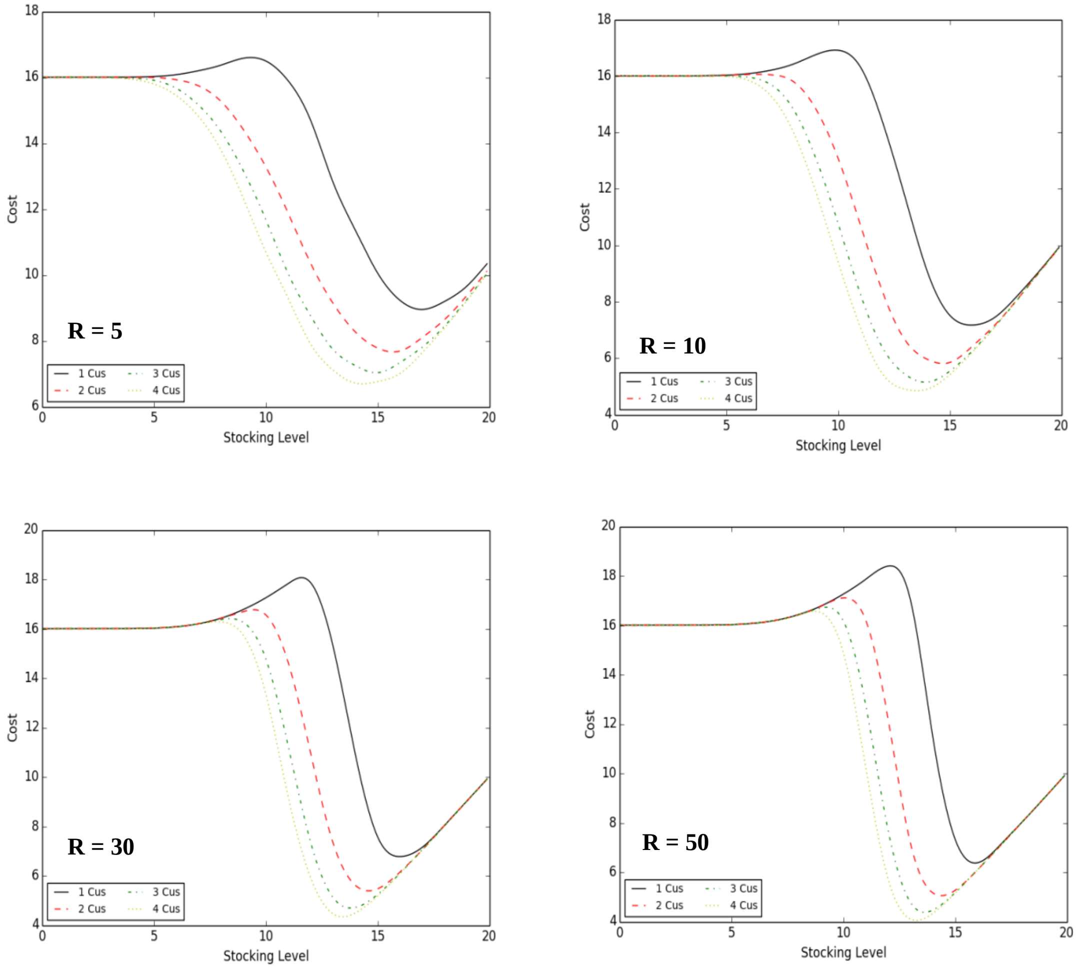


Figure 6.1: The results of Example 1. Cost of fixed-sum penalties: comparison of one, two, three and four-customer cases with varying base stock levels and review period durations.

Figure 6.1 shows the results of one, two, three and four-customer cases examined under a lump-sum penalty. It can be seen that the expected total cost changes when the base stock level increases for several performance review phase lengths. When we increase *S*, the penalty decreases and is lowered for the multiple-customer cases more than the one-customer cases. We also noticed the following pattern: when we increased the base stock level *S*, the cost function is first quite flat and then starts to rise. As *S* is further increased, a marked drop in cost is observed, followed by a steep and continued rise. For all performance review phase durations, this pattern was consistent and corresponds with what Liang and Atkins (2013) found in their study of the lump-sum penalty SLA in the single-customer case. We also examined, in numerical example 2, the increase in customer demand, and found that the pattern also holds when customer demand increases.

Figure 6.2: The results of Example 2. Cost of fixed-sum penalties: comparison of one, two, three and four-customer cases with varying base stock levels and review period durations.

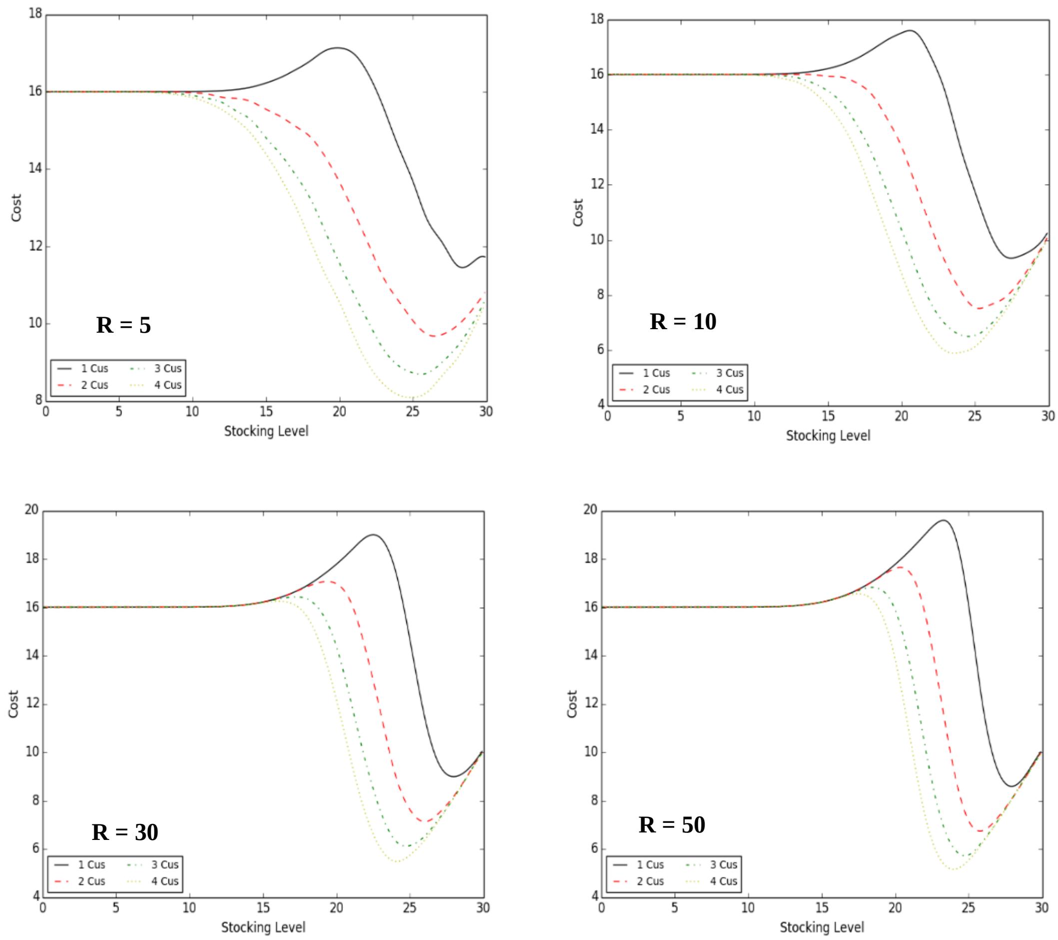


Table 6.1: Optimal base-stock and cost for fixed sum penalty contract – Numerical Example 1

R	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
5	*17*	*15*	*15*	*14*	*8.98*	*7.70*	*7.05*	*6.72*
10	*16*	*15*	*14*	*14*	*7.26*	*5.86*	*5.20*	*4.88*
30	*16*	*15*	*14*	*13*	*6.68*	*5.42*	*4.66*	*4.47*
50	*16*	*14*	*14*	*13*	*6.45*	*5.22*	*4.12*	*4.07*

Table 6.2: Optimal base-stock and cost for fixed sum penalty contract – Numerical Example 2

R	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
5	*28*	*28*	*25*	*24*	*11.41*	*9.90*	*8.68*	*8.11*
10	*28*	*25*	*25*	*24*	*9.47*	*7.52*	*6.54*	*5.85*
30	*28*	*26*	*25*	*24*	*8.99*	*7.07*	*6.24*	*5.50*
50	*28*	*26*	*25*	*24*	*8.72*	*6.76*	*5.76*	*5.11*

Under the lump-sum penalty SLA, Tables 6.1 and 6.2 show the base stock levels (*S*) that minimise the expected total cost. When the number of customers increases, the optimal total cost decreases too. As mentioned, aggregated demand is the same in the one-customer case and in all multiple-customer cases, but despite this equality of demand, we see that having the aggregate demand split over a greater number of customers will assist the supplier to lower costs. The explanation for this observed effect is that in instances where there is insufficient stock to satisfy the aggregate demand, when the demand distribution is split amongst multiple customers there is a chance that at least some of the customers will have their demand satisfied in full and hence a penalty need not be paid to these parties. We also found that a lower base stock level is needed to achieve the cost minima when the number of customers is increased. In Figures 6.1 and 6.2 this effect can be clearly observed. These results will assist the supplier and provide the following management insight under a lump-sum penalty SLA. If the aggregate demand is divided between more customers, the base stock level necessary for optimal performance will be lower and lower total costs will be incurred.

Finally, we examine the effect of increasing demand (for one and multiple-customer cases) on the supplier's expected cost. In comparing Tables 6.1 and 6.2 when the optimal base stock level is employed, it can be seen that an increase in demand leads to an increase in expected cost. This effect is observed in the one-customer and in all multiple-customer cases, and across all performance review phase durations.

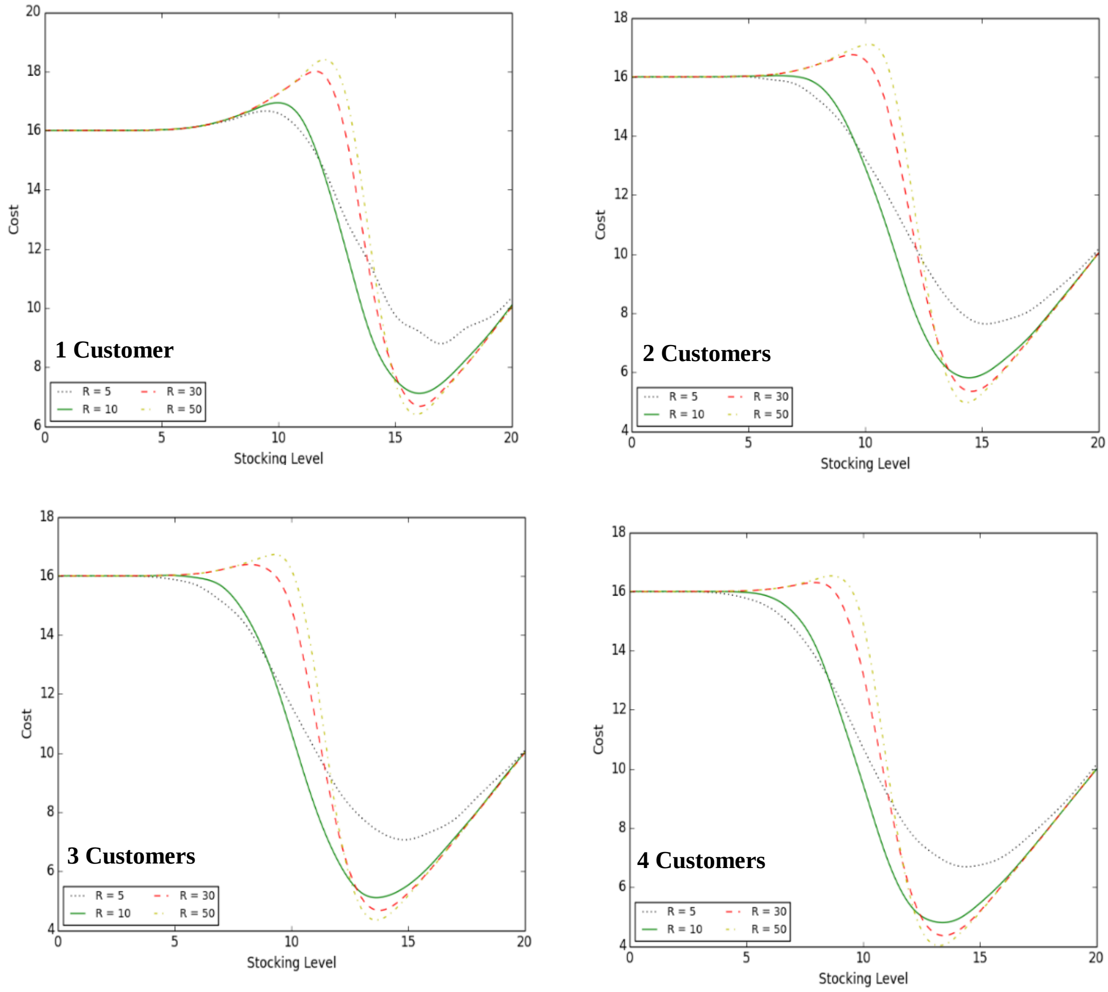


Figure 6.3: The expected total cost of lump-sum penalty, comparison of different performance review phase durations.

Figures 6.3 and 6.4 display the relationship between the expected total cost and the performance review phase duration of R for the supplier, when $R = \{5, 10, 30 \text{ and } 50\}$ (the same as in numerical example 1 for Figure 6.1, and numerical example 2 in Figure 6.2). The optimal expected total cost becomes lower as the performance review phase duration rises for all (single or multiple-customer) cases.

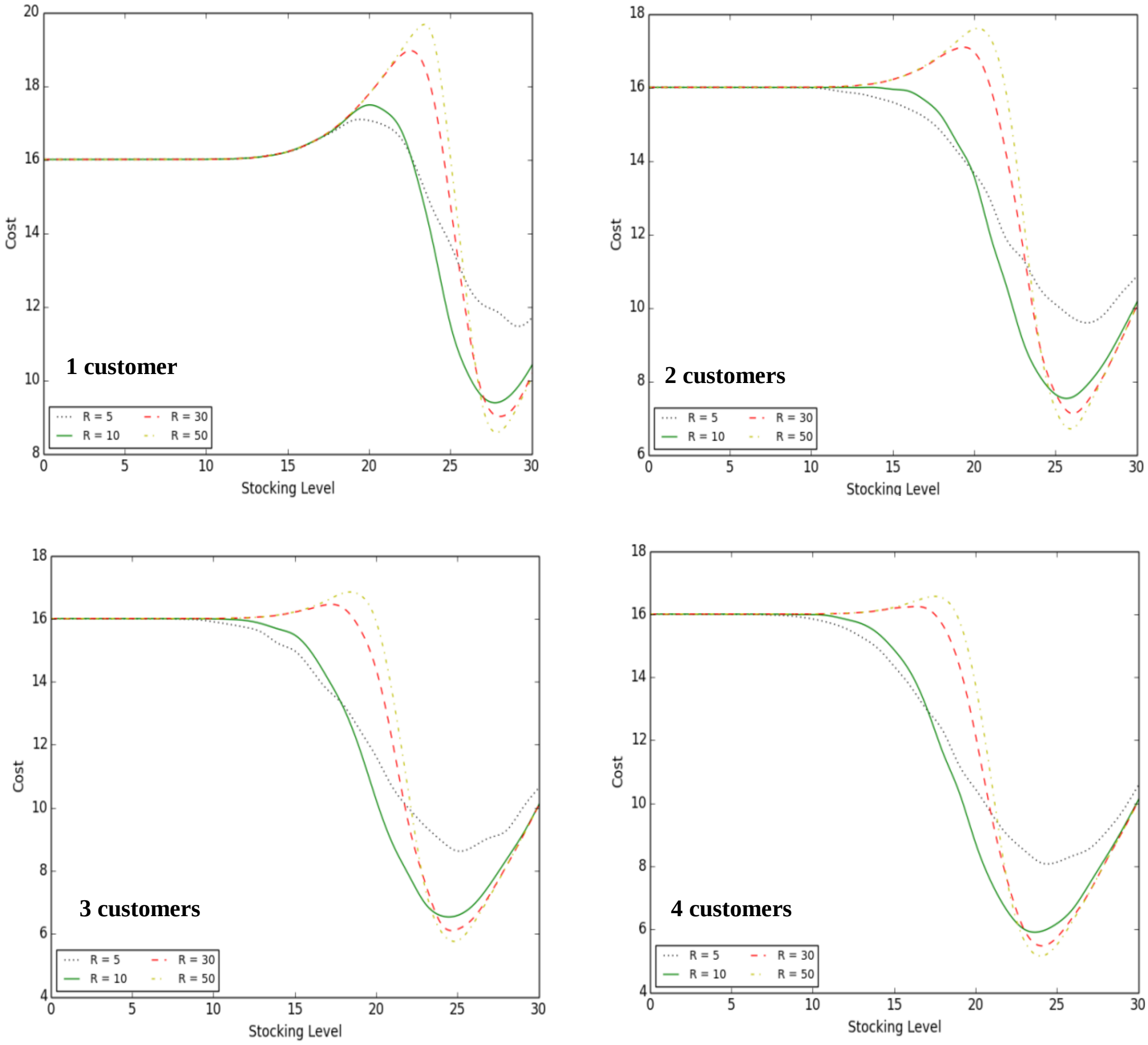


Figure 6.4: The expected total cost of lump-sum penalty when increasing the demand, comparison of different performance review phase durations.

It can be seen that the optimal total cost decreases as the number of customers increase, and for more customers a lower base stock level is necessary to fill demand. In real-life multiple customer scenarios, it is also likely that in the event of a stock-out, the supplier can elect to fill demand for the highest penalty customer first. However, an increase in demand leads to an increase in expected costs (and profits) as the supplier expands.

6.7.2 Numerical study for linear penalty SLA

In this subsection, the results for SLAs with a linear penalty are presented. All notations and parameters are similar to the numerical examples presented in Section 6.7.1.

Numerical Example 3

In this numerical example, all parameters are similar to numerical example 1, except the penalty coefficients. The penalty structure is assumed to be linear (see equations (6.5) and (6.6)). We defined the penalty parameter as $K = 4 \times R$. The demand distribution of each customer is Erlang with shape $\zeta_i = \frac{10}{n}$ where n is the number of customers ($i = 1, \dots, n$) and the scale parameter of the Erlang distribution is set to one in all experiments.

Numerical Example 4

To examine the impact of increased demand, the fourth numerical example is similar to numerical example 3. The only difference is the shape parameter of the demand distribution. In this example $\zeta_i = \frac{20}{n}$ where n is the number of customers.

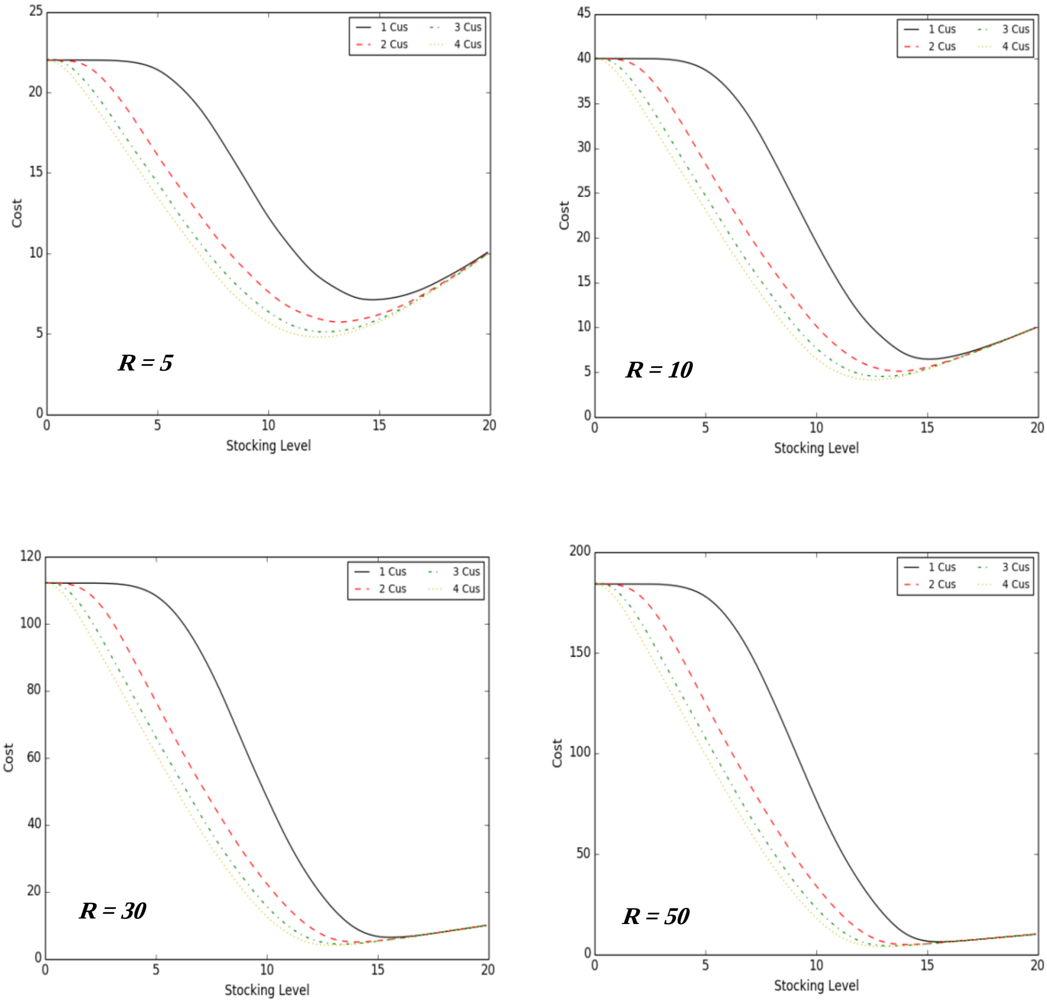


Figure 6.5: The results of Numerical Example 3, cost of linear penalties: comparison of one, two, three and four-customer cases with varying base stock levels and review period durations.

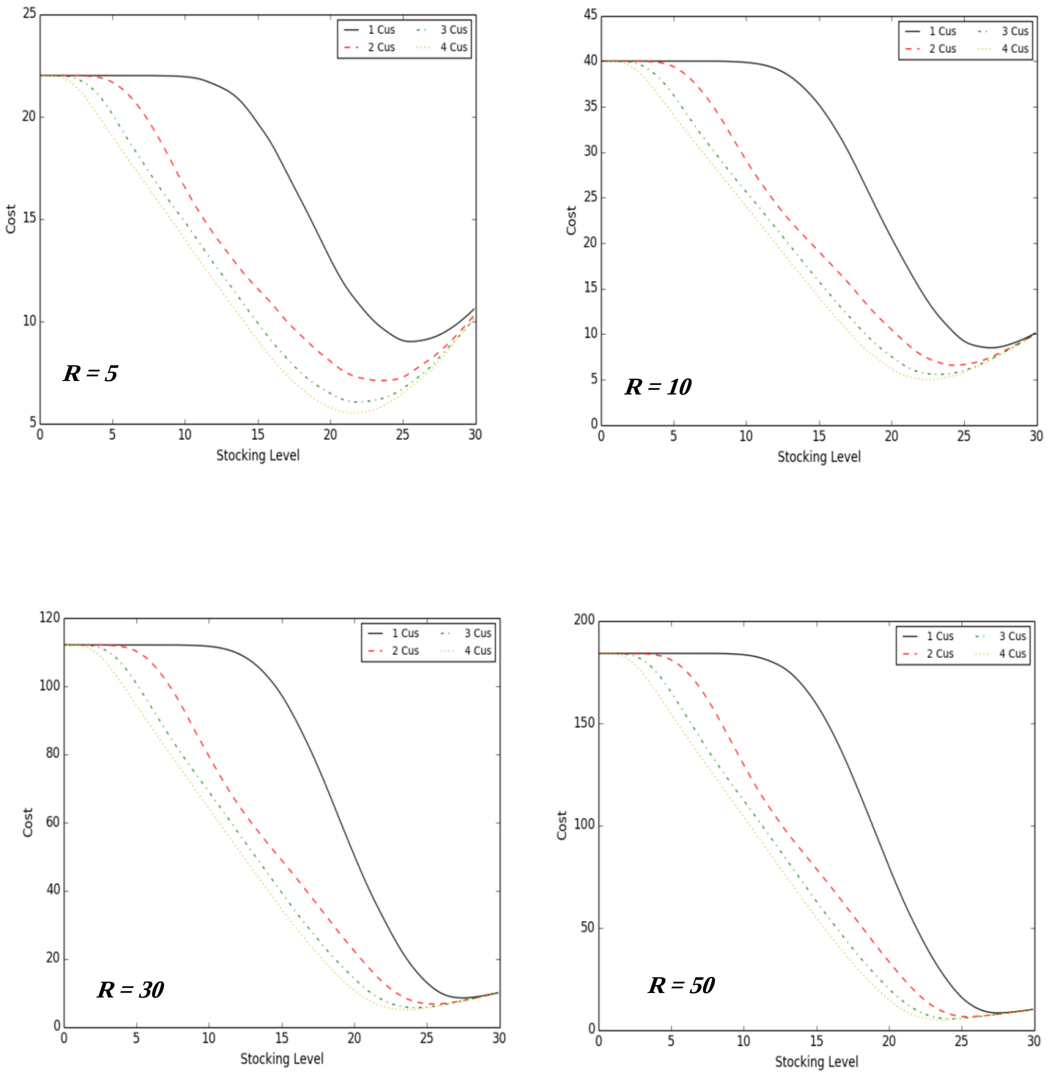


Figure 6.6: The results of Numerical Example 4, cost of linear penalties: comparison of one, two, three and four-customer cases with varying base stock levels and review period durations.

In Section 6.8.1 we presented the numerical results for the lump-sum penalty SLA. In contrast to the previous results, Figures 6.5 and 6.6 display that in multiple-customer cases under a linear penalty, the supplier's cost function is unimodal. This result is analogous with Liang and Atkins' (2013) result for a supplier with a single customer under a linear penalty SLA. When compared to the previous results in Figures 6.1 and 6.2 for the lump-sum penalty SLA, Figures 6.5 and 6.6 show that the cost of choosing too small a base stock level *S* is much higher when there is a linear penalty in place. This result occurs because under a linear penalty the expected cost rises based on how much the performance target is missed by, while under a lump-sum SLA penalty, once the target has been missed, any additional performance deterioration is costless. This phenomenon has been already mentioned by Liang and Atkins (2013) for the one-customer case.

Table 6.3: Optimal base-stock and cost for linear penalty contract (Numerical Example 3)

R	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
5	*15*	*13*	*13*	*12*	*7.17*	*5.80*	*5.15*	*4.81*
10	*15*	*14*	*13*	*13*	*6.53*	*5.10*	*4.42*	*4.11*
30	*16*	*14*	*13*	*13*	*6.41*	*4.94*	*4.44*	*4.01*
50	*16*	*14*	*13*	*13*	*6.22*	*4.87*	*4.25*	*3.77*

Table 6.4: ***Optimal base-stock and cost for linear penalty contract (Numerical Example 4)***

R	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
5	*26*	*24*	*22*	*21*	*9.02*	*7.06*	*6.04*	*5.49*
10	*27*	*25*	*23*	*23*	*8.48*	*6.62*	*5.54*	*5.01*
30	*28*	*25*	*24*	*24*	*8.63*	*6.75*	*5.66*	*5.08*
50	*27*	*25*	*24*	*24*	*8.39*	*6.52*	*5.42*	*4.86*

Tables 6.3 and 6.4 show that as the number of customers increases, the optimal expected total cost decreases, as does the base-stock level needed to achieve the cost minima. These findings are different from what was observed for the lump-sum penalty SLA in Section 6.8.1. Conversely, in a linear penalty SLA, the results specify that the supplier's optimal expected total cost is not monotone with respect to the different performance review phase duration (R). This can be seen in Tables 6.3 and 6.4; that is, as R increases, the optimal expected total cost may decrease or increase. Under a lump-sum penalty SLA, the supplier always benefitted from a longer R. Similarly to the lump-sum penalty SLA, in comparing Tables 6.3 and 6.4 it is obvious that under a linear penalty SLA, an increase in demand leads to an increase in expected cost.

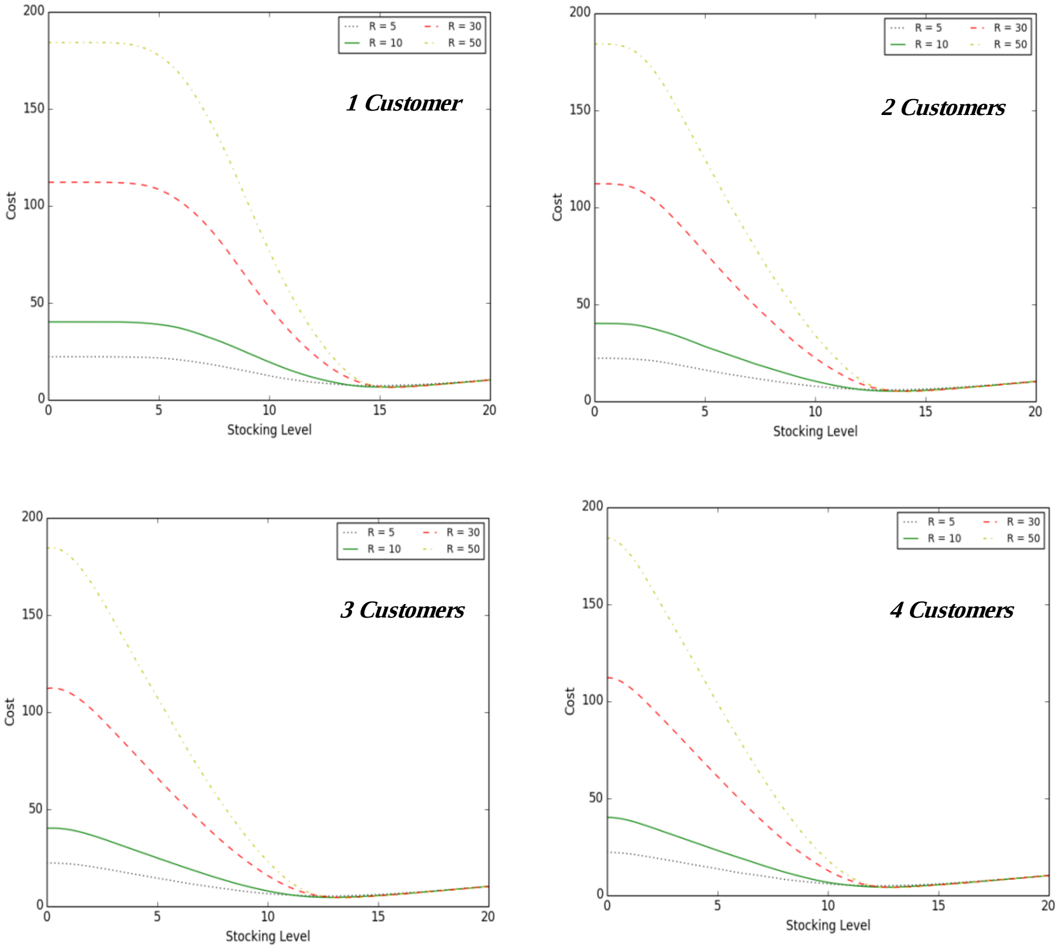


Figure 6.7: The expected total cost of linear penalties, comparison of different performance review phase durations.

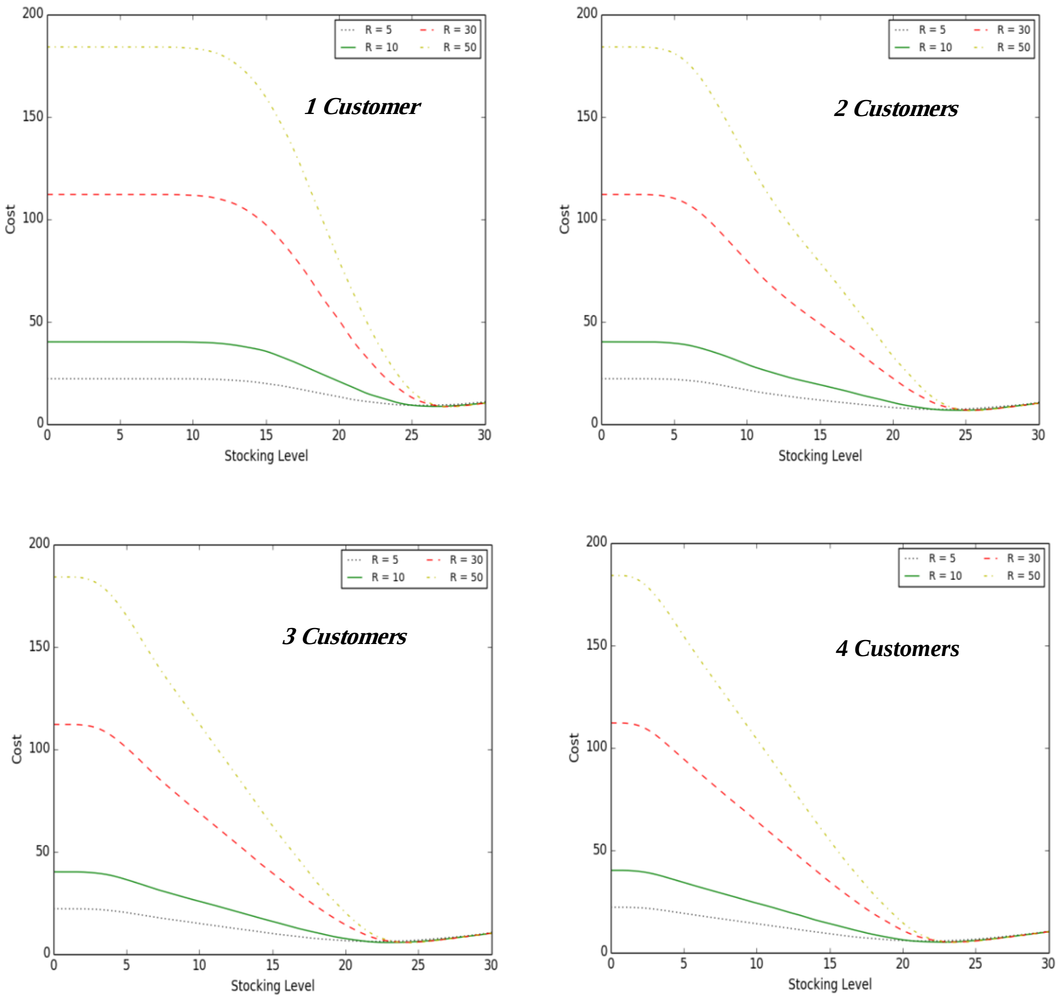


Figure 6.8: The expected total cost of linear penalties when increasing demand, comparison of different performance review phase durations.

The relationship between the expected total cost and the performance review phase duration *(R)* for all cases given a linear penalty structure are elucidated in Figures 6.7 and 6.8. It can be seen that it is beneficial to the supplier to have a shorter review phase length, as the supplier can achieve a lower penalty than for a longer review phase.

6.8 Studying the SLA of Ready Rate Threshold

Previously, the ready rate performance threshold was fixed as $\alpha = 0.90$. In this section, the two forms of penalties (lump sum and linear) are examined for one customer or more customers, with the ready rate as $\alpha = 0.60$, 0.70, 0.80 and 0.90, and the review period as $R = 30$ in all scenarios.

6.8.1 Lump-sum penalty SLA

Numerical example 5

This numerical example uses the lump-sum penalty. The parameters are (α, K) with one or more customers, and the demand distribution is *Erlang* with shape parameter $\zeta = 10$ and the scale parameter 1:

$R = 30$, $h = 1$, lump-sum $(\alpha, K) \in \{(0.60, 200), (0.70, 170), (0.80, 130), (0.90, 110)\} \subset \Theta$.

Figure 6.9 shows how the suppliers' cost function changes when dealing with single or multiple customers under different ready rate performance thresholds. In Figure 6.9 we observed the following pattern: for both cases (one and multiple-customers), when the ready-rate threshold is increased and the penalty is decreased, the expected total cost is lower. In all cases as S get larger, the expected total cost is initially quite steady, then as S increases a marked drop in cost is observed, followed by a steep and continued rise. This was observed when $\alpha = 0.60$, 0.70, 80. However, for both cases when $\alpha = 0.90$, the expected total cost rises up without decreasing, as long as S increases, whereas the optimal base stock level $S = 0$. This pattern was consistent through all multiple-customer cases modelled, and mirrored what Liang and Atkins (2013) observed in their study of the lump-sum penalty SLA for the single-customer case.

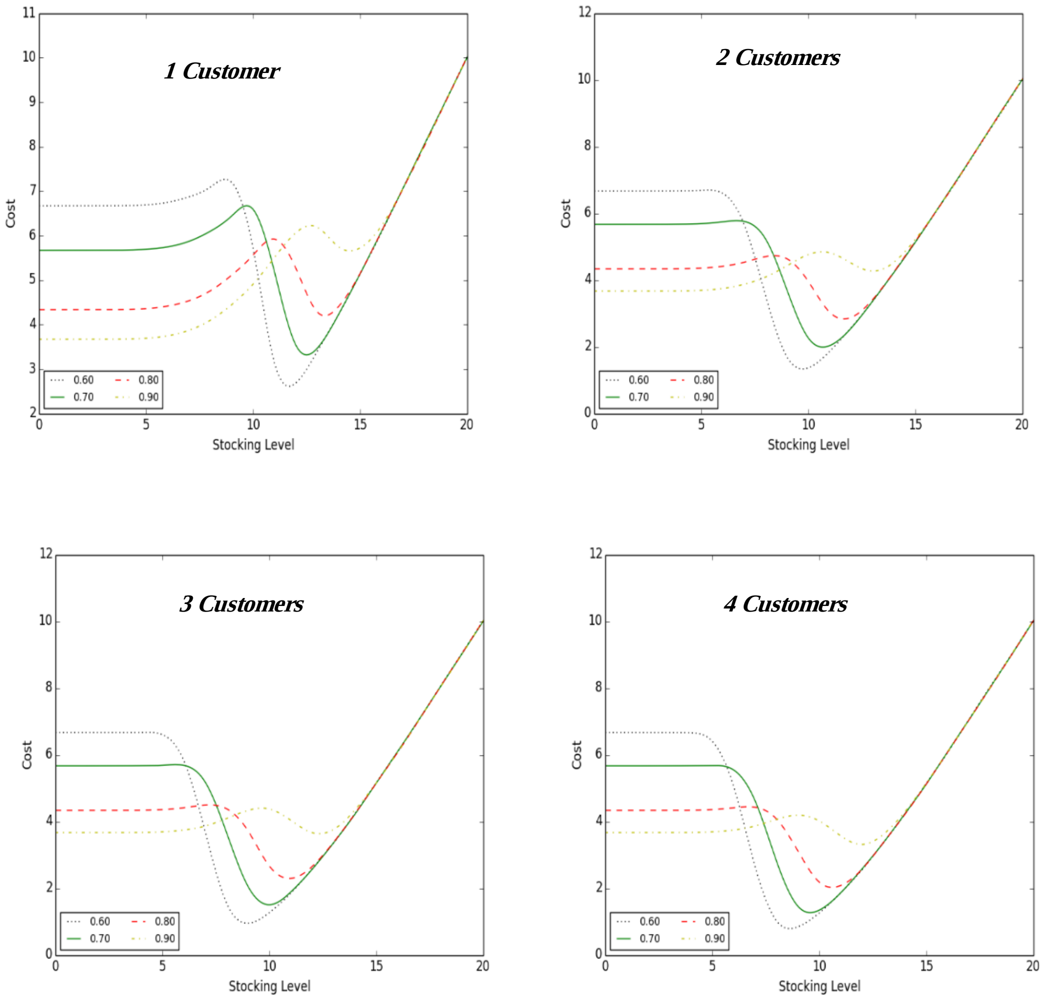


Figure 6.9: The expected total cost of lump-sum penalty, comparison of single and multiple customers when ready rate performance α=0.60, 70, 80 and 0.90.

Under different ready rate thresholds, Table 6.5 display the base stock levels (S) that minimise the expected total cost for the lump-sum penalty SLA. Table 6.5 shows that the optimal expected total cost decreases as the number of customers increases. It can be seen from Table 6.5 that spreading the demand over more customers will lead to lower costs and base stock levels. In examples where there is insufficient stock to satisfy the aggregate demand, when the demand distribution is divided over multiple customers there is a chance that at least some customers will have their demand satisfied in full, so a penalty need not be paid to these parties. Moreover, as the number of customers increases, a lower base stock level is needed to realise the cost minima. This result can be observed in Figure 6.9.

Table 6.5: Optimal base stock and cost for fixed penalty contracts, ready rate thresholds

Ready rate thresholds	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
0.60	12	10	9	9	2.58	1.38	0.94	0.84
0.70	13	11	10	10	3.47	2.02	1.50	1.36
0.80	13	12	11	11	4.40	2.89	2.32	2.08
0.90	0	0	12	12	3.67	3.67	3.67	3.33

6.8.2 Linear penalty

Numerical example 6

In this section, the effect of a supplier's ready rate performance threshold is studied for SLA contracts with a linear penalty. Except for the penalty coefficients, the parameters remain the same as the numerical example of the lump-sum penalty. The ready rate thresholds for linear penalties are $(\alpha, K) \in \{(0.60, 120), (0.70, 100), (0.80, 80), (0.90, 60)\} \subset \Theta$.

In contrast to the results for the lump-sum penalty SLA presented in Section 6.7.1, Figure 6.10 shows that for both cases (single and multiple customers) under a linear penalty with different ready rate performance thresholds, the expected total cost is unimodal. This finding is consistent with Liang and Atkins' (2013) results for a supplier with a single customer under a linear penalty SLA. In comparison to the results in Figures 6.9 for the lump-sum penalty SLA, in Figures 6.10 we found that the cost of choosing too small a base stock level S is much higher when there is a linear penalty in place. This arises because under a linear penalty, the expected cost rises based on how much the performance target is missed by, while under a lump-sum penalty, once the target has been missed any additional performance deterioration is costless. This phenomenon has been previously noted for the one-customer case by Liang and Atkins (2013).

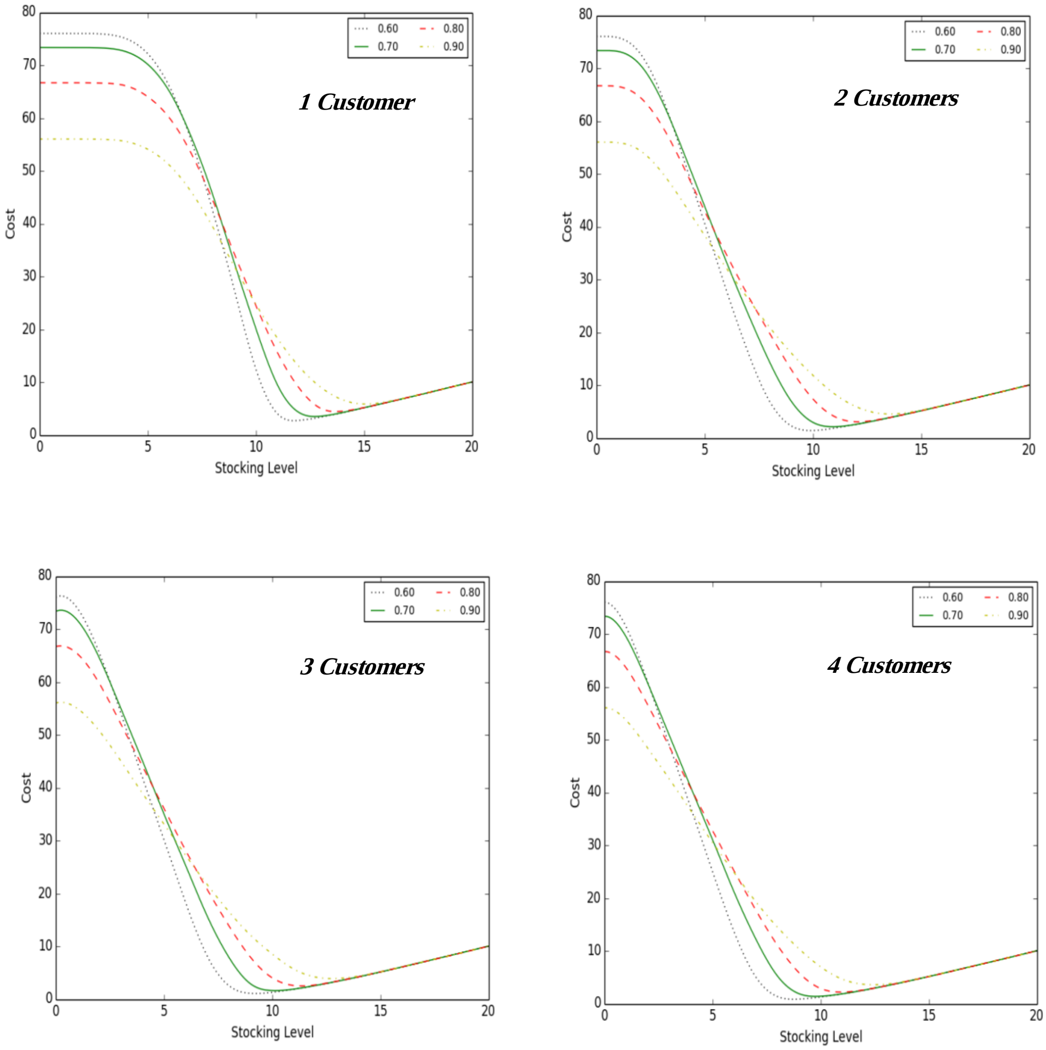


Figure 6.10: The expected total cost of linear penalty, comparison of single and multiple customers when ready rate performance α=0.60, 70, 80, and 0.90.

Table 6.6: Optimal base stock and cost for linear penalty contracts, ready rate thresholds

Ready rate thresholds	Optimal S				Expected total cost			
	One customer	Two customers	Three customers	Four customers	One customer	Two customers	Three customers	Four customers
0.60	12	10	9	9	2.71	1.41	1.06	0.87
0.70	13	11	10	10	3.51	2.12	1.64	1.40
0.80	14	12	11	11	4.49	3.05	2.55	2.22
0.90	15	14	12	12	5.88	4.63	3.89	3.63

The results in Table 6.6 show that under a linear penalty SLA it is preferable for a supplier to deal with multiple customers, as this will lead to lower expected costs and lower inventory necessary to avoid penalties.

6.9 Chapter Summary

In this chapter, we examined SLAs that employ the ready rate as the performance measure from the perspective of a supplier in both the one and multiple-customer contract cases. We examined the supplier's cost function for both lump-sum and linear penalty SLAs, and in particular, we considered the influence of several factors, such as the base stock level, the two different types of penalty, the ready rate performance threshold and the performance review phase duration. The results of this chapter provide insights that can benefit suppliers in the design and negotiation of future SLAs, and in devising strategies for realising compliance and hence avoiding higher penalties once an SLA is in place.

Our main findings for a supplier with multiple customers are as follows. First, under a lump-sum penalty contract, it is beneficial for the supplier to have a longer performance review phase. Under a linear penalty contract, the relationship between performance review phase duration and the expected total cost is not monotone, so we cannot definitively say whether the supplier will benefit from a shorter or a longer review phase duration. Second, for both penalty types (lump-sum and linear), if the optimal base stock level is implemented, having the aggregate demand split among multiple customers leads to lower expected costs. Moreover, the desired base stock level necessary for optimal performance will be lower as the number of customers is increased. Finally, under different ready rate performance thresholds, the expected total cost decreased much lower and faster in the case of multiple-customer contracts when base stock-level S increased. This is demonstrated under lump-sum and linear penalties in both Figures 6.9 and 6.10.

Chapter 6 contains the first research to analyse the SLAs with ready rate measures in a multiple-customers setting over different review phases, and different ready rate performance thresholds. Using a methodology based on a numerical study approach (for one and two customers) and Monte Carlo Simulation for more than two customers, cost results under different SLA contracts were systemically examined.

This chapter finalises the objectives of the thesis. The next chapter discusses the implications of the thesis, conclusions and possible avenues for further research.

Chapter 7: Conclusion and Future Work

This Chapter of this dissertation discusses general conclusions on meeting the objective and the findings from an integrated approach to analysing service level measures. A review of the current literature found that, while considerable attention has been paid to aspects of supply chain (SC) dynamics, there has been no integrated or concerted attempt to understand the effects on multiple customers on service levels such as: the FR, the ready rate as determined by varied performance review periods, and replenishment periods. Subsequent sections present the contributory models of the thesis, which will assist the supplier before engaging in any type of SLA contract. Recommendation and opportunities for future research are discussed in the penultimate section.

7.1 Thesis Summary

Chapter 1 explained the objectives of the research, and set out the issues of conceptualisation and nomenclature for world trade as it has evolved over the decades. The aim was to produce models of an efficient SC from the perspective of one supplier and to compare the outcomes from modelling the supplier's interactions with one or more customers. To establish the contracts for the supplier and customers, measures used, in this dissertation, were FR, ready rate and a base stock policy. The scenarios established for the simulations were for static demand across (finite horizon) replenishment review periods. This modelling approach allows for the production of a series of plots to assist in establishing performance measures for the contract. Further, two distinct customer prioritisation schemes were adopted as firms' rationing policies: FCFS and PLFR. A modelling approach adapted from Thomas (2005) was used to build the simulations.

The first objective was to model FR distributions for a supplier with a base stock policy and a single-customer, which was then extended to test for multiple customers, each with their own agreement. The customer's total demand was assumed to be constant, different across review periods to establish a target FR. The second objective was to explore the impact of lead time on the FR over finite horizons with variable review period lengths. The third objective assessed the variability of these results on consistency and asymptotic normality to determine an optimal stock level. The fourth objective was to model ready rate SLAs performance for a supplier with a single-customer, which was then extended to test for multiple customers to examine the impact of the base stock level, the type of penalty (lump-sum or linear), and differences in ready rate thresholds on the supplier's costs for multiple customers.

Chapter 2 presented the concept of FR and issues in its calculation (Nahmias & Olsen 2015; Vollmann, Berry & Whybark 1997). Johnson et al. (1995) observed that the traditional expression for line item FR should exceed 95% to avoid underestimation. Guijarro et al. (2012) approximated the FR, outperforming Johnson et al.'s (1995) estimation, while others used FR as service level measures for call centres (Hasija, Pinker & Shumsky 2008; Milner & Olsen 2008).

The review period used for FR calculation, from one replenishment to the next, was also of interest to researchers. In the standard periodic review, the interval is a single unit of time. In calculating FR real instances, the interval is $(R \geq 1)$. Banerjee and Paul (2005) and Katok, Thomas and Davis (2008) found that, in a base stock inventory system, the average FR decreased when the review period increased. Dubois et al. (2013) and Wan, Evers and Dresner (2012) used a service industry (retail) setting to estimate FR on storage capacity and product variety. Choi et al. (2004) argued that the end customer cannot be guaranteed delivery because of variables along the SC and between supplier and customer, which included inventory, storage and transport issues. This view was supported by Zhang and Sobel (2012), who advocated for shorter SCs.

The concept of finite and infinite horizons attracted more studies. Chen et al. (2003) showed that the expected value of the FR for a finite horizon is higher than that for an infinite horizon. Using volume FR and order FRs, Larsen (2011) found that volume FR was more accurate for an infinite horizon, and an order FR for a finite horizon model. Boyaci and Gallego (2001) and Shang and Song (2006) developed infinite horizon continuous review inventory models.

The incentives of reward and penalties for supplier performance were also substantially researched. Yin and Ma (2015) considered bonuses for the supplier, while Liang and Atkins (2013) found linear penalties a greater incentive than either reward or set penalties.

The second performance measure considered in this thesis, ready rate, had not been subject to the same depth of research as the FR concepts, although it was popular earlier (Feeney & Sherbrooke 1966; Rosling 2002; Schneider 1981; Silver 1972). However, Larsen and Thorstenson (2008, 2014) consistently used ready rate as a measure, as did Liang and Atkins (2013). Other recent studies included Srivathsan and Kamath (2012) and Rossetti et al. (2013). Rossetti and Xiang (2010) commented that there were issues in unpredictable demand and variable reorder quantities, so the ready rate and FR could differ markedly.

Minimising lead time can reduce high inventory levels and increase customer satisfaction. For example, a stock-out could occur (Bertazzi et al. 2013; Li & Liu 2013). Modelling lead time has engendered interest in a number of recent studies (Heydari 2014; Louly & Dolgui 2013; Priyan & Uthayakumar 2014). Spiegler et al. (2016) found that as the lead time increased, inventory orders were made higher to compensate. Muckstadt and Sapra (2010) advocated for establishing continuous reviews of lead times.

Chapter 3 focused on the first of the research objectives, exploring the FR over a number of review period durations with multiple customers. The FR methodology was established that demand was the same for a single customer, or as aggregated demand for multiple customers, and these were investigated for both service policies, FCFS and PLFR. The supplier's FR for inventory was measured over varying performance review period lengths for single and multiple customers with zero lead time. The simulation showed that in both the single and multiple-customer cases, when the performance review period increases, the average FR decreases. Further, it was found that the probability of exceeding a target FR is dependent on *T*, the length of time between performance reviews. For a shorter *T*, less stock is required for multiple customers than for a single customer. Other findings for this objective were that for a negative correlation of multiple customers' demands, a higher base stock level is required than in the case of independent customers (not correlated), while a lower inventory level is needed when the correlation between customers' demands is positive. The last finding under this objective was that a customer with higher average demand receives better service.

Chapter 4 examined objective 2: exploring the effect of varying lead times on the FR. In scenarios of one large single customer and multiple smaller customers, the chapter findings were that as lead times increase, the supplier needs more inventory for a defined service level. There is a higher probability of exceeding a fixed FR with longer lead times, and this probability is higher for multiple customers rather than one large customer, given the same aggregate demand. However, for any lead time if customer demands are correlated with high positive variance, for a supplier having one large customer is always preferable to multiple smaller customers. Further findings from this set of simulations showed that negative correlation in demand results in lower required base stock levels. Multiple customers with either negatively or positively correlated demand are more profitable for the supplier for short review periods, then beyond a threshold review period duration one larger customer is preferable. Lead times have no effect on the larger customer receiving a better service level than smaller customers.

Objective 3 concerned reliability measures for the FR, the topic of Chapter 5. This was to assess consistency and asymptotic normality to determine an optimal base stock level for single or multiple customers. The findings were that an average FR decreases with performance review duration, although variations in the FR occur due to fluctuating or seasonal demand. Reliability was established with a set of equations, then numerically tested first for one customer, then extended to two customers. The results were that the upper bounds for reliability increased with increasing performance review duration and decreased as confidence levels rose, reaching symmetry over the long-run contract FR. In summary, the optimal stock levels for reliability in the SC were established.

The last objective concerning the ready rate and using penalties to test six scenarios for a range of customers was considered in Chapter 6. For multiple customers with similar lump-sum penalty contracts, longer performance review periods were found to be more advantageous, as there was a set penalty and no further costs (Liang & Atkins 2013). Assuming a scenario with a proportional linear penalty preferred contract performance review period durations could not be calculated using the simulation, due to uncertainty (non-monotonic cost function). For both types of penalty, multiple customers were preferable to a single customer as, all variables being equal, the supplier was more likely to be able to service at least some customers from inventory, thereby being able to elect the most advantageous conditions. Further, optimal base stock levels are lower as the number of customers increase, and the penalty costs incurred are likewise lower for multiple customers.

7.2 Models Developed

The overall contribution of this dissertation is the development and comparison of numerous SLAs models for a single and multiple customer. This thesis comprises a series of contiguous studies, the contributory models appearing in the thesis are listed and described in this sub-section.

7.2.1 Fill Rate-based model

A new model of FR was developed for multiple customers over various periods. The variables were customer numbers, firm service policies, review periods and two customer demand distributions (Erlang and Normal). The impact of correlated customer demands on the finite horizon FR was analysed with aggregated mean and variance of demand held constant for one customer or more. Multiple customers were assessed for different demand distributions, with the base stock level set for a long-run FR for overall demand across all customers. Customers were serviced according to either FCFS or PLFR policies.

7.2.2 Lead time

A model to analyse the effect of lead times on the FR performance of a base stock system for one or more customers was developed. A further model investigated correlated customer demands and lead times.

The lead time model analysed the finite horizon FR and required base stock levels for one or more customers, for various review period lengths. For a long-run FR, as the lead time increases, more stock is required. However, the probability of exceeding the target FR (the PS) rises as the lead time increases. This effect is especially evident for shorter review period durations. The customer demand model addressed two situations, first where the marginal demand distribution of each correlated demand customer is unchanged; second, when in various correlations the marginal demand distribution of each customer has changed but the total variance is fixed in all cases.

7.3 Ready rate based model

A ready rate model was constructed to analyse the supplier's costs for one or more customers for both lump-sum and linear penalty contracts. Variables included base stock level, the type of penalty (lump-sum and linear) and review period duration. When three and four customers were estimated, a more complex model was required to simulate the supplier's cost function.

7.3.1 Recommendations for future research

In this section, we provide some ways that this thesis could be extended. For example, in Chapter 3, several directions for future research arose. In the multiple-customer case we considered two different policies for fulfilling customers' demands: an FCFS policy and a second policy (called PLFR throughout) based on prioritising customers based on current measured FR performance. One extension would be analysing the penalty structure in SLAs (such as lump-sum and linear penalties) when the supplier has multiple customers, and then designing an optimal service policy based on this analysis.

Chapter 4 could be extended by examining the effect of lead time together with different policies for fulfilling customers' demands, such as an FCFS policy, a second PLFR policy based on prioritising customers, and an optimal service policy based on analysis of the penalty structure. Another extension would be investigating the impact of lead time on other service level measures, such as the ready rate.

In Chapter 5, our work can extend in other directions. There are several possible extensions of the model considered in this chapter, such as including lead times and shortages, and in considering correlated demands and non-static stock levels. Other possible directions could be to investigate other demand distributions and other base stock-policies.

In Chapter 6 there are several possible directions for future research. One is to use an analytical approach to better understand and obtain formal, provable results regarding ready rate behaviour and resultant penalties in the multiple (more than two) customer case. We suspect that the difficulties attendant in such work partially explains why all analytical studies so far have been limited to the single-customer case. Another possible research topic is to expand on the present work to consider the effect of non-zero lead times, both fixed and random. In multiple-customer cases it could be worthwhile to investigate the effect of correlated demand distributions (positive and negative) on supplier costs. Finally, the design of optimal SLAs for a supplier with multiple customers would also appear to be worthwhile.

This thesis was intended to bring together concepts measuring service levels in SCs or supply networks. While organisations may be highly efficient in developing some elements of trade services, the complexity arising from the multitude of variables, performance measures and cost estimates are such that models developed to date, are at best fragmentary and inconclusive. Even with several years' work, this study is but a step forward in the evolution of firms' trading arrangements. The models and insights presented in this thesis are increasingly important due to the prevailing business environment featuring volatility of trade, mergers of firms, loss of corporate knowledge and the resultant disruption of commerce and regional economies. We trust that the contents of this study add to greater knowledge of the effects of the variables and measures for service contracts.

References

Abbasi B, Hosseinifard SZ, Alamri O, Thomas D & Minas JP. (2017), 'Finite time horizon fill rate analysis for multiple customer cases', *Omega Journal* (forthcoming).

Alamri O, B Abbasi, JP Minas & P Zeephongsekul 2017, 'Service level agreements: ready-rate analysis with lump-sum and linear penalty structures', *Journal of the Operational Research Society* (forthcoming).

Acquisition Central. 2008. Performance-Based Service Acquisition. Available at: http: //www.acquisition.gov/sevensteps/library/NIH_PBSCTemplateandSamples.pdf (accessed date January 28, 2017)

Acquisition Central. Performance-Based Service Acquisition. http://www.acqnet.gov/comp/seven˙steps/library/NIH˙ PBSCTemplateandSamples.pdf.

Aktas, E & Ulengin, F 2016, 'Penalty and reward contracts between a manufacturer and its logistics service provider', *Logistics Research*, vol. 9, no. 1, item 8.

Altendorfer, K & Minner, S 2015, 'Influence of order acceptance policies on optimal capacity investment with stochastic customer required lead times', *European Journal of Operational Research*, vol. 243, no. 2, pp. 555–565.

Alwan, LC, Xu, M, Yao, DQ & Yue, X 2016, 'The dynamic newsvendor model with correlated demand', *Decision Sciences*, vol. 47, no.1, pp. 11–30.

Anderson, ET, Gavan JF & Duncan S (2006), 'Measuring and mitigating the costs of stockouts', *Management Science*, vol. 52, no. 11, pp. 1751–1763.

Archibald, T 2007, 'Modelling replenishment and transshipment decisions in periodic review multilocation inventory systems', *Journal of the Operational Research Society*, vol. 58, no. 7, pp. 948–956.

Arıkan, E, Fichtinger, J & Ries, J 2014, 'Impact of transportation lead-time variability on the economic and environmental performance of inventory systems', *International Journal of Production Economics*, vol. 157, pp. 279–288.

Axsäter, S 2006, *Inventory control*, 2nd edn, Springer, New York NY.

Banerjee, A & Paul, A 2005, 'Average fill rate and horizon length', *Operations Research Letters*, vol. 33, no. 5, pp. 525–530.

Beamon, B. M. 1998. Supply chain design and analysis: Models and methods. *International Journal o f Production Economics, 55, 281-294*.

Ben-Daya, M, Hassini, E, Hariga, M & AlDurgam, M 2013, 'Consignment and vendor managed inventory in single-vendor multiple buyers supply chains', *International Journal of Production Research*, vol. 51, no. 5, pp. 1347–1365.

Bensaou, M. 1999, 'Portfolios of buyer-supplier relationships', *Sloan Management Review*, vol. 40, no. 4, pp. 35–44.

Berling, P & Marklund, J 2014, 'Multi-echelon inventory control: an adjusted normal demand model for implementation in practice', *International Journal of Production Research*, vol. 52, no. 11, pp. 3331–3347.

Bertazzi, L, Bosco, A, Guerriero, F & Lagana, D 2013, 'A stochastic inventory routing problem with stock-out', *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 89–107.

Bijvank, M, Huh, W, Janakiraman, G & Kang, W 2014, 'Robustness of order-up-to policies in lost-sales inventory systems', *Operations Research*, vol. 62, no. 5, pp. 1040–1047.

Blackburn J 2012, 'Valuing time in supply chains: establishing limits of time-based competition', *Journal of Operations Management*, vol. 30, no. 5, pp. 396–405.

Boyaci, T & Gallego, G 2001, 'Serial production/distribution systems under service constraints', *Manufacturing & Service Operations Management*, vol. 3, no. 1, pp. 43–50.

Brandenburg, M, Govindan, K, Sarkis, J & Seuring, S 2014, ‘Quantitative models for sustainable supply chain management: developments and directions’, *European Journal of Operational Research*, vol. 233, no. 2, pp. 299–312.

Cario, MC & Nelson, BL 1997, ‘Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix’, Tech. Rep., Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL.

Chen, F & Krass, D 2001, ‘Inventory models with minimal service level constraints’, *European Journal of Operational Research*, vol. 134, no. 1, pp. 120–140.

Chen, J, Lin, DK & Thomas, D 2003, ‘On the single item fill rate for a finite horizon’, *Operations Research Letters*, vol. 31, no. 2, pp. 119–123.

Choi, KS, Dai, JG & Song, J 2004, ‘On measuring supplier performance under vendor-managed-inventory programs in capacitated supply chains’, *Manufacturing & Service Operations Management*, vol. 6, no.1, pp. 53–72.

Christopher, M, 2016. *Logistics & supply chain management,* 5th edn, Pearson UK, Harlow UK.

Constantin, A, Ioan, A & Romania, I 2016, ‘Inventory management service level and safety stock’, viewed 23 December 2016, http://www.jopafl.com/uploads/issue9/INVENTORY_MANAGEMENT_SERVICE_LEVEL_AND_SAFETY_STOCK.pdf.

Cox, D 1983, ‘Point processes and renewal theory: a brief survey’, in J Skwirzynski (ed.), *Electronic systems effectiveness and life cycle costing,* Springer-Verlag, Berlin-Heidelberg, Germany, pp. 107–112

CVS 2013, ‘Routing guide & supplier compliance instructions’, viewed 20 August 2015, http://www.cvssuppliers.com/sites/suppliers.sub.cvsc.addventures.com/files/CVS%20Routing%20Guide%20Supplier%20Compliance%20Instructions%20-%20version%2048.pdf.

Disney, S, Gaalman, G, Hedenstierna, C & Hosoda, T 2015, 'Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand', *International Journal of Production Economics*, vol. 170B, pp. 501–512.

Dubois, T, Allaert, G & Witlox, 2013, 'Determining the fill rate for a periodic review inventory policy with capacitated replenishments, lost sales and zero lead time', *Operations Research Letters*, vol. 41, no. 6, pp. 726–729.

Dvoretzky, A, Kiefer, J & Wolfowitz, J 1956, 'Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator', *Annals of Mathematical Statistics*, vol. 33, pp. 642–669.

Ellram, L, Tate, W, & Billington, C 2004, Understanding and managing the services supply chain, *Journal of Supply Chain Management*, vol. 40, no. 4, pp. 17–32.

Feeney, G & Sherbrooke, C 1966, '($s-$ 1, s) inventory policy under compound Poisson demand', *Management Science,* vol. 12, no. 5, pp. 391–411.

Fragkos, I & De Reyck, B 2016, 'Improving the maritime transshipment operations of the Noble Group', *Interfaces,* vol. 46, no. 3, pp. 203–217.

Ganeshan, R, Jack, E, Magazine, MJ & Stephens, P 1999, 'A taxonomic review of supply chain management research', Quantitative models for supply chain management. Springer.

Ghiana, G, Laporte, G, Musamanno, R 2004, 'Introducing logistics systems', in R Webber & S Ross (eds), *Introduction to logistics systems planning and control*, John Wiley & Sons, pp. 7–15

Giard, V & Sali, M 2013, 'Optimal stock-out risk for a component mounted on several assembly lines in case of emergency supplies', *Proceedings of 7th IFAC Conference on Manufacturing Modelling, Management, and Control, International Federation of Automatic Control, St Petersburg, Russia*, IFAC Proceedings, vol. 46, no. 9, pp. 58–63.

Goetschalckx, M 2011, *Supply chain engineer*ing, Springer, New York, NY.

Gong, Q, Lai, KK & Wang, S 2008, Supply chain networks: closed Jackson network models and properties, *International Journal of Production Economics*, vol. 113, pp. 567–574.

Gruen, TW, Corsten, DS & Bharadwaj, S 2002, *Retail out-of-stocks: a worldwide examination of extent, causes and consumer responses*, Washington DC, Grocery Manufacturers of America.

Guajardo, JA, Cohen, MA, Netessine, S, Kim, SH 2012, 'Impact of performance-based contracting on product reliability: an empirical analysis', *Management Science*, vol. 58, no. 5, pp. 961–979.

Guijarro, E, Cardós, M & Babiloni, E 2012, 'On the exact calculation of the fill rate in a periodic review inventory policy under discrete demand patterns', *European Journal of Operational Research*, vol. 218, no. 2, pp. 442–447.

Gupta, A & Maranas, CD 2003, 'Managing demand uncertainty in supply chain planning', *Computers & Chemical Engineering*, vol. 27, pp. 1219–1227.

Hadley, G & Whitin, T 1963, *Analysis of inventory systems,* Prentice Hall, Englewood Cliffs, NJ.

Halldórsson, Á, Larson, P & Poist, R 2008, 'Supply chain management: a comparison of Scandinavian and American perspectives', *International Journal of Physical Distribution and Logistics Management*, vol. 38, no. 2, pp. 126–142.

Hasija, S, Pinker, E & Shumsky, R 2008, 'Call center outsourcing contracts under information asymmetry', *Management Science*, vol. 54, no. 4, pp. 793–807

Hefley, B & Loesche, E 2010, *eSourcing capability model for client organizations (eSCM-CL),* Van Haren Publishing, Zaltbommel, Netherlands:

Herbon, A, Levner, E & Cheng, TC 2014, 'Perishable inventory management with dynamic pricing using time–temperature indicators linked to automatic detecting devices', *International Journal of Production Economics*, vol. 147, pp. 605–613.

Heydari, J 2014, ‘Lead time variation control using reliable shipment equipment: an incentive scheme for supply chain coordination’, *Transportation Research Part E: Logistics and Transportation Review*, vol. 63, pp. 44–58.

Hira, D 2009, *System simulation*, Chand & Co., New Delhi, India.

Holmes, S & France, M 2002, ‘Boeing’s secret’, *Business Week*, May 20.

Ignaciuk, P 2015, ‘Discrete-time control of production-inventory systems with deteriorating stock and unreliable supplies’, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 2, pp. 338–348.

Ingalls, R 1998, ‘The value of simulation in modeling supply chains’, in *Proceedings of the 30th conference on winter simulation*, IEEE Computer Society Press, Washington, DC, pp. 1371–1376.

Ittner, CD, Larcker, DF, Nagar, V, Rajan, MV, 1999, ‘Supplier selection, monitoring practices, and firm performance’, *Journal of Accounting and Public Policy*, vol. 18, pp. 253–281.

Jaber, M, Glock, C & El Saadany, A 2013, ‘Supply chain coordination with emissions reduction incentives’, *International Journal of Production Research*, vol. 51, no. 1, pp. 69–82.

Jain, A, Rudi, N, & Wang, T 2014, ‘Demand estimation and ordering under censoring: stock-out timing is (almost) all you need’, *Operations Research*, vol. 63, no. 1, pp. 134–150.

Johnson, M & Scudder, G 1999, ‘Supporting quick response through scheduling of make- to- stock production/inventory systems’, *Decision Sciences*, vol. 30, no. 2, pp. 441–467.

Johnson, M, Lee, H, Davis, T & Hall, R 1995, ‘Expressions for item fill rates in periodic inventory systems’, *Naval Research Logistics*, vol. 42, no. 1, pp. 57–80.

Julka, N, Srinivasan, R & Karimi, I 2002, ‘Agent-based supply chain management—1: framework’, *Computers & Chemical Engineering*, vol. 26, no. 12, pp. 1755–1769.

Kanet JJ, Gorman MF &Sto¨ßlein M 2010, 'Dynamic planned safety stocks in supply networks', *International Journal of Production Research*, vol. 48, no. 22, pp. 6859– 6880.

Katok, E, Thomas, D & Davis, A 2008, 'Inventory service-level agreements as coordination mechanisms: the effect of review periods', *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 609–624.

Kendall M & Stuart A 1969, 'Standard errors', in *The advanced theory of statistics*, 3rd edn, Charles Griffin & Company, London and High Wycombe, pp. 228–245.

Kim, JJ 2016, New and extended results in renewal and queueing theories, doctoral thesis, Royal Military College of Canada, Kingston, Ontario.

Kosorok, M 2008, *Introduction to empirical processes and semiparametric inference*, Springer, New York, NY

Kouki, C, Jemaï, Z & Minner, S 2015, 'A lost sales (r, Q) inventory control model for perishables with fixed lifetime and lead time', *International Journal of Production Economics*, vol. 168, pp. 143–157.

KPMG 2010, *New on the horizon: revenue recognition for food, drink and consumer goods companies*, viewed 10 June 2014, https://www.kpmg.com/cn/en/IssuesAndInsights/ArticlesPublications/Documents/Revenue-recognition-for-food-drink-consumer-goods-O-201008.pdf.

Kroger 2015, *Standard vendor agreement*, viewed 6 September 2016, http://www.thekrogerco.com/docs/default-document-library/standard-vendor-agreement-(sva)-document-merchandising.pdf?sfvrsn=12.

Kulkarni, V 2016, *Modeling and analysis of stochastic systems*, 3rd edn, CRC Press, Boca Raton, FL.

Kumar, A & Sharman, G 1992, 'We love your product, but where is it?' *McKinsey Quarterly*, vol. 1, no. 3, pp. 24–44.

Larsen, C & Thorstenson, A 2008, ‘A comparison between the order and the volume fill rate for a base-stock inventory control system under a compound renewal demand process’, *Journal of the Operational Research Society,* vol. 59, no. 6, pp.798–804.

Larsen, C & Thorstenson, A 2014, ‘The order and volume fill rates in inventory control systems’, *International Journal of Production Economics*, vol. 147 no. A, pp. 13–19.

Larsen, C 2011, ‘Derivation of confidence intervals of service measures in a base-stock inventory control system with low-frequent demand’, *International Journal of Production Economics,* vol. 131, no. 1, pp. 69–75.

Lee, HL & Billington, C 1992, ‘Managing supply chain inventory: pitfalls and opportunities’, *Sloan Management Review*, vol. 33, no. 3, p. 65.

Lee, YH, Cho, MK, Kim, SJ & Kim, YB 2002, ‘Supply chain simulation with discrete–continuous combined modeling’, *Computers & Industrial Engineering*, vol. 43, no. 1, pp. 375–392.

Lehman, B & Weibel, S 1999, ‘Fundamental theorems of averaging for functional differential equations’, *Journal of Differential Equations*, vol. 152, no. 1, pp. 160–190.

Lewis, R & Forquer, Q 2007, ‘UPS and HP: Value creation through supply chain partnerships’, Richard Ivey School of Business, The University of Western Ontario, Case Study.

Li, C & Liu 2013, ‘A robust optimization approach to reduce the bullwhip effect of supply chains with vendor order placement lead time delays in an uncertain environment’, *Applied Mathematical Modelling*, vol. 37, no. 3, pp. 707–718.

Liang, L & Atkins, D 2013, ‘Designing service level agreements for inventory management’, *Production and Operations Management*, vol. 22, no. 5, pp. 1103–1117.

Liang, L 2009, Designing performance based contracts in supply chains, doctoral dissertation, University of British Columbia.

Louly, M & Dolgui, A 2013, ‘Optimal MRP parameters for a single item inventory with random replenishment lead time, POQ policy and service level constraint’, *International Journal of Production Economics*, vol. 143, no. 1, pp. 35–40.

Mapes, J 1993, ‘The effect of capacity limitations on safety stock’, *International Journal of Operations & Production Management,* vol. 13, no. 10, pp. 26–33.

Melnyk, S, Narasimhan, R & DeCampos, H 2014, ‘Supply chain design: issues, challenges, frameworks and solutions’.

Milner, J & Olsen, T 2008, ‘Service-level agreements in call centers: perils and prescriptions’, *Management Science*, vol. 54, no. 2, pp. 238–252.

Mirzahosseinian H & Piplani R 2011, ‘A study of repairable parts inventory system operating under performance-based contract’, *European Journal of Operational Research*, vol. 214, no. 2, pp. 256–261.

Mirzahosseinian, H & Piplani, R 2013, ‘Impact of operational systems on supplier’s response under performance-based contracts’, *Procedia Computer Science*, vol. 16, pp. 639–647.

Mostberger, P 2006, *Supply chain controlling*, Montreux, Switzerland.

Muckstadt, J & Sapra, A 2010, *Principles of inventory management, when you are down to four, order more,* Springer, New York, NY.

Nahmias, S & Olsen, T 2015, *Production and operations analysis,* 7th edn, Waveland Press, Long Grove, IL.

Neely, A, Adams, C & Kennerley, M 2002, *The performance prism: The scorecard for measuring and managing business success,* FT Prentice-Hall, London.

Nelson, B 2013, *Foundations and methods of stochastic simulation,* Springer, New York, NY.

Nebres, P., 2011. *RENEWAL THEORY AND ITS APPLICATIONS.* Vancouver.

Niaki, STA & Abbasi, B 2006, NORTA and neural networks based method to generate random vectors with arbitrary marginal distributions and correlation matrix, *Proceedings of the 17th IASTED International conference on Modeling and Simulation.* Omega, vol. 40, no. 5, pp. 511–524.

Oblicore Inc. 2007, ‘Service level management survey: results, trends and analysis (report)’, Oblicore Inc., Cambridge, MA.

Park, K & Kyung, G 2014, ‘Optimization of total inventory cost and order fill rate in a supply chain using PSO’, *International Journal of Advanced Manufacturing Technology*, vol. 70, pp. 1533-1541.

Pauls-Worm, K, Hendrix, E, Alcoba, A & Haijema, R 2016, ‘Order quantities for perishable inventory control with non-stationary demand and a fill rate constraint’, *International Journal of Production Economics*, vol. 181A, pp. 238–246.

Persson, F. and Olhager, J., 2002. Performance simulation of supply chain designs. *International journal of production economics*, 77(3), pp.231-245.

Porteous, A, Rammohan, S & Lee, HL, 2015, ‘Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties’, *Production and Operations Management*, vol. 24, no. 9, pp. 1402–1413.

Priyan, S & Uthayakumar, R 2014, ‘Optimal inventory management strategies for pharmaceutical company and hospital supply chain in a fuzzy–stochastic environment’, *Operations Research for Health Care*, vol. 3, no. 4, pp. 177–190.

Qi, Y, Ni, W & Shi, K 2015, ‘Game theoretic analysis of one manufacturer two retailer supply chain with customer market search’, *International Journal of Production Economics*, vol. 164, pp. 57–64.

Radner, R 1985, ‘Repeated principal-agent games with discounting’, *Econometrica*, vol. 53, no. 5, pp. 1173–1198.

Ren, X, Li, H & Che, M 2016, ‘Perishables’ inventory control model under time-varying and continuous demand’, *Academic Journal of Manufacturing Engineering*, vol. 14, no. 1, pp. 36–45.

Rievley, JD 2001, 'Performance-based launch services contracts: better, master, cheaper?' in *Proceeding of 38th Space Congress*, pp. 29–36.

Rite Aid 2013, *Supplier compliance guide*, viewed 12 November 2014, https://raportal.riteaid.com.

Rosling, K 2002, 'Inventory cost rate functions with nonlinear shortage costs', *Operations Research*, vol. 50, no. 6, pp. 1007–1017.

Ross, S 1970, *Applied probability models with optimization applications*, Holden-Day, San Francisco, CA.

Rossetti, M & Xiang, Y 2010, 'Simulating backlog and load building processes in a two-echelon inventory system', in B Johansson, S Jain, J Montoya-Torres, J Hugan & E Yücesan (eds), *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, MD, IEEE Press, Piscataway, NJ.

Rossetti, M, Shbool, M, Varghese, V & Pohl, E 2013 (8–11 December), 'Investigating the effect of demand aggregation on the performance of an (R, Q) inventory control policy', in R Pasupathy, SH Kim, A Tolk, R Hill & M Kuhl (eds), *Proceedings of the 2013 Winter Simulation Conference*, Washington, DC, IEEE Press Piscataway, NJ.

Rushton, A, Croucher, P & Baker, P 2014, *The handbook of logistics and distribution management: understanding the supply chain*, 5th edn, Kogan Page Publishers, London, UK.

Sandvig, J & Allaire, J 1998, Vitalizing a service parts inventory, *Production and Inventory Management Journal*, vol. 39, no. 0897-8336), pp. 67–71.

Schneider H 1981, 'Effect of service-levels on order-points or order-levels in inventory models', *International Journal of Production Research*, vol. 19, no. 6 pp. 615–631.

Schwartz, J & Rivera, D 2010, 'A process control approach to tactical inventory management in production-inventory systems', *International Journal of Production Economics*, vol. 125, pp. 111–124.

Selviaridis, K & Norrman, A 2015, ‘Performance-based contracting for advanced logistics services: challenges in its adoption, design and management’, *International Journal of Physical Distribution & Logistics Management*, vol. 45, no. 6, pp. 592–617.

Settanni, E, Newnes, LB, Thenent, NE, Parry, GC, Bumblauskas, D, Sandborn, P & Goh, YM 2016, ‘Applying forgotten lessons in field reliability data analysis to performance-based support contracts’, *Engineering Management Journal*, vol. 28, no. 1, pp. 3–13.

Shang, KH & Song, JS 2006, ‘A closed-form approximation for serial inventory systems and its application to system design’, *Manufacturing & Service Operations Management*, vol. 8, no. 4, pp. 394–406.

Shin, H, Collier, DA, Wilson, DD 2000, ‘Supply management orientation and supplier/buyer performance’, *Journal of Operations Management*, vol. 18, pp. 317–333.

Sieke, M, Seifert, R & Thonemann, U 2012, ‘Designing service level contracts for supply chain coordination’, *Production and Operations Management*, vol. 21, no. 4, pp. 698–714.

Silver, E 1972, ‘'Inventory allocation among an assembly and its repairable subassemblies’, *Naval Research Logistics (NRL)*, vol. 19, no. 2, pp. 261–280.

Song, DP, Dong, JX & Xu, J 2014, ‘Integrated inventory management and supplier base reduction in a supply chain with multiple uncertainties’, *European Journal of Operational Research*, vol. 232, no. 3, pp. 522–536.

Spiegler, V, Naim, M & Syntetos, A 2016, ‘On the analysis of lead-time disturbances in production and inventory control models’, Paper presented at the *19th International Working Seminar on Production Economics*, Innsbruck, Austria.

Srivathsan, S & Kamath, M 2012, ‘Performance modeling of a two-echelon supply chain with inventory visibility’, in G Lim & J Herrmann (eds), *Proceedings of the*

Industrial and Systems Engineering Research Conference, Orlando, FL. Institute of Industrial and Systems Engineers, Norcross, GA.

Syntetos, A, Babai, M & Gardner, E 2015, 'Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping', *Journal of Business Research*, vol. 68, no. 8, pp. 1746–1752.

Taleizadeh, A & Noori-daryan, M 2016, 'Pricing, manufacturing and inventory policies for raw material in a three-level supply chain', *International Journal of Systems Science*, vol. 47, no. 4, pp. 919–931.

Tempelmeier, H & Bantel, O 2015, 'Integrated optimization of safety stock and transportation capacity', *European Journal of Operational Research* vol. 247, no. 1, pp. 101–112.

Tempelmeier, H 2000, 'Inventory service-levels in the customer supply chain', *OR-Spektrum*, vol. 22, no. 3, pp. 361–380.

Tempelmeier, H, 2007, 'On the stochastic uncapacitated dynamic single-item lotsizing problem with service level constraints', *European Journal of Operational Research*, vol. 18, no. 1, pp. 184–194.

Thomas, D 2005, 'Measuring item fill-rate performance in a finite horizon', *Manufacturing & Service Operations Management*, vol. 7, no. 1, pp. 74–80.

Thonemann, U, Behrenbeck, K, Diederichs, R, Grosspietsch, J, Küpper, J & Leopoldseder, M 2003, 'Supply chain champions', *Financial Times Deutschland/Gabler*, Wiesbaden, Germany.

Tijms, H.C., 2003. *A first course in stochastic models*. John Wiley and sons.

Toys "R" Us, 2013, 'Vendor guide', viewed 28 August 2014, https://vendorconnect.toysrus.com.

Tzafestas, S & Kapsiotis, G 1994, ‘Coordinated control of manufacturing/supply chains using multi-level techniques’, *Computer Integrated Manufacturing Systems*, vol. 7, no. 3, pp. 206–212.

Vigoroso, M 2005, *Maximizing profitability with optimized service parts planning*, Aberdeen Group, Boston, MA.

Vollmann, T, Berry, W & Whybark, D 1997, *Manufacturing planning and control systems,* 4th edn, Irwin/McGraw-Hill, New York, NY.

Wan, X, Evers, P & Dresner, M 2012, ‘Too much of a good thing: the impact of product variety on operations and sales performance’, *Journal of Operations Management*, vol. 30, no. 4, pp. 316–324.

Wang, T, Chen, Y, & Feng, Y 2005, ‘On the time-window fulfillment rate in a single-item min-max inventory control system’, *IIE Transactions*, vol. 37, no. 7, pp. 667–680.

Wang, X & Disney, S, 2017, ‘Mitigating variance amplification under stochastic lead-time: the proportional control approach’, *European Journal of Operational Research*, vol. 256, no. 1, pp. 151–162.

Webster, J 2015, ‘Whose trolley now? Workplace liability moves up the supply chain’, *Proctor*, vol. 35, no. 4, pp. 16–18.

Yan MR, Chien, KM & Yang, TN 2016, Green component procurement collaboration for improving supply chain management in the high technology industries: a case study from the systems perspective, *Sustainability*, vol. 8, no. 2, p. 105.

Yin, Z & Ma, S 2015, ‘Incentives to improve the service level in a random yield supply chain: the role of bonus contracts’, *European Journal of Operational Research*, vol. 244, no. 3, pp. 778–791.

Zhang, C & Zhang, CH 2007, ‘Design and simulation of demand information sharing in a supply chain’, *Simulation Modelling Practice and Theory*, vol. 15, no. 1, pp. 32–46.

Zhang, D 2013, ‘An integrated production and inventory model for a whole manufacturing supply chain involving reverse logistics with finite horizon period’, *Omega*, vol. 41, no. 3, pp. 598–620.

Zhang, F 2010, ‘Procurement mechanism design in a two-echelon inventory system with price-sensitive demand’, *Manufacturing & Service Operations Management*, vol. 12, no. 4, pp. 608–626.

Zhang, J & Sobel, M 2012, ‘Interchanging fill rate constraints and backorder costs in inventory models’, *International Journal of Mathematics in Operational Research*, vol. 4, no. 4, pp. 453–472.

Zhang, Q, Sundaramoorthy, A, Grossmann, I & Pinto, J 2016, ‘A discrete-time scheduling model for continuous power-intensive process networks with various power contracts’, *Computers & Chemical Engineering*, vol. 84, pp. 382–393.

Zipkin, PH 2000, *Foundation of inventory management*, McGraw Hill, New York.