A Comparative Study of Bootstrapping Techniques for Inventory Control

by Lauren Frederick

Submitted in fulfilment of the requirements for the degree of

Magister Scientiae

to be awarded at the

Nelson Mandela University

April 2020

Supervisor: Professor Gary Sharp

Co-Supervisor: Mrs Erin Bromley-Gans

NELS N M NDELA

UNIVERSITY

DECLARATION BY CANDIDATE

NAME: Lauren Frederick

STUDENT NUMBER: 321404867

QUALIFICATION: Magister Scientiae

TITLE OF PROJECT: A Comparative Study of Bootstrapping Techniques for Inventory Control.

DECLARATION:

In accordance with Rule G5.6.3, I hereby declare that the above-mentioned treatise/ dissertation/ thesis is my own work and that it has not previously been submitted for assessment to another University or for another qualification.

SIGNATURE:

DATE: 02.12.2019

Acknowledgements

First of all I would like to thank Professor Gary Sharp for his constructive feedback, guidance and support. Furthermore I would like to thank Erin Bromley-Gans for the countless hours she dedicated towards my thesis along with the invaluable insights and advice she provided. This thesis would not have been possible without either of my supervisors.

I would also like to thank DSV South Africa for providing partial funding for my thesis, this enabled me to focus on the thesis without any financial concerns.

Finally, I would like to thank my fiancé, Tristan and mother, Michele for their unwavering support which kept me motivated throughout my studies.

Abstract

Setting correct inventory levels is an important business consideration in order to minimise inventory investment while at the same time ensuring sufficient inventory levels to meet customer demand. Inventory management has a significant impact on both financial and customer service aspects of a business. Selecting appropriate inventory levels requires that products' lead time demand be accurately estimated in order to calculate the reorder point. The purpose of this study was to empirically determine whether bootstrapping methods used to estimate the lead time demand distribution and reorder point calculation could match or even outperform a standard parametric approach. The two bootstrapping methods compared in this research included variations of those presented by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015]. These were compared to the standard parametric approach common in practice which makes use of the Normal distribution for modelling lead time demand. The three reorder point calculation methods were each incorporated into the inventory policy simulations using data supplied by a South African automotive spare parts business. The simulations covered a period of twelve months and were repeated for multiple service levels ranging from 70 to 99 percent. Results of the simulations were compared at a high level as well as for groups of items identified using segmentation techniques which considered different item demand and lead time characteristics. Key findings were that the Normal approximation method was far superior in terms of the service level metric, while the variation of the Bookbinder and Lordahl [1989] method adopted in this study presented possible cost benefits at lower service levels.

Keywords: Inventory management, Stochastic lead time, Lead time demand distribution, Bootstrap, Simulation, Automotive spare parts

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of Variance
ARIMA	Autoregressive Integrated Moving Average
B&L	Modified Bookbinder and Lordahl (1989) reorder point calculation
CSL	Cycle Service Level
CV	Coefficient of Variation
EOQ	Economic Order Quantity
ICSL	Implemented Cycle Service Level
IP	Inventory Position
LTD	Lead Time Demand
MAD	Median Absolute Deviation
R&M	Modified do Rego and de Mesquita (2015) reorder point calculation
RCSL	Realised Cycle Service Level
RFR	Realised Fill Rate
RHC	Realised Total Holding Cost
RSOH	Realised Average Daily Stock on Hand
SBA	Syntetos-Boylan Approximation
SDA	Single Demand Approach
SKU	Stock Keeping Unit
TCSL	Target Cycle Service Level

Chapter 1

Introduction

Inventory is an important business asset necessary for the fulfilment of many business processes in both industrial and commercial organisations [Nenes et al., 2010]. The practice of holding inventory is necessitated by the differences between demand and supply rates. It is therefore essential for inventory to be managed optimally in order to minimise costs/maximise profits while ensuring the highest possible level of customer satisfaction.

Optimal inventory management results in increased cashflow, reduced holding costs and reduced risk of obsolescence. This is achieved by not investing in too much stock or stocking the wrong items. Inventory management is responsible for ensuring that products are available when and where the customer wants them and as such improves customer service and revenue. Good inventory management also reduces the need to expedite orders or utilise costly air freight [Waters, 2003; Caplice, 2017]. Improvements in inventory management can therefore have significant benefits to the financial position and reputation of a business and consequently, improvement of inventory optimisation techniques is an important area of research.

The specific inventory management practices implemented in an organisation are defined by an operating doctrine also known as an inventory control policy. The inventory control policy addresses the key questions of when to replenish inventory and how much to order at each replenishment [Hadley and Whitin, 1963].

In a perfect world where both demand and supply are known, inventory management would simplify significantly. However it is very rarely, if ever, the case in practice that both aspects are known. The automotive spare parts industry is a prime example of how both supply and demand aspects of a business can experience uncertainty. This is largely due to the global nature of their supply chains and demand behaviour, which is dependent on numerous external factors. Inventory control policies account for this uncertainty by adjusting the level of inventory at which the next order for stock must be placed so as to meet a desired level of service. This is referred to as the reorder point.

A key consideration in the determination of the reorder point is the distribution of the Lead Time Demand (LTD). LTD is the demand which occurs from the time of placing an order for stock up until the time it is received and available for purchase or collection by the customer. Standard inventory control policies common in both theory and practice simply assume a Normal distribution for this variable. This is referred to as the Normal approximation. Despite its simplicity and robust nature, this assumption is a simplification of the LTD distribution, which in practical situations is rarely Normally distributed [do Rego and de Mesquita, 2015].

This research aimed to investigate the appropriateness of the Normal method for modelling LTD by comparing it to two nonparametric bootstrapping techniques, both of which have shown promising results in similar research. This was done by means of an empirical study using demand and lead time data from a South African automotive spare parts business. The company requested to remain anonymous and was thus referred to as "Company A" in this study.

There have been numerous research efforts which suggest alternative methods for estimating the LTD distribution, including both parametric and nonparametric alternatives to the Normal distribution. This study investigated two nonparametric bootstrapping approaches proposed by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015]. Bootstrapping methodologies were chosen as, with an increase in computing power, computationally intense methods such as bootstrapping have become more feasible for practical implementation. Additionally, bootstrapping methods do not require the distributional assumptions which are often not met by LTD data [Syntetos et al., 2015].

Bookbinder and Lordahl [1989] investigated one of the earliest occurrences of bootstrapping for use in inventory control. Their study presented and compared a bootstrapping procedure for the estimation of the reorder point to the Normal approximation using LTD data simulated from populations with varying shapes. The researchers found that the bootstrapping approach produced good results in situations where LTD originated from a non-standard distribution. It was based on these promising results on certain types of LTD distributions that this method was chosen for evaluation in the current study.

The current study built on the work of Bookbinder and Lordahl [1989] by implementing their bootstrapping method in an empirical study using industry data (as opposed to the simulated data used in their original study). The current study also extended the method of Bookbinder and Lordahl [1989] to include a jittering procedure, as proposed by do Rego and de Mesquita [2015]. The jittering procedure served to expand the possible values of LTD and hence the variance. This procedure was also especially beneficial for items in which the number of distinct LTD values was limited.

The research by do Rego and de Mesquita [2015] was chosen for inclusion in this study as it represented a more recent, advanced bootstrapping technique which showed promising results on automotive spare parts data. In their research, the authors assessed the performance of the bootstrapping method against numerous parametric methods by means of inventory policy simulations, utilising automotive spare parts demand data from a Brazilian auto manufacturer. In addition to the LTD distribution investigation, the authors also addressed a number of other aspects of spare parts demand forecasting and inventory control including item classification. Both the bootstrapping method and item classification methods from do Rego and de Mesquita [2015] were adopted in the present research.

The current study extended the research of do Rego and de Mesquita [2015] by using an alternative service level metric and implementing the authors' recommendation to remove outliers from the data. Additionally, an alternative nonparametric approach was taken to simulate lead time values.

The key objective of this research was to determine whether at least one of the bootstrapping approaches could outperform the Normal approximation. In this context, outperformance would mean achieving an acceptable service level at the lowest cost. In addition, this study aimed to determine whether certain methods were more suited to particular groups of items. This was done through numerous item segmentation techniques based on both demand variability and inter demand arrival times[1], demand variability and lead time characteristics. The Normal method was utilised as the comparative benchmark method, not only because of its general prevalence in industry and literature, but also because it was

[1]The inter demand arrival times were simply the number of days between subsequent demand periods.

the method utilised by the company from which the data was obtained.

Although much research has been done on the estimation of the LTD and reorder point calculations, results have been shown to be dependent on the type of demand data [Zhou and Viswanathan, 2011; Syntetos et al., 2015]. As such this research will provide valuable insight into the applicability of methods to the data of a South African auto manufacturer (Company A). Additionally, by segmenting the data according to different characteristics, this study sought to identify which specific attributes were suited to which method.

An introduction to inventory management is provided in Chapter 2, with focus on the most common inventory policies implemented in practice and their theoretical basis. Following this, Chapter 3 presents a review of previous research targeting those studies which suggested alternatives to the Normal approximation, especially bootstrapping methods, for the estimation of the LTD distribution. A short description of the data used for the purposes of this research is provided in Chapter 4 and is followed by a detailed description of the methodology carried out for the analysis in Chapter 5. Chapter 6 presents and discusses the inventory policy simulation results at a high level, including all items, then for groups of items where groups were determined by segmentation techniques accounting for demand volume, variance and frequency as well as lead time variance. Finally, Chapter 7 draws key conclusions from the results found in Chapter 6 and suggests in what situations each of the three methods would be recommended. Recommendations for future research are also included in Chapter 7.

Chapter 2

An Introduction to Inventory Management

In this chapter a thorough background to the terminology and fundamental concepts of inventory management is discussed. A number of different control policies and the assumptions which accompany them are presented. Where necessary, the methods presented here assume a parametric distribution for the LTD, specifically the Normal distribution. Extensions which make use of alternative parametric or nonparametric distributions are presented in Chapter 3.

The motives for carrying inventory may be compared to macro-economic theory motives for liquidity. This theory identifies three motives for liquidity: namely transactional, precautionary and speculative [Gudum, 2002]. Transactional motives are built on the desire to meet predictable expenditure when income may be inconsistent. In terms of inventory, this would be equivalent to holding stock to fill known demand when stock supply may be inconsistent. Precautionary motives for liquidity are those related to the ability to cover unpredictable expenditure. In terms of inventory, such expenditure would result from unpredictable inventory movements that arise from uncertainties in demand. Finally, speculative motives for liquidity are those based on opportunities for profit gain as a result of price changes or interest. In inventory management, such a profit gain could arise when inventory is purchased ahead of an expected commodity price increase [Gudum, 2002]. When considering these motives for holding inventory it is apparent that the practice of holding inventory is in itself not a goal, but rather a means of ensuring a smooth flow of goods through a supply chain, whilst containing costs and exploiting financial opportunities [Emmett and Granville, 2007].

Inventory is managed by means of an inventory control policy. This policy is described by a set of mathematical models which represent the replenishment rules that are subject to the specific characteristics and assumptions of the inventory system. The large number of possible combinations of assumptions and characteristics result in a multitude of inventory control policies each of which can vary significantly in terms of both size and complexity. It is therefore not feasible to address all possible models, and this chapter rather focuses on models derived from some of the more common inventory system characteristics and assumptions. These include: number of items, number of periods (planning horizon), nature of demand processes (deterministic or stochastic, stationary or dynamic), nature of the lead time process (deterministic or stochastic), handling of stock-outs (back-orders or lost sales), obsolescence (shelf-life considerations), and nature of the information available at any given point in time [Hadley and Whitin, 1963; Gudum, 2002].

For the determination of an inventory control policy, it is necessary to first define two key inventory related concepts. The first of these is stock type. Inventory is typically divided into five types of stock. Despite inconsistent nomenclature among sources, definitions for each stock type are, for the most part, similar. The nomenclature used throughout this study is consistent with that of Waters [2003]. The five stock types defined in this text are: cycle stock, safety stock, seasonal stock, pipeline stock, and other stock. Cycle stock is the stock on hand between orders. Safety stock is the stock held in case of emergencies, resulting from uncertainties in demand or supply. Pipeline stock is stock on order that has not yet been received. Seasonal stock is stock which enables uninterrupted operations during periods with seasonal demand variations. Finally, stock which does not fall into any of the above groups, is termed other stock [Waters, 2003].

The second inventory related concept which is key in determining an inventory control policy involves inventory costs. There are essentially four components which make up the total cost of inventory. These are: the purchasing cost, ordering cost, holding/carrying cost, and the shortage cost [Waters, 2003]. The most intuitive of these is the purchasing cost. This is simply the value paid per unit of stock ordered. The ordering cost includes all costs associated with the acquisition of stock. Examples of such costs include: wages of personnel involved in the procurement process, rental of the floor space used by the purchasing department, and the cost of receiving and inspecting incoming stock. The third component of the total inventory cost is the holding or carrying cost. As the name suggests, this consists of all costs associated with holding inventory. Holding cost can make up as much as 30 percent of the total inventory cost and can be further subdivided into: opportunity costs, insurance costs, property taxes, storage costs, as well as obsolescence and deterioration costs. Finally, shortage costs also known as stock out costs, are those cost which arise from carrying insufficient stock quantities. These costs are usually estimated and include the cost of losing customers and decreases in production as a result of insufficient stock [Narayan and Subramanian, 2009].

The following sections discuss the determination of fundamental inventory control policies which are commonly employed in practice. Alternatives and extensions to these are discussed in the review of literature in Chapter 3.

2.1 Fundamental Inventory Control Policies

As mentioned previously, inventory control policies comprise the set of models and rules used to determine the quantity and timing of orders. There are two main branches of control policies, namely continuous and periodic review policies. Each solve for the order timing and quantity using different approaches, the details of which are presented in Section 2.1.2.1.

Each control policy, whether it be periodic or continuous, operates under a set of assumptions. These assumptions represent the characteristics of the inventory system and form the foundation of each particular policy. The specific characteristics of the inventory system thus determine which assumptions may be made and consequently which control policy is appropriate. In certain instances, assumptions may be made in order to simplify the policy calculations and are therefore not true reflections of the actual system. Due to the large number of possible characteristic combinations it is infeasible to discuss all the resulting policies and their assumptions.

For the purposes of this study, the policies resulting from two key policy differentiating inventory system characteristics, namely the forms of the demand, and lead time, processes will be investigated. Each of these can present as either stochastic or deterministic, depending on the particular inventory system. The sections which follow discuss the fundamental policies which are appropriate for three different combinations of the lead time and demand processes. These are: deterministic demand and instantaneous (inherently deterministic) lead time, stochastic demand and deterministic lead time, and finally the case where both demand and lead time are stochastic.

2.1.1 Known Inventory System Variables

2.1.1.1 Deterministic Demand and Instantaneous Lead Time (Economic Order Quantity)

In the case of deterministic demand and instantaneous lead time, the continuous and periodic review policies yield the same result [Waters, 2003]. This special case policy is known as the Economic Order Quantity (EOQ).

The EOQ is the most traditional and simplistic of the control policies. Its aim is to determine a fixed order quantity by minimising inventory costs [Waters, 2003]. This order quantity is then placed when the Inventory Position (IP) is zero, also referred to as the reorder point. This is only possible when the lead time is instantaneous as this allows stock levels to be replenished as soon as the system runs out of stock. The history behind the derivation of this formula is unfortunately surrounded by much confusion. R.H. Wilson [Wilson, 1934] is frequently credited for the derivation of the EOQ model, however, it was actually formulated in 1913 when the model was first presented by Ford Whitman Harris [Erlenkotter, 1990].

The EOQ policy simplifies inventory control calculations by making some basic assumptions about the inventory system. These assumptions are:

- Constant, continuous and deterministic demand.
- All costs are known and constant.
- Shortages are not possible.
- Lead time is instantaneous (zero).
- Items are considered in isolation (independent).
- Reorder costs and purchase price are not affected by order quantity.
- Each order is delivered in full and in a single delivery.

These assumptions are somewhat unrealistic but do significantly simplify calculations. Despite this, the EOQ is still considered to be a good guideline for the determination of the optimal order quantity [Waters, 2003]. The EOQ policy and its assumptions result in an idealised stock level pattern which can be seen in Figure 2.1. This pattern represents what is referred to in this study as the IP.

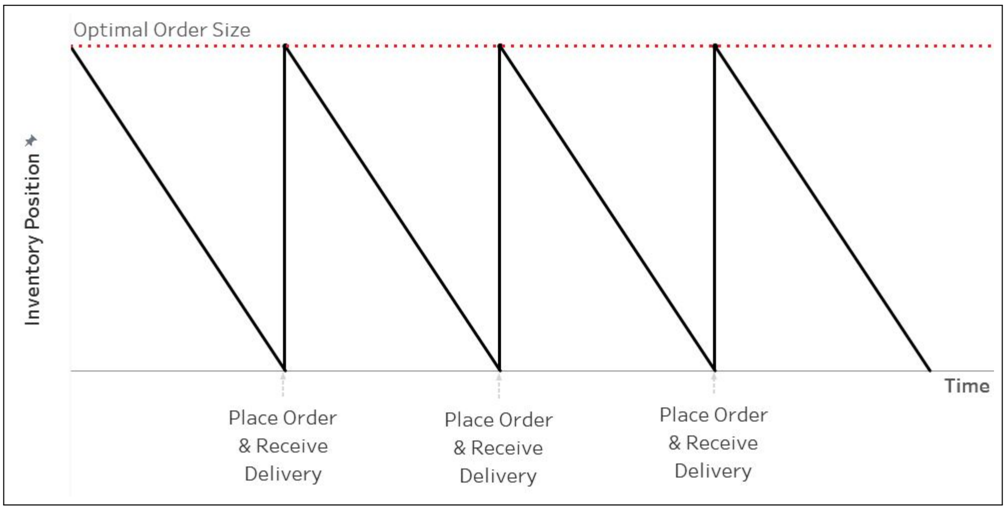

Figure 2.1: EOQ ideal stock level pattern [Waters, 2003]

Figure 2.1 is a popular representation of stock on hand, and is often referred to as the "saw tooth inventory diagram". Each spike in the stock level represents when an order is placed and received. The time between replenishments is known as the replenishment cycle time. The assumption of zero lead time under this policy results in the instantaneous receipt of goods when ordered. The diagonal lines in the diagram represent decreasing stock on hand, also referred to as cycle stock. Between replenishments, this is shown as a single straight line with a constant gradient - a result of the assumption of known, constant and continuous demand. Finally, the height of the replenishment line is determined by the optimal order quantity, as calculated from the EOQ policy [Waters, 2003].

Calculation of this order quantity is the fundamental result of the EOQ policy and any further policy parameters are derived from this quantity. The EOQ is the quantity which minimises the total costs associated with inventory. The total cost usually comprises of the unit costs, reordering costs, holding/carrying costs, and shortage costs. The details of each of these costs were discussed in the introduction to this chapter. However, under the assumptions of this policy, shortages are not possible and thus their cost is not included in the total cost. The derivation of the EOQ follows three steps. In the first step, the total cost of a single stock cycle is determined by adding the unit, ordering and holding cost components. Each cost component is calculated as follows:

$$Unit\,Cost\,Component = UC \times Q, \tag{2.1}$$

$$Reorder\,Cost\,Component = RC, \tag{2.2}$$

$$\begin{aligned} Holding\,Cost\,Component &= HC \times Average\,Stock\,Level \\ &= HC \times \left(\frac{Q}{2} \times T\right) = \frac{HC \times Q \times T}{2}. \end{aligned} \tag{2.3}$$

Thus the total cost for a single stock cycle is given by:

$$TC = UC \times Q + RC + \frac{HC \times Q \times T}{2}. \tag{2.4}$$

Where UC represents the unit cost of an item, Q is the order quantity, RC is the cost of placing an order or the reordering cost, HC is the cost of holding and item in the warehouse, and T is the replenishment cycle length, which is the length of time between consecutive orders.

The next step in the derivation determines the total cost per unit time. This is found by dividing the total cost for a single stock cycle by the cycle length, T. The resulting equation is:

$$\begin{aligned} Total\,Cost\,Per\,Unit\,Time &= TC_{Unit\,Time} \\ &= \frac{UC \times Q}{T} + \frac{RC}{T} + \frac{HC \times Q}{2}. \end{aligned} \tag{2.5}$$

One of the fundamental characteristics of the EOQ policy is that the amount of stock which enters a cycle must equal the amount of stock that leaves it. Thus Q, the order quantity, must equal $D \times T$ or $D = \frac{Q}{T}$, where D is demand per unit time. Substituting this into Equation 2.5, the total cost equation becomes:

$$TC_{Unit\,Time} = UC \times D + \frac{RC \times D}{Q} + \frac{HC \times Q}{2}. \tag{2.6}$$

In the final step, the cost per unit time is minimised by setting the derivative of the total cost per unit time, with respect to Q, equal to zero. Solving for Q yields the EOQ, which is denoted by Q_0 and is given by the following equation:

$$Q_0 = \sqrt{\frac{2 \times RC \times D}{HC}}, \tag{2.7}$$

where RC is the reorder cost, D is demand in a given period of time and HC is the holding cost for one unit of an item for a single period of time. The calculation above determines the optimal order size, fulfilling the question of how much to order. The optimal timing of orders is a function of Q_0, and can be calculated using the following equation:

$$\begin{aligned} T_0 &= \frac{Q_0}{D} \\ &= \sqrt{\frac{2 \times RC}{D \times HC}}, \end{aligned} \tag{2.8}$$

where T_0 is the optimal replenishment cycle length and determines the time between orders, RC is the reorder cost, HC is the holding cost and D the demand over a given period. From the results obtained in Equation 2.7, the optimal cost per unit time can be derived by substituting Q_0 into the total cost equation per unit time given by Equation 2.6.

When evaluating the results of the EOQ, it is only necessary to consider relevant costs, as not all costs are dependent on order size. Thus the unit cost, which is independent of order quantity under the EOQ policy assumptions, is excluded from the optimal relevant cost equation. The optimal relevant cost equation, after substituting Q_0, is given by:

$$\begin{aligned} Relevant\, Costs &= RelC_0 \\ &= HC \times Q_0 \\ &= \sqrt{2 \times RC \times HC \times D}, \end{aligned} \tag{2.9}$$

where $RelC_0$ denotes the relevant cost per unit time component of the total cost per unit time, and all other variables are as defined above. The relationship between inventory cost and the EOQ is graphically represented in Figure 2.2. As can be seen from this figure, the minimum total cost corresponds to the order quantity at the point where the holding and reordering cost curves intersect. This point corresponds to the value of the optimal order quantity, Q_0.[1]

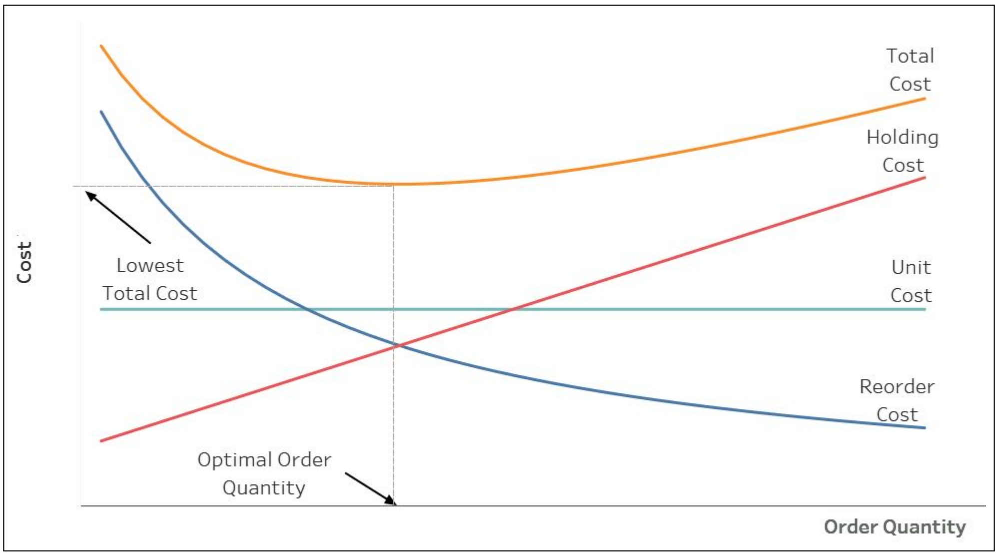

Figure 2.2: Total cost as a function of order quantity [Waters, 2003]

The popularity of the EOQ policy can be attributed to its simplicity as well as to its robust nature. This robustness allows for small adjustments of the order quantity without significant impacts on the total relevant cost. This is due to the stability of the cost curve around the EOQ.

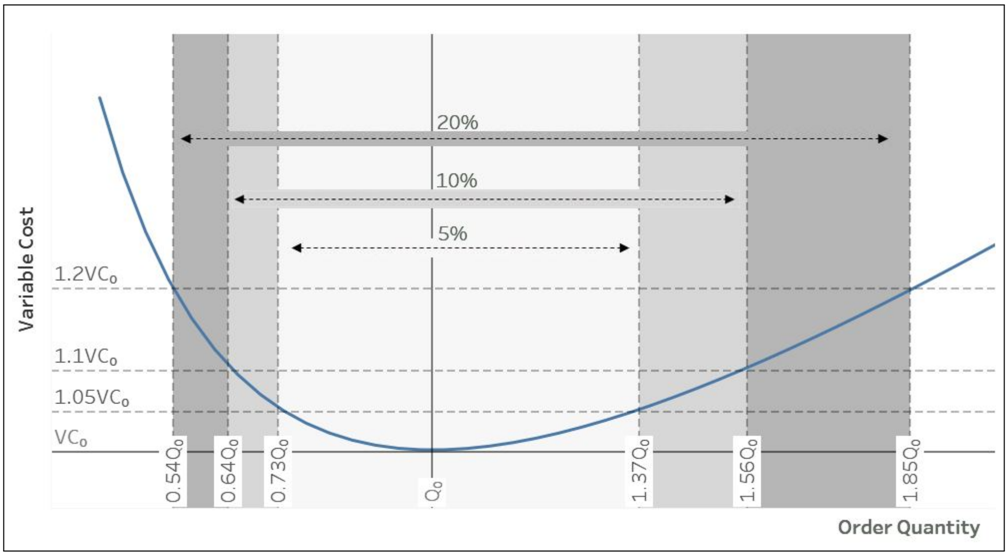

Figure 2.3: Relevant inventory cost curve around the economic order quantity [Waters, 2003]

[1]The derivation of this result is not necessary for the purposes of this research but may be found in Waters [2003].

For example, Figure 2.3 shows that an order quantity of 37 percent higher, or 27 percent lower, than the EOQ will result in a minimal cost increase of 5 percent. The stability of cost around the EOQ is especially important when used in a system which violates the deterministic demand assumption. Uncertain demand requires a forecast which in turn introduces forecast error. However, as a result of the cost stability around the EOQ, the introduction of forecast error does not necessarily have a large effect on inventory costs. Figure 2.4 illustrates the relationship between inventory cost and forecast error. The particular calculation of forecast error used here was mean percentage forecast error which is given by

$$\frac{100\%}{n} \times \sum_{t=1}^{n} \frac{Actual_t - Forecast_t}{Actual_t}.$$

This result makes it possible to use the mean demand when calculating the EOQ when demand is stochastic, thereby significantly simplifying the calculations.

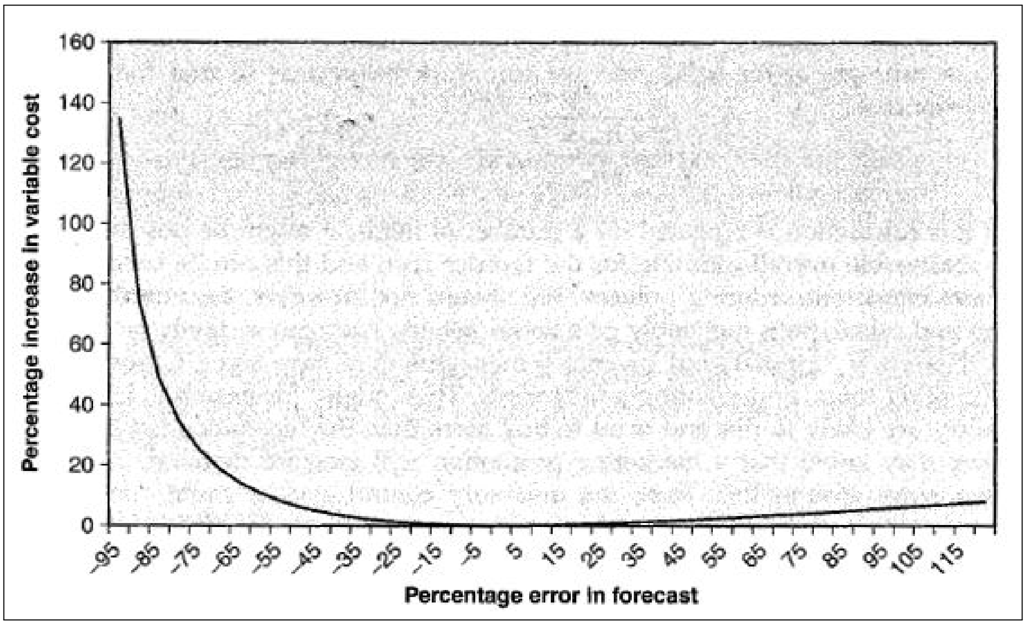

Figure 2.4: Effect of percentage forecast error on variable inventory cost [Waters, 2003]

It is evident from Figure 2.4 that largest cost variations occur when under forecasting. For example a 50 percent under estimation of demand resulted in a 6 percent increase in costs, whilst an over estimation of 50 percent only resulted in a 2 percent increase in costs.

Despite the robust nature of the EOQ, it is not always sufficient, especially when the assumptions of known constant demand and instantaneous lead time are severely violated. This necessitates the use of alternative control policies which allow for uncertainty. Under the assumptions of known constant demand and instantaneous lead time, the periodic and continuous review control policies are equivalent. However, differing recommendations for order quantity and timing are obtained for the two methods when uncertainty is introduced. Inventory policies which account for uncertainty in the inventory system will be discussed in the sections that follow.

2.1.2 Uncertain Inventory System Variables

As discussed in the previous section, the EOQ policy operates under the assumption that demand and all other inventory system variables are known. However, in practice these assumptions are unrealistic. Businesses inevitably experience some form of uncertainty caused by uncontrollable external sources. This can present in many forms, such as: price changes, new competition, supply chain disruptions, law amendments, new product availability etc. These affect four key inventory system characteristics, namely: demand, lead time, cost, and supply. The policies employed for these systems must therefore be based on the assumption of stochastic demand and lead time. Further assumptions may also need to be made for example, in the case where demand during lead time, also known as LTD, is stochastic it becomes necessary to assume that shortages are possible. The reason being that when demand is stochastic, it is possible that higher than expected demand may occur within the lead time, resulting in shortages.

Continuous and periodic review policies each account for inventory system uncertainties in their own way. The two policies are therefore no longer equivalent as they were under the assumptions of the EOQ policy. In these policies the uncertain variables have stochastic distributions and are modelled using stochastic models [Waters, 2003]. The following sections discuss the details of periodic and continuous review policy methodologies. These discussions will focus on inventory systems where demand is a stochastic process and lead time is constant, as well as where both variables form stochastic processes.

2.1.2.1 Stochastic Demand and Constant Lead Time

Continuous Review Policy

In the standard continuous review policy, order timing is determined by continuously tracking the IP, where IP is the sum of the stock on hand and the pipeline stock, less back orders. Pipeline stock, also referred to as stock on order, is stock which has been ordered but not yet received and back orders are any orders which could not be replenished from stock in a previous cycle [Narayan and Subramanian, 2009]. When the IP drops below a specific level, referred to as the reorder point, an order is triggered. The size of this order is either a pre-calculated fixed quantity or dependent on a maximum stock level. By considering all three stock components as the IP, the policy ensures that enough stock is ordered to account for back orders and prevents unnecessary stock being ordered when it may already be en route. The calculation of the fixed order quantity and reorder point is discussed in the sections below.

Two common continuous review policies are investigated in this study. They are the "order point, order quantity system" also referred to as the (s, Q) system and the "order point, order up to level system" also referred to as the (s, S) system [Caplice, 2017]. Both the (s, Q) and (s, S) policies operate on the following set of assumptions:

- Stochastic and continuous demand.
- Constant and deterministic lead time.
- All costs are known and constant.
- Shortages are possible.
- Items are considered in isolation (independent).
- Reorder costs and purchase price are not affected by order quantity.
- Each order is delivered in full and in a single delivery.
- Demand is Normally distributed with mean μ_D and variance σ_D^2.

Continuous Review Policy- The (s, Q) policy

The (s, Q) policy, illustrated in Figure 2.5, demonstrates the stock pattern which results from the policy's methodology. Over time the IP, represented by the solid black lines in Figure 2.5, decreases at a variable rate as customer demand is filled. When the IP equals or falls below the level marked s, the reorder point, an order of size Q is placed. These orders are illustrated by the dotted blue lines in Figure 2.5. The determination of the reorder point s is explained in detail below. The order quantity is always a fixed value Q, and is not affected by the size of the demand which caused the IP to drop to or below s. Thus when an order of Q is placed, the IP does not always increase to the same level but rather a varying level, calculated as $IP + Q$. An example of this can be seen when the first order is placed in Figure 2.5. In practice, the value for Q is typically estimated using the EOQ which was discussed in Section 2.1.1.1 [Caplice, 2017]. However, alternative methods have been discussed in literature which simultaneously solve for s and Q. These are not discussed in this study and may be found in Hadley and Whitin [1963].

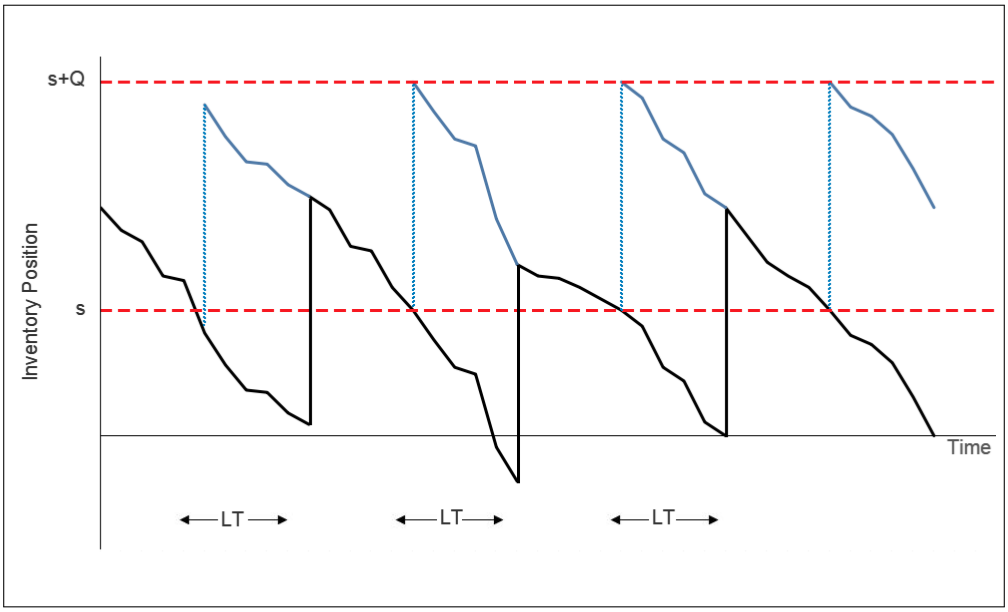

Figure 2.5: (s, Q) Continuous review policy stock levels [Caplice, 2017]

The calculation for s when Q is based on the EOQ is given by:

$$s = \mu_{LTD} + SS, \tag{2.10}$$

where μ_{LTD} is the expected LTD and SS is safety stock. Safety stock is additional inventory which is kept to allow for the continued fulfilment of customer orders when unexpected demand occurs [Waters, 2003]. In a system with non zero lead time, stock is only received lead time periods after the order is placed. By including demand which occurs during the lead time period in the reorder point calculation, customer orders may still be replenished whilst waiting for stock to arrive. If there was no uncertainty in the system, the reorder point would be equal to mean LTD and safety stock would be zero. However, when this assumption is not realistic, the reorder point is raised through the addition of safety stock. Safety stock and μ_{LTD} account for the presence of stochastic demand and positive lead time.

Inventory managers are thus faced with the complex problem of finding a level of safety stock which balances the risk of stocking out with the risk of overstocking. There are different schools of thought surrounding the choice of the best method to achieve this balance and hence multiple methods for the calculation of safety stock, which can be categorised into five groups [Silver and Peterson, 1985]. The groups are defined by whether the safety stock calculation is based on: the use of a common factor, costing of shortages, service considerations, effects of disservice on future demand, and aggregate considerations [Silver and Peterson, 1985].

This study will discuss safety stock calculations based on service considerations. In this approach to safety stock calculation, a control parameter known as the service level is introduced. This parameter constrains the safety stock calculation i.e. the safety stock value is determined such that the level of service specified by management is obtained, while costs are minimised or profits maximised. An example of a service level which may be set is that 95 percent of demand be satisfied by stock. This study will use the service level measure known as the Cycle Service Level (CSL). This is defined as a specified probability of no stock out[1] per replenishment cycle. This is equivalent to the fraction of cycles in which a stock out does not occur. The safety stock calculation corresponding to this service level measure is given by the following equation:

$$SS = k \times \sigma_{LTD}, \tag{2.11}$$

where k is the safety factor corresponding to the service level and σ_{LTD} is the standard deviation of LTD.

Equation 2.11 is essentially made up of two parts, namely the standard deviation

[1]Stock out refers to the non-availability of an item [Narayan and Subramanian, 2009].

of LTD and the service factor. To determine the values of these, the distributional assumption for demand must first be considered. It is common practice for aggregated demand, which consists of a large number of individual demands, to be assumed to have a Normal distribution with mean or expected demand μ_D and standard deviation σ_D. It should be noted that in the case of intermittent demand, it is sometimes assumed that demand is generated from a Poisson or Negative Binomial distribution. Alternative assumptions such as these are discussed in the following chapter. It follows from the properties of the Normal distribution that, when lead time is constant, LTD has a Normal distribution with mean $LT \times \mu_D$ represented by μ_{LTD} and the variance is $\sigma_D^2 \times LT$ represented by σ_{LTD}^2 [Waters, 2003]. Equation 2.11 then becomes $SS = k \times \sigma_D \times \sqrt{LT}$. To determine the second component of the SS calculation, the value of the safety factor k, the required no stock out probability or CSL must first be specified by management. The definition of CSL is the probability that LTD is smaller than or equal to the reorder point. This may be represented by $CSL = P[X \leq s]$, where X is LTD and s is the reorder point. Under the assumption of Normally distributed demand, the equation for CSL can be transformed to the unit Normal case as follows:

$$CSL = P[X \leq s] = P[\frac{(X - \mu)}{\sigma} \leq k], \tag{2.12}$$

and the value for k, also referred to as the z-score, is found using the inverse standard Normal tables. The fill rate is a popular alternative service level metric to the CSL. This metric is defined as the fraction of demand met from stock on hand/cycle stock, where cycle stock is the stock on hand between replenishments. The equation for fill rate is given by $1 - \frac{E[US]}{Q}$ where $E\left[US\right]$ is the expected units short or unfilled demand and Q is the order quantity which essentially represents the cycle stock [Caplice, 2017]. This study makes use of CSL for the determination of the reorder point and fill rate is only used as a performance measure. As a result the calculation of the reorder point using this service level metric is not detailed in this study. Further details may be found in Caplice [2017] and do Rego and de Mesquita [2015].

Once the safety stock has been determined one can calculate the reorder point using Equation 2.10. As discussed, under the assumption of Normality and fixed lead time, μ_{LTD} is given by $LT \times \mu_D$ and $SS = k \times \sigma_D \times \sqrt{LT}$. Substituting this into Equation 2.10 gives:

$$s = (LT \times \mu_D) + \left(k \times \sigma_D \times \sqrt{LT}\right), \tag{2.13}$$

where μ_{LTD} is the mean LTD, LT is lead time, μ_D is mean demand, k is the z-score corresponding to the CSL and σ_D is the standard deviation of demand. The reorder point thus determines the amount of stock required to fill customer LTD (μ_{LTD}) as well as during periods of irregular demand (SS). Placing an order at this point reduces the risk of shortages [Waters, 2003].

Continuous Review Policy - The (s, S) policy

The (s, S) policy, also referred to as the min-max policy, is an alternative to the (s, Q) policy. Both policies operate under the same assumptions but have different order quantity parameters. As with the (s, Q) policy, the (s, S) policy requires an order to be placed when the IP falls below a minimum value/reorder point, s. The reorder point is calculated in the same way for both policies, however, the order quantity requires a different calculation. The (s, S) policy does not have a fixed order quantity as in the (s, Q) policy. The order quantity is calculated as the difference between current IP and a maximum stock level, S [Caplice, 2017]. This maximum level is usually set by management and should include at least LTD and safety stock [Narayan and Subramanian, 2009]. Silver and Peterson [1985] suggest setting $S = s + Q$. The policy will therefore always replenish stock up to the maximum level denoted by S. The stock behaviour for this type of policy can be seen in Figure 2.6. The solid black lines represent the stock on hand levels, which decrease over time as customer demand is filled, and increase when ordered stock arrives. The blue lines represent the IP.

The difference in stocking behaviour between the (s, S) and (s, Q) policy can be seen when the demand, which causes the IP to drop below the reorder point, is larger than a single unit. In this case the value for IP may be a number of units below s. When the order size is fixed, as in the (s, Q) policy, the IP will not always increase to the same level when an order is placed but rather to $IP + Q$. In the case of the (s, S) policy however, the IP always returns to the level S when an order is placed as the order size is $S - IP$, and not fixed.

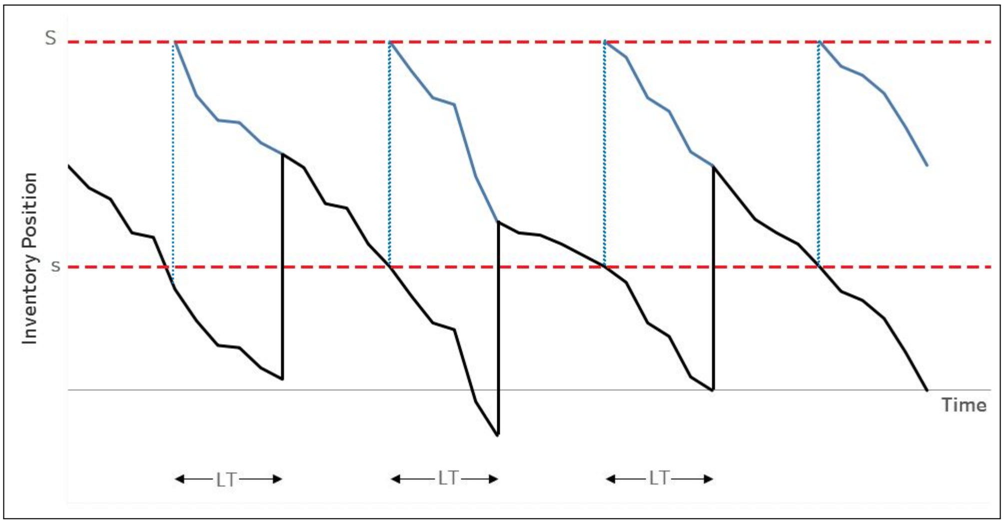

Figure 2.6: (s, S) Continuous review policy stock levels [Caplice, 2017]

Periodic Review Policy

Both periodic and continuous review policies require the review of the IP in order to determine whether to place an order. However, the timing of these reviews differs between the two types of policies. As discussed in the section on continuous review (s, Q) policy, under its methodology the order timing is fully dependent on the IP. Periodic review policies on the other hand have predetermined review intervals which dictate the order timing. As with the continuous review policy there are a number of different periodic review policy methods available. One of the commonly used periodic review policy is the "order up to level" policy also referred to as the (R, S) policy [Caplice, 2017].

Under this policy the IP is reviewed periodically every R periods. Each review is represented by a blue line in Figure 2.7. The value of R is largely a managerial decision based on convenience [Waters, 2003]. However, this decision may also be based on T_0 which is derived from the EOQ policy calculations presented in Section 2.1.1.1. Upon review of the IP every R periods, an order is placed, if $IP < S$. The size of this order is equal to the difference between the IP and a target stock level, S. The target stock level has three components: LTD, review period demand, and safety stock. This is similar to the maximum stock level in the (s, S) continuous review policy. However, with the additional constraint of periodic review timing, one must also include demand which occurs during the review period. This is an essential consideration in the periodic review policy, as during this period no orders are placed and thus inventory on hand must satisfy customer orders

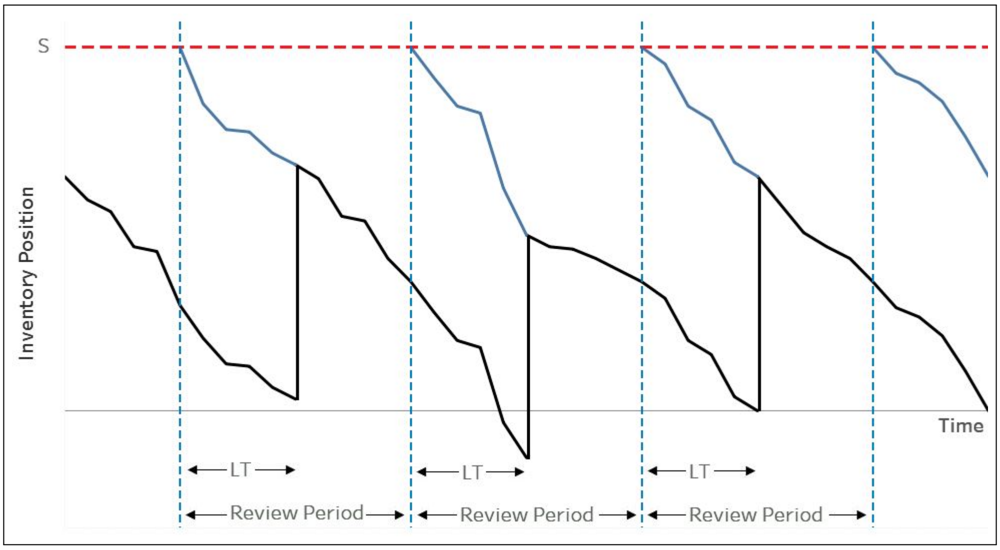

Figure 2.7: (R, S) Periodic review policy stock levels [Caplice, 2017]

[Waters, 2003; Emmett and Granville, 2007; Narayan and Subramanian, 2009; Caplice, 2017]. The formula for the calculation of S is as follows:

$$S = \mu_{LTD+R} + SS, \tag{2.14}$$

where S is the target stock level, μ_{LTD+R} is the mean demand over the lead time plus review period and SS is the safety stock. The formula for safety stock in this case is much the same as that given in Equation 2.11, with the inclusion of the review period in the calculation of the standard deviation of LTD. The resulting equation is as follows:

$$S = \mu_{LTD+R} + (k \times \sigma_{LTD+R}), \tag{2.15}$$

where S and μ_{LTD+R} are as defined for Equation 2.14, k is the z-score corresponding to the CSL and σ_{LTD+R} is the standard deviation of the demand over the lead time plus review period. As with the continuous review policy, it is assumed that demand has a Normal distribution, with mean μ_D and variance σ_D^2, and that R and LT are constant values. It follows from the properties of Normal random variables that the variation of demand over lead time and review period is simply the variance of demand multiplied by $R + LT$. Similarly, the mean demand over the lead time and review period is the mean demand multiplied by $R + LT$ [Waters, 2003; Emmett and Granville, 2007; Narayan and Subramanian, 2009; Caplice, 2017]. The standard deviation and mean are thus given by:

$$\sigma_{LTD+R} = \sigma_D \times \sqrt{R + LT}, \tag{2.16}$$

$$\mu_{LTD+R} = (R + LT) \times \mu_D, \tag{2.17}$$

where σ_D is the standard deviation of demand, μ_D is the mean demand, R is the review period and LT is the lead time [Waters, 2003; Emmett and Granville, 2007; Narayan and Subramanian, 2009; Caplice, 2017]. Substituting Equations 2.16 and 2.17 into the formula for S given in Equation 2.15, results in the following equation:

$$S = (R + LT)\,\mu_D + \left(k \times \sigma_D \times \sqrt{R + LT}\right), \tag{2.18}$$

where all parameters are defined previously.

Section 2.1.2.2 which follows, explores inventory policy extensions for variable lead time which are commonly used in practice.

2.1.2.2 Stochastic Demand and Lead Time

Thus far all control policies discussed have assumed that lead time is either instantaneous, or greater than zero, but deterministic. However in practice, as with stochastic demand, lead time is more often than not also stochastic. The variability of lead time can be the result of numerous factors including: weather delays, port congestion, supplier issues, labour strikes etc. [Vernimmen et al., 2007]. This additional uncertainty requires that further modifications be made to the control policies for stochastic demand and deterministic lead time presented in Section 2.1.2.1. The policy extensions for the case of stochastic lead time and demand are discussed below.

The most common policy extension for stochastic lead time operates under the same assumptions as those for the case where only demand is stochastic, except for the additional assumption that lead time is also stochastic. The stochastic lead time is incorporated into the policy calculations through the LTD mean and variance parameters. Under the assumption of both lead time and demand being random variables, the control policy parameters become:

$$\mu_{LTD} = E[\sum_{i=1}^{LT} D_i] \tag{2.19}$$
$$= E[D] \times E[LT]$$
$$= \mu_D \times \mu_{LT},$$

$$\sigma^2_{LTD} = V[\sum_{i=1}^{LT} D_i] \tag{2.20}$$
$$= E[LT] \times Var[D] + (E[D])^2 \times Var[LT]$$
$$= \mu_{LT} \times \sigma^2_D + \mu^2_D \times \sigma^2_{LT},$$

$$\sigma_{LTD} = \sqrt{\mu_{LT} \times \sigma^2_D + \mu^2_D \times \sigma^2_{LT}}, \tag{2.21}$$

where i represents time, LT represents the random variable lead time measured in some unit of time, D represents the random variable demand measured in units per time interval, μ_D and σ^2_D are the mean and variance of demand respectively and μ_{LT} and σ^2_{LT} are the mean and variance of lead time respectively. In Equations 2.19 and 2.20, μ_{LT} and σ^2_{LT} represent the mean and variance of the number of time periods (such as days, weeks or months) per lead time and as such are unitless multipliers in the equations [Caplice, 2017].

Under the assumption that LTD is Normally distributed these parameters can be directly substituted into the reorder point equations for the continuous review policies found in Equations 2.10 and 2.11.

In the case of the periodic review policy the parameters become:

$$\mu_{LTD+R} = \mu_D \times \mu_{LT} + R \times \mu_D, \tag{2.22}$$

$$\sigma_{LTD+R} = \sqrt{\mu_{LT} \times \sigma^2_D + R \times \sigma^2_D + \mu^2_D \times \sigma^2_{LT}}, \tag{2.23}$$

where R is the review period. These can be substituted into Equation 2.15 to find the reorder point.

This chapter presented some of the standard inventory control policies used in practice and the theory behind them. Three main groups of policies were identified based on the assumptions of the form of the demand and lead time processes. These groups included: policies in which both demand and lead time processes were deterministic, policies in which the demand process was stochastic and the lead time process was deterministic, and policies in which both demand and lead time processes were stochastic. Both periodic and continuous review policies were considered for each of these groups.

The chapter which follows presents some alternatives to the standard policies presented in this chapter. The focus will be on previous research which has considered the stochastic nature of the demand and lead time processes in inventory systems and the related policies.

Chapter 3

Literature Review

This chapter investigates alternative approaches and extensions to the standard control policies introduced in Section 2.1. It is not feasible to provide a comprehensive overview of all existing research due to its sheer volume. This literature review therefore briefly discusses some of the parametric and nonparametric alternatives to the standard control policies. And, given that the objective of this research was to compare a standard control policy with bootstrapping approaches, the main emphasis of this chapter is on research that has implemented bootstrapping methods for the purposes of the reorder point calculation. Specific attention is paid to the studies which first presented the two bootstrapping methods adopted in this study. Furthermore, due to the nature of the data used here, focus will be on past studies investigating inventory control in the spare parts industry. do Rego and de Mesquita [2015] identify five areas of research which make up spare parts inventory control research namely: item classification, demand time bucket selection, demand forecasting, LTD distribution, and parameter revision frequency. Although the primary focus of this study is the LTD distribution item classification and parameter revision frequency will also be covered in both the literature review and later chapters.

3.1 Parametric and Nonparametric Distributions for Lead Time Demand

The LTD distribution is an essential consideration in the determination of inventory control policy parameters, and is thus a popular topic in literature. Existing research covers a wide range of methods for modelling the LTD distribution including both parametric and nonparametric techniques. The most common parametric approach is to use the Normal distribution for the LTD distribution [do Rego and de Mesquita, 2015]. Due to its simplicity and availability in commercial software this has become the common approach in practice [Silver and Peterson, 1985; Bookbinder and Lordahl, 1989]. The policies presented in Section 2.1 were all based on the assumption that LTD has a Normal distribution. This is also referred to as the Normal approximation for LTD.

Despite its popularity in practice, many researchers have questioned the accuracy of this approximation. Eppen and Martin [1988] discussed potential errors and concluded that the use of this approach would likely result in stock out probability estimation errors. Chopra et al. [2004] built on this research, discussing the implicit Normal approximation assumption, which suggests that a decrease in lead time variability will result in a lower reorder point. The study showed that this was in fact not always the case and when the CSL was greater than 50 percent this assumption did not hold. This was especially evident when the Coefficient of Variation (CV) for demand was high. This conclusion was supported by findings of empirical research conducted by Tyworth and O'Neill [1997]. Although their study concluded that the Normal approximation method was robust, this only held true for demand with a low CV. The Normal approximation approach did not perform as expected and was inappropriate for demand with a high CV [Tyworth and O'Neill, 1997].

A range of alternatives to the Normal approximation for LTD have been investigated by other researchers. One such alternative investigated by a number of authors, including Gudum [2002], was to model LTD with a compound distribution. This approach assumed that in the case where both lead time and demand were stochastic and stationary with no order crossing, the variability of both could be incorporated by determining the distribution of the lead time and demand separately. These distributions were then combined to form the compound distribution for LTD. The parameters of the two constituent distributions were determined from either historical data or through forecasting. The mean and variance of the compound distribution for independent lead time and demand, under the constraints of a continuous review policy, were determined as in Equations 2.19 and 2.20 respectively (Section 2.1.2.2). The compound density function was then integrated to determine the reorder point with respect to a service level constraint

[Gudum, 2002].

Gudum [2002] provided a summary of the common combinations of theoretical distributions for lead time and demand, as well as the resulting compound distribution for LTD. Distributions for demand included: Normal, Exponential, Poisson and Negative Binomial, whilst lead time distributions included: Gamma, Exponential, Geometric and Normal. Combining different pairs of these resulted in known forms for LTD such as: Approximate Gamma, Truncated Exponential, Exponential, Hermite, Negative Binomial, Geometric, and Logarithmic Poisson Gamma.

Gudum [2002] also discussed a guide for choosing a distribution for lead time and demand. This guide considered two factors, whether demand was classified as slow or fast moving and whether supply was regular or often delayed. Regular supply lent itself to more symmetric distributions while supply which experienced delays often warranted a skewed distribution. Table 3.1 represents the appropriate choice of Gamma, Poisson and Normal distributions for lead time and demand of different forms, according to the recommendations of Gudum [2002].

Table 3.1: Demand and lead time distribution combinations [Gudum, 2002]

Type of lead time	Variable	Demand per time unit		
		Slow-moving	$\longleftrightarrow$	Fast-moving
Symmetric (regular supply)	Demand	Poisson	Gamma	Normal
	Lead time	Normal	Normal	Normal
Skewed (often delays)	Demand	Poisson	Gamma	Normal
	Lead time	Gamma	Gamma	Gamma

There are numerous other distributions for LTD which are described in the literature. These include: standard, mixture, Empirical, and Bimodal distributions. Table 3.2 provides a list of some corresponding studies for each distribution. Of the parametric distributions summarised here, the Gamma distribution for LTD showed good results, while the Poisson distribution was only suitable for a small portion of Stock Keeping Units (SKU). Thus if a single distribution was required for all items, the Poisson distribution would not model LTD well for the majority of items. Despite the many models proposed, the Normal distribution remains the most widely implemented in practice. The Normal distribution is also a common choice for comparison with alternative distributions and was applied in a num-

Table 3.2: LTD modelling methods

Modelling Method		Reference
Parametric	Negative Binomial	Hadley and Whitin [1963]; Zhou and Viswanathan [2011]; do Rego and de Mesquita [2015]
	Gamma	Nenes et al. [2010]; do Rego and de Mesquita [2015]
	Mixture distribution	Cobb [2013]; Cobb et al. [2015]
	Bimodal	Das [2013]
	Poisson	Porras and Dekker [2008]; Nenes et al. [2010]
	Compound	Croston [1972]; Ehrhardt [1979]; Krever et al. [2005]; do Rego and de Mesquita [2015]
Nonparametric	Min-max distribution free approach	Moon and Gallego [1994]; Kumar and Goswami [2015]
	Empirical distribution	Porras and Dekker [2008]
	Bootstrap	Bookbinder and Lordahl [1989]; Willemain et al. [2004]; Hua et al. [2007]; Porras and Dekker [2008]; Zhou and Viswanathan [2011]; do Rego and de Mesquita [2015]; Syntetos et al. [2015]

ber of the studies listed in Table 3.2. Therefore, due to its prevalence in both industry and literature as well as its use at Company A in particular, the Normal distribution was chosen in this study as the comparative method to bootstrapping approaches.

In practice, there are often only small samples available from which to fit the LTD distribution. This is especially true in spare parts industry data where demand is sparse and lead times are often long. The limited data thus lends itself to distribution free approaches [Bookbinder and Lordahl, 1989]. Amongst the nonparametric methods used for modelling LTD are the Min-max distribution free approach and bootstrapping. These are both popular choices in literature and a summary of the studies which utilised them is provided in Table 3.2 [Ruiz-Torres and Mahmoodi, 2010].

3.1.1 Bootstrapping Distributions for Lead Time Demand

As this investigation focused on bootstrapping, a detailed review of studies making use of this approach is provided. A decade after the introduction of bootstrapping by Bradley Efron [Efron, 1979], Bookbinder and Lordahl [1989] suggested the use of the bootstrapping technique in an inventory control setting. Their study assumed that: both lead time and demand were stochastic, LTD was stationary, there was no order crossing, and that back ordering was allowed. This nonparametric technique was used to estimate a percentile, p, of the LTD distribution which corresponded to a specified CSL. The percentile estimate replaced the classic formula shown in Equation 2.10, Section 2.1.2.1, for the reorder point. The estimate was calculated by following a procedure of sampling with replacement from a LTD sample of size n. From this single sample, a family of bootstrap samples, each of size n, was generated. The p^{th} percentile of each bootstrap sample was calculated and the expected value of these percentiles provided the bootstrap estimate for the p^{th} percentile of the LTD distribution. The reorder point was then set to this value.

Bookbinder and Lordahl [1989] compared their bootstrapping methodology to the Normal approximation of the LTD distribution. For the comparison, LTD samples were generated from data simulated from seven different distributions with varying CV and sample size which included: Uniform, Truncated Normal, Log-Normal, Two-Point and Bimodal distributions. The performance of each method was measured by comparing the percentile estimates of each approach to the true p^{th} percentiles of the seven distributions. The study concluded that when LTD had a non standard distribution, especially Bimodal, the bootstrap method outperformed the Normal approximation in terms of both cost and service level. Both methods were found to perform equally for LTD generated from standard distributions with CV greater than or equal to 1. However, the bootstrap method was not recommended when LTD had a standard shape and a CV smaller than, or equal to, 0.5. In this case, minor cost savings resulted from implementing the Normal approximation. This was due to the bootstrapping method resulting in an underestimation of the reorder point.

A further application of bootstrapping for inventory control parameter calculations was presented by Willemain et al. [2004]. This method considered only fixed lead times and accounted for three common characteristics of intermittent demand: autocorrelation, frequently repeated values, and short series. Willemain et al. [2004] suggested that autocorrelation was present between the inter demand arrival times. That is, a zero or non-zero demand occurrence was dependent on whether the previous demand occurrence was zero or non-zero. This autocorrelation was modelled by a two state, first order Markov process. The state transition probabilities were estimated from historical demand data and a forecast of the sequence of zero, and non-zero occurrences estimated over a fixed lead time. Each of the non-zero values in the sequence was replaced by a positive value, randomly sampled from the historical demand data. Due to the small sample sizes present when dealing with intermittent demand, very few unique values of demand were observed. Randomly sampling exclusively from the previously observed values was therefore unrealistic. As such Willemain et al. [2004] introduced a jittering process to be applied to the randomly sampled demand observations. The jittered demand values were calculated using the following equation:

$$Jittered\,demand = 1 + INT\left\{X^{*} + z\sqrt{X^{*}}\right\},$$

where X^* was the randomly sampled demand size, z was randomly generated from the unit Normal distribution and the INT function was used to obtain the integer component of the calculated value. In addition to this if the jittered demand value was less than or equal to zero the randomly sampled demand size X^* was used.

The forecasted demand values were then summed over the lead time to generate a single predicted value of LTD. This process was repeated one thousand times to generate the LTD distribution. This study did not go on to calculate reorder point values but rather measured the accuracy of the LTD distribution and compared this to LTD distributions generated from two alternative techniques. In each of the alternative techniques, the LTD distributions[1] were assumed to be Normal. However, the mean and variance values were calculated using different forecasting techniques, namely Croston's and exponential smoothing [Croston, 1972].

Willemain et al. [2004] compared the results of the three methods for nine different industrial datasets. The bootstrapping procedure was found to be the most accurate of the three methods. Croston's and exponential smoothing provided no significant advantages over each other. Although increases in the lead time

[1]In contrast to the traditional accuracy measures for point forecasts, Willemain et al. [2004] presented a method for measuring the accuracy of the entire distribution. The details of this measure were not required for the purposes of the current research. Refer to Willemain et al. [2004] for a detailed description of this measurement.

length resulted in decreased accuracies for the bootstrapping method, it still outperformed the two alternative methods. In addition to the clear accuracy improvements, the bootstrapping method in the study could also be extended to account for stochastic lead time, by sampling from historical lead time data, as with demand [Willemain et al., 2004].

The study did however identify some flaws with the method presented. The first concerned the jittering procedure, which was inappropriate in cases where order multiples exist. The second problem occured when demand was non-stationary (the study's bootstrapping methodology assumed that demand was stationary). This assumption was considered appropriate for the industrial data used, however Willemain et al. [2004] pointed out that for retail demand data or any data, exhibiting clear trend and seasonality, the stationarity assumption was not appropriate. A suggested solution was to apply the bootstrapping method to each different phase of the seasonal cycle.

Application of the Willemain et al. [2004] bootstrapping method can be found in empirical studies by Hua et al. [2007], Porras and Dekker [2008] and Syntetos et al. [2015]. Using data from a Netherlands oil refinery, Porras and Dekker [2008] conducted an empirical study comparing the Willemain et al. [2004] bootstrapping methodology with Empirical, Normal and Poisson distributed LTD. Results of this study showed that the Normal distribution for LTD performed best overall, while the Empirical LTD distribution unexpectedly outperformed the Willemain et al. [2004] bootstrapping method. Hua et al. [2007] proposed an extension to the Willemain et al. [2004] method. In this extension, zero and non zero demand sequences, which did not exhibit strong autocorrelation, were forecasted using logistic regression rather than a two state Markov chain. Using spare parts demand data from a Chinese petrochemical company, this method was compared with the Willemain et al. [2004] bootstrapping method and showed improved accuracy under the majority of the different lead times.

Syntetos et al. [2015] compared the results of the Willemain et al. [2004] bootstrapping method to parametric methods. These methods used simple exponential smoothing, Croston's method and the Syntetos-Boylan Approximation (SBA) method to obtain the mean and variance of the LTD distribution, which was assumed to be Negative Binomial. The authors made use of two demand datasets: one from a jewellery retailer, and another from an electronics manufacturer. They found that for the jewellery data, which were less intermittent and had shorter lead times, the bootstrapping method was slightly superior to the parametric methods. However, on the electronics data, which had longer lead times, more outliers and were more erratic, the parametric methods were superior.

Zhou and Viswanathan [2011] presented an alternative bootstrapping methodology to that of Willemain et al. [2004]. The methods were much the same and both considered fixed lead times. The key difference between the two was that the sequence of zero and non-zero demands was no longer generated using a two state Markov chain. Rather, the demand arrival sequence was generated by randomly sampling the historical observed inter demand arrival times. The Zhou and Viswanathan [2011] bootstrapping method thus required historical data for both demand size and intervals. The first step of this method was to randomly sample a value from the inter demand arrival time historical data. This value was then added to the horizon which kept track of time relative to the lead time. While the horizon was less than or equal to the lead time, a demand size value was generated by randomly sampling from the historical demand size data. These two steps were repeated until the horizon was greater than the lead time. At this point, the randomly generated demand sizes were summed over the lead time. The result was a single estimate of LTD. As in Willemain et al. [2004], these three steps were repeated one thousand times. The resulting LTD estimates were then used to generate a LTD distribution.

Unlike Willemain et al. [2004], Zhou and Viswanathan [2011] measured the performance of this methodology by calculating the averages of the: total inventory cost, inventory level, order fill rate, and stock out rate, all of which resulted from implementing an order-up-to-level policy with parameters calculated from the bootstrapped LTD distribution. These results were then compared with two parametric methods for the order-up-to-level calculation. The parametric methods used the Babai and Syntetos [2007] variation of Croston's forecasting technique for estimating the mean and variance of LTD, when LTD was Negative Binomial and Normally distributed respectively. Calculations were first done using simulated data with one thousand data points available for model fitting and another thousand for performance measurement. For the randomly generated data, the bootstrapping method had lower average inventory levels and costs than the parametric methods. Although both the bootstrapping and parametric methods achieved a service level greater than 95 percent, the parametric methods achieved lower stock out and higher fill rates. This was likely the cause of the higher costs present when using the parametric methods. The differences in performance between the methods also became less significant when the CV of the demand size or interval increased.

Zhou and Viswanathan [2011] also conducted a second set of comparisons between models, based on real industry data. This data set had significantly fewer data points than the simulated data set. This significantly impacted the performance measures, which showed that the parametric approaches outperformed the bootstrapping method. Surprisingly, the method which assumed Normally distributed LTD performed the best of the three methods. However, the bootstrapping method did still have one advantage over the parametric methods, in that it achieved service levels closer to 95 percent. Zhou and Viswanathan [2011] concluded that the result was due to the limited data and higher variance. Further research on the data limit for application of the bootstrapping method was recommended.

An alternative bootstrapping method for determining the LTD distribution, which is discussed and applied here, is that of do Rego and de Mesquita [2015]. This method was essentially a stochastic lead time extension of the Zhou and Viswanathan [2011] method, and included an alternative jittering procedure to that suggested by Willemain et al. [2004]. Unlike any of the previously discussed research efforts, every LTD prediction began with a positive demand value. This positive demand value represented the demand quantity which resulted in the replenishment trigger. As with Zhou and Viswanathan [2011], this method required historical demand data which included both demand size and intervals. The do Rego and de Mesquita [2015] method also required a lead time distribution; the one used was a triangular distribution with parameters estimated from historical lead time data.

The first step to obtain a LTD value was to randomly select a lead time from this distribution. In the second step, a demand size was randomly sampled from the historical data and the jittering process applied. Then a demand interval was randomly sampled from historical data and its value added to the horizon, which then tracked the amount of time which had passed and compared this to the randomly sampled lead time in the third step. If the horizon was smaller than, or equal to, the lead time, steps two and three were repeated until the horizon was greater than the randomly sampled lead time. Once the horizon was greater than the lead time, the jittered demand values were summed to form a single simulated value of LTD. This procedure was repeated two thousand times and the resulting LTD simulations used to generate the LTD distribution. The jittering procedure applied to the randomly sampled demand sizes was as follows:

$$Jittered\, demand = INT\left\{0.5 + X^* + z\sqrt{X^*}\right\},$$

where X^* was the randomly sampled demand size, z was randomly generated from the unit Normal distribution and the INT function was used to obtain the integer component of the calculated value. In addition to this if the jittered demand value was less than or equal to zero the randomly sampled demand size X^* was used.

do Rego and de Mesquita [2015] compared this bootstrapping method to numerous parametric methods. These included methods which assumed the LTD distribution to be Normal, Gamma and Negative Binomial, with distribution parameters calculated using either a simple moving average approach or the SBA. In addition to these, a method which considered individual order data, referred to as the Single Demand Approach (SDA), was also compared to the bootstrapping method. The SDA considers three compound LTD distributions namely: Poisson-Normal, Poisson-Gamma and Negative Binomial. The distribution parameters for these compound distributions were calculated using the latest 36 months of individual order data [do Rego and de Mesquita, 2015].

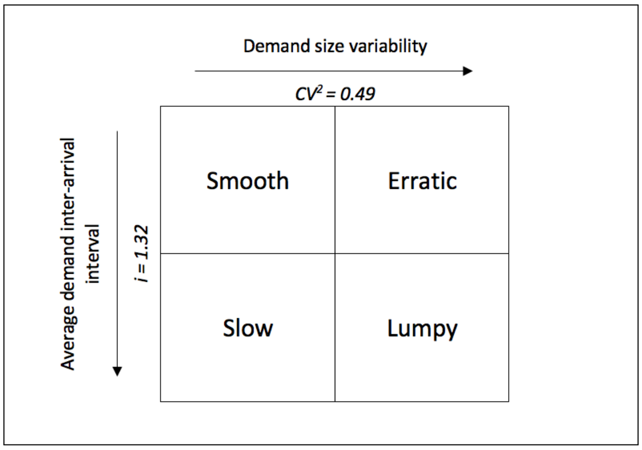

Figure 3.1: Demand classification [Syntetos et al., 2005]

In contrast to the other studies which used CSL discussed in this section, do Rego and de Mesquita [2015] used a target fill rate service level measure to determine the parameters for a continuous review (s, nQ) policy. The analysis made use of spare parts data for more than 10000 SKUs. The conclusions and recommendations of the comparison between methods were specific to the demand type of the SKU and target fill rate selected for the SKU. The demand type was classified as per the procedure presented by Syntetos et al. [2005]. In this procedure, demand was classified into four distinct groups: smooth, erratic, slow and lumpy. This was based on CV and average demand interval ($\bar{I}$). Cut off values for the groups were $CV^2 = 0.49$ and $\bar{I} = 1.32$. Figure 3.1 represents the classification of the four groups.

Based on the observed fill rate and total inventory cost incurred, do Rego and de Mesquita [2015] recommended the following:

- For all fill rates tested (i.e. 80, 90, 95, 99 percent), erratic demand is best modelled using a Gamma LTD distribution, with parameters calculated using the SBA.
- For all fill rates tested, lumpy demand is best modelled using the bootstrapping methodology.
- For fill rates 80 percent and 90 percent, smooth demand is best modelled using the bootstrapping methodology.
- For higher fill rates 95 percent and 99 percent, smooth demand is best modelled using a Gamma LTD distribution, with parameters calculated using the SBA.
- For slow demand items, the Negative Binomial LTD distribution is recommended with parameters calculated using a simple moving average for an 80 percent fill rate and the SDA for fill rates of 90 percent and 95 percent.
- For a 99 percent fill rate, slow demand items are best modelled using the bootstrapping methodology.

Their study also discussed how often the LTD distribution and its parameters should be revised. This issue was not covered by any of the other three studies discussed in this section (i.e. Willemain et al. [2004]; Hua et al. [2007]; Porras and Dekker [2008]; Zhou and Viswanathan [2011]). do Rego and de Mesquita [2015] considered both monthly and semi annual dynamic updating procedures for all of the methods compared in their analysis. The conclusion being that the LTD distribution generated and policy parameters calculated using bootstrapping should be updated semi-annually, while all other methods should undergo monthly updating of the LTD distribution and policy parameters.

do Rego and de Mesquita [2015] suggested that further research comparing the bootstrapping technique presented in their study to that of Willemain et al. [2004] and Zhou and Viswanathan [2011] should be conducted to ensure the benefits of their method.

This concludes the review of literature relevant to this current study. The current research compared the bootstrapping methods outlined by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] to a parametric method – specifically the Normal approximation method. Although many parametric approaches have been proposed in literature, the reasons for selecting the Normal approximation were threefold. Firstly, this method was the most commonly implemented one in practice. Secondly, this method was a popular choice among academic research for comparative purposes. Finally, this method was the one utilised by Company A. Also, unlike the parametric approaches discussed in this chapter, in the current study parameters of the LTD distribution were estimated using an automatic Autoregressive Integrated Moving Average (ARIMA) forecasting methodology [Hyndman et al., 2019], details of which are discussed in later chapters. This was chosen to more closely represent the automatic forecasting method employed by Company A. The bootstrapping methods evaluated in this study, namely those proposed by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015], were chosen as they represented a very early methodology in the application of bootstrapping to inventory control which was not tested on empirical data, and an advanced extension of the bootstrapping methodology developed far more recently and which showed promising results. Unlike many of the other methods discussed, the Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] methods allowed for stochastic lead times, which is an accurate representation of the uncertainty of supply experienced by Company A. The demand classification employed by do Rego and de Mesquita [2015], as defined by Syntetos et al. [2005], was also utilised in the current research to evaluate performance.

The next chapter discusses the data used for this study. Following this, Chapter 5 outlines the methodology used for the implementation and simulation of the chosen parametric and bootstrapping techniques for the reorder point calculation which were chosen for this research.

Chapter 4

The Data

This chapter introduces the datasets used for the purposes of this study and describes the data cleaning and manipulation required to render the data suitable for analysis. As previously mentioned, data were acquired from an automotive spare parts company based in South Africa. The key variables included in these datasets were historical, demand and lead time observations. The demand observations covered the period January 2013 to December 2017. As stated, each observation represented demand which was chosen over sales. Customer purchase orders were always recorded, irrespective of the stock on hand and whether an order was filled or not. This resulted in a better estimate of true demand for an item, as opposed to sales, which only recorded the customer orders which could be filled. Therefore, as sales would have distorted the true demand for an item, customer orders were instead used to represent demand. The data used to obtain the lead time observations covered the period October 2013 to April 2016.

The historical demand data consisted of daily demand observations for a total of 73670 items. The data were obtained from Company A, in the form of 60 tab delimited text files, one for each month of the period 2013 to 2017. Each text file included the fields: item code, date, and demand quantity, all of which were required for this study. In each required file the demand quantity recorded for each item code-date combination represented the month to date cumulative demand for that given item and day. However, for the purposes of the present study's analysis this was not sufficient, as single day demand quantities were required. These were calculated by subtracting the month to date demand of the previous day from the month to date demand for each day. In some cases, due to long weekends and public holidays, the cumulative values were not reset at the beginning of the month. Additional logic was thus built into the calculation to account for this. The resulting values were combined into a single dataset, which represented the daily demand for all items over the period January 2013 to December 2017. A subset of the resulting dataset for one item is provided in Table 4.1. The actual item code is not given in this table for the purpose of confidentiality.

Table 4.1: Demand data subset for item code 100xxxx

Item Code	Date	Demand Quantity
100xxxx	2013-01-03	0
100xxxx	2013-01-04	0
100xxxx	2013-01-07	1
⋮	⋮	⋮
100xxxx	2017-12-31	0

The demand dataset underwent further filtering which ensured that only items which were active (that is were available for sale) for the full five year period from January 2013 to December 2017 were included. Items which did not satisfy this criteria were identified by means of item creation date, as well as logical conditions, which were required due to a lack of information on item redundancy dates. The conditions were based on the following assumptions:

- Items with zero demand for a period of six or more consecutive months during the period 2013 to 2017 were redundant and not sold throughout the period.
- An item with fewer than 58 demand observations, whether zero or not, between the first and last positive demand occurrence was not sold throughout the full period of 2013 to 2017.

Any items with a creation date after 1 January 2015, or satisfying the conditions of the above two assumptions, were subsequently removed from the dataset. Although these conditions were somewhat strict the benefits of this approach supported its use. Firstly, the techniques applied in the analysis did not consider demand patterns for new and redundant items. The above conditions filtered out such items, therefore ensuring the chosen techniques were only applied to demand patterns of items in the mature phase of their lifecycle. Secondly, because of the computational burden of carrying out daily simulations of the three reorder point methods implemented in this study, a reduction in the number of items in turn reduced the time required for the analysis. Applying the above conditions resulted in the omission of 61845 items.

Finally, outliers in the demand history were identified and a second demand dataset was created in which outliers were replaced with the median demand to date. The identification of outliers was done through the use of the Median Absolute Deviation (MAD) method [Hampel, 1974]. This was chosen over the classic Box and Whisker [Tukey, 1977] plot fences as it provided a more robust method of identifying outliers. It thus followed that any demand observations satisfying the criteria $\frac{x_i - \widetilde{x}}{MAD} > |\pm 3|$ were considered outliers, where x_i was the individual demand observation at time i, $\widetilde{x}$ was the median demand to date and MAD was calculated using the formula $median\{|x_i - \widetilde{x}|\}$ with x_i and $\widetilde{x}$ as defined previously [Leys et al., 2013]. This second dataset was used in the analysis to compare the effect of outlier replacement on the simulation results for each reorder point calculation method.

Historical data for the second variable required for analysis in this study, lead time, were obtained in the form of supplier purchase order history from Company A. As with the demand data, these data were received in the form of tab delimited text files. Combining the supplier purchase order history files resulted in a dataset consisting of all orders placed during October 2013 to April 2016. This comprised a total of 45824 items. Each observation in the dataset represented a purchase made by Company A from their suppliers, and included both when the order was placed as well as when it was received. The time between these two events was split further by five time gates, consisting of:

1. Date the item was ordered, also known as the purchase order date.
2. Date the item was on loaded onto the vessel.
3. Date the item arrived at its final destination port.
4. Date the item arrived at the warehouse from where it was to be collected.
5. Date the item was ready for finding and extracting from the warehouse by warehouse personnel and sale to customers.

The lead time for each item's purchase order was calculated as the time between the purchase order date (1) and the date the order arrived in the warehouse (4). Although the final time gate (5) would have enabled a more accurate measure of the lead time, this field was only incorporated into the data half way through the period October 2013 to April 2016 and thus could not be used. Negative lead times (assumed to be as a result of incorrect recording) were identified and removed from the data. A subset of the final lead time dataset for a single item is presented in Table 4.2.

Table 4.2: Lead time data subset for item code 100xxxx

Purchase Order Number	Item Code	Purchase Order Date	Date Arrived in Warehouse	Lead time
3500xxxx71	100xxxx	2014-04-23	2014-06-03	41
3500xxxx06	100xxxx	2014-06-10	2014-08-12	63
3500xxxx19	100xxxx	2014-07-08	2014-08-12	35
⋮	⋮	⋮		⋮
3500xxxx65	100xxxx	2016-03-09	2016-04-08	30

The demand dataset was filtered from the resulting lead time dataset. The filtering ensured that only items present in both the demand and lead time datasets were kept. Differing from the research of do Rego and de Mesquita [2015] was the inclusion of items which had fewer than three lead time observations in the test dataset. The reason being, that in this study lead time samples were drawn from the actual lead time values whilst do Rego and de Mesquita [2015] made use of a distribution for their sampling which required a minimum of three lead time observations. The resulting datasets included records on 8056 items.

As demand data were available for the period 2013 to 2017 and lead time for the period 2013 to 2015, the period 2013 to 2015 (including observations on both variables) was chosen for the analysis. The period January 2013 to December 2014 was used for initialisation of the model parameters and was referred to as the training data. January 2015 to December 2015 was used for evaluation of model performance and referred to as the test data. January 2015 to December 2015 thus formed the period over which policy simulations were completed.

This concluded the cleaning and manipulation of the historical demand and lead time datasets required for use in the analysis. Further details of the methodology are presented in Chapter 5.

Chapter 5

Methodology

This chapter provides a detailed overview of the methodology for the analysis carried out in this study. The purpose of this analysis was to compare three reorder point calculation approaches. These were: a variation of the classical bootstrapping method proposed by Bookbinder and Lordahl [1989], a variation of an advanced bootstrapping technique advocated by do Rego and de Mesquita [2015], and the Normal approximation approach which is both common in industry and was adopted by Company A. The theoretical basis of inventory policies under the Normal assumption were discussed in detail in Chapter 2, and previous studies proposing alternative bootstrapping methodologies for calculating reorder points, including those of Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015], were discussed in Chapter 3. This chapter discusses details of the application of the three methodologies to the industry data that was provided by Company A.

Each approach was incorporated into a continuous review inventory control policy. Making use of Company A's data, a simulation of the stock on hand levels and costs was carried out for each resulting policy over a predefined simulation period. The stock on hand levels and costs obtained from the simulation of each policy were then compared in order to determine the best approach. The sections which follow include a high level overview of the simulation, a detailed discussion of each step, as well as the relevant theory for both the statistical and inventory related components of the analysis.

5.1 Analysis Overview

Three key processes governed the policy simulations performed in this study. The first of these was the overall simulation process flow. With the exception of a few modifications, this process was similar to that adopted by do Rego and de Mesquita [2015]. The two remaining processes, referred to as the Tactical and Operational management processes, formed sub-procedures of this overall simulation flow. The purpose of the Tactical management process was to determine the reorder point, s, making use of one of the reorder point approaches chosen for comparison in this study. The resulting reorder point was then used as an input parameter for the Operational management process. The inventory control policy simulation was performed from this process. The remainder of this section discusses the overall process flow and the following sections provide an in-depth explanation of the Tactical and Operational management processes respectively.

5.1.1 Overall Simulation Process Flow

Figure 5.1 illustrates the modified do Rego and de Mesquita [2015] overall simulation process flow. Processes are represented by rectangles and decisions by diamonds.

The aim of the first process (see Figure 5.1) was to define the policy version. This was achieved by specifying the service level and reorder point calculation technique. A total of 27 policy versions were simulated in this study. Each policy differed in its calculation of the reorder point, CSL and whether or not the demand history, used for the calculation of policy parameters, included outliers or not. Details of the methodology used for the detection of outliers in the demand history were discussed in Chapter 4. As previously defined in Section 2.1.2.1, CSL is the probability that LTD is smaller than or equal to the reorder point. At the time of acquiring data from Company A, the business made use of CSL as its service level metric and aimed to achieved a CSL of 95 percent for all products. For this reason the key CSL considered in this study was 95 percent. However, in an attempt to thoroughly compare all reorder point calculation methods, additional CSLs were also considered, ranging from 70 to 99 percent depending on the reorder point method.

Thereafter the Tactical management process was undertaken. As previously men-

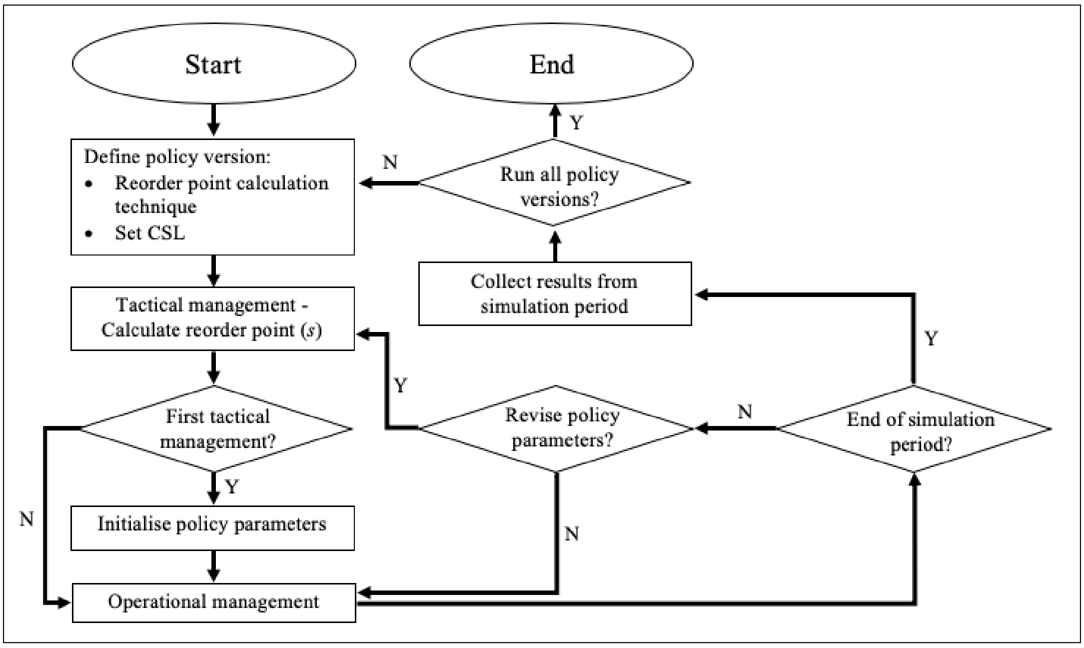

Figure 5.1: Modified do Rego and de Mesquita [2015] Overall simulation process flow

tioned, this process comprised the calculation of the reorder point as specified in the previous process. The resultant reorder point was subsequently used as input for the Initialise policy parameters and Operational management processes which followed.

After the calculation of the reorder point, a decision task determined whether the policy parameters required for the Operational management step needed initialising or not. Initialising the policy parameters included:

- Calculating the order quantity, Q.
- Setting customer and supplier backorders to zero.
- Setting the starting IP, to $s + Q$, where s was the reorder point and Q the order quantity.

Parameter initialisation was only required after the initial calculation of the reorder point. This step was therefore excluded from any repetitions of the Tactical management process.

The initialised policy parameters were then applied as inputs for the Operational management process. This process was essentially where the simulation of the inventory policy, defined in the first process, took place. Each policy took the form of a continuous review policy denoted by (s, nQ) differing only in its calculation of s which was dependent on the reorder point method. The (s, nQ) policy, also adopted by do Rego and de Mesquita [2015], was a variation of the (s, Q) policy (see Section 2.1.2) further details of which are discussed in Section 5.1.4. The stock on hand levels, which depended on demand and incoming stock orders, were simulated for each day of the simulation period. Each day's results were recorded and a continuously updating horizon was incremented by a single day.

The policy simulation ran from 1 January 2015 to 31 December 2015. At the beginning of each "day" of the policy simulation a decision was required which determined one of three outcomes. The first possible outcome was that the simulation proceeded with a reiteration of the Operational management process. The second outcome required a revision of the policy parameters , thereby temporarily terminating the Operational management process and reverting to the Tactical management process. In this step the reorder point and order quantity were recalculated using updated demand and lead time data. These data had been revised during the last calculation of said parameters. The third and final outcome was complete termination of the simulation. Result collection and further policy version considerations followed.

The s and Q parameters were revised monthly for the purposes of this study. do Rego and de Mesquita [2015] identified small cost reductions when reorder point calculations based on bootstrapping parameters were revised semi-annually. Given the small magnitude of these cost reductions it was decided to rather implement monthly parameter revisions. This also allowed the occurrence of non-stationary demand to be accounted for to an extent.

This process was repeated for each of the three inventory policies compared in this study. Further details of the reorder point calculations and inventory policy are discussed in the following sections.

5.1.2 Tactical Management - Reorder Point Estimation

Three reorder point calculation approaches were selected for comparison in this study. These consisted of the Normal approximation technique introduced in Section 2.1.2, a variation of a classical bootstrapping method described by Bookbinder and Lordahl [1989] and a variation of a modified bootstrapping method presented by do Rego and de Mesquita [2015]. Each approach was guided by the process flow chart seen in Figure 5.2.

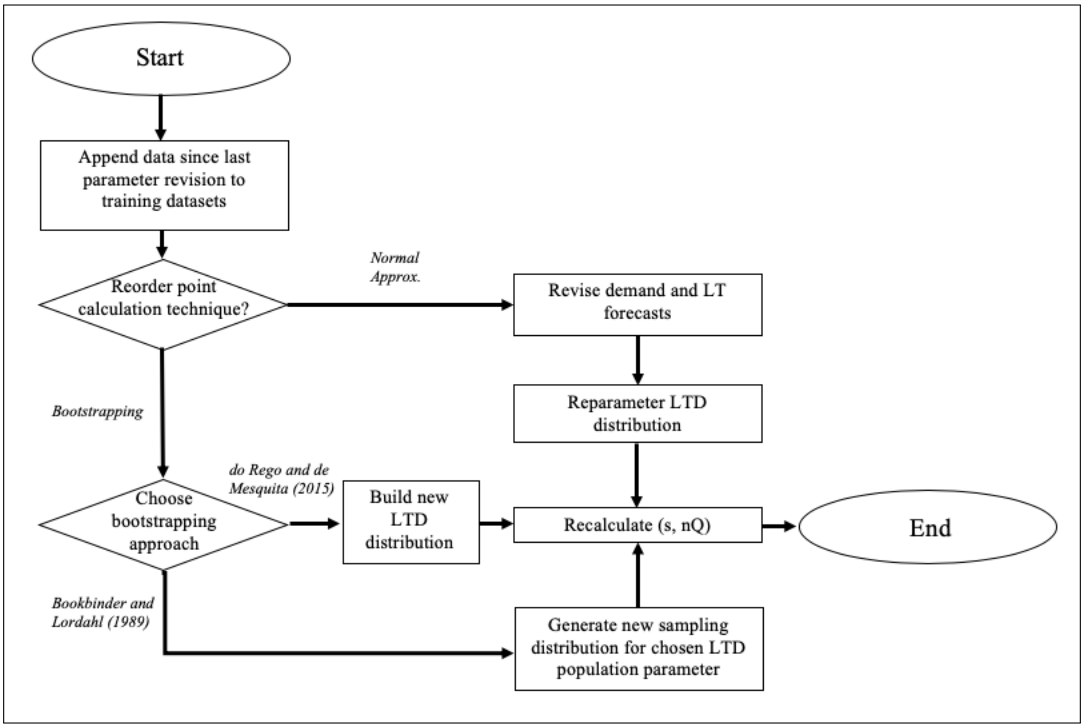

Figure 5.2: Tactical simulation flow chart

As seen in Figure 5.1, the Tactical management step was repeated multiple times throughout the simulation. The purpose of this was, on its first iteration, to calculate the initial policy parameters, s and nQ. Every repetition thereafter revised parameters incorporating the additional lead time and demand observations which had occurred since the previous parameter revision had taken place. Appending these new data to the previous training data sets resulted in an updated training data set which was then used for the parameter estimation.

The Tactical management process iterated with the Operational management process but was only executed on the first iteration of the Operational management processes and at the beginning of every new simulation month thereafter. The first iteration of the Tactical management process made use of demand and lead time data for the period January 2013 to December 2014. This was referred to as the training dataset.

The second step in the Tactical management process varied according to the chosen calculation method of the reorder point. Either the Normal approximation or a bootstrapping method was chosen.

As mentioned above, two bootstrapping methods were employed in this study. Therefore when bootstrapping was chosen as the reorder point calculation method either the modified Bookbinder and Lordahl [1989] or the modified do Rego and de Mesquita [2015] approach was adopted. The modified do Rego and de Mesquita [2015] approach made use of bootstrapping to build an updated LTD distribution from which the reorder point was determined. The Bookbinder and Lordahl [1989] approach in contrast made use of bootstrapping in a more classical sense and a sampling distribution of a LTD population parameter, chosen to represent the reorder point, was generated.

If the Normal approximation was chosen as the reorder point calculation technique, the first step in the method was to revise both the demand and lead time forecasts. Each updated forecast was then used to re-parameterise the LTD distribution and to calculate the reorder point.

Details of each of the three reorder point calculation approaches are discussed in Sections 5.1.2.1 and 5.1.2.2. In addition to this, a description of the methodology used to estimate the second inventory policy parameter, order quantity, is provided in Section 5.1.3.

5.1.2.1 Normal Approximation Approach

The decision to include the Normal approximation approach as part of this study was based on its prevalence in industry and academic research, as well as its application at Company A. Through the use of industry data, this study aimed to determine which of the three approaches was superior in a practical environment. It was thus a natural choice to include an approach commonly used in practice to supply a meaningful comparison for the proposed alternatives. The details of the methodology of the Normal approximation approach are discussed below.

The theory presented in Section 2.1.2 showed that the first step in estimating the reorder point under the Normal approximation approach was to estimate the mean and variance of the LTD. Together, these parameters represented the shape and spread of the Normal distribution used to approximate the LTD distribution. The formulae for μ_{LTD} and σ_{LTD} were given in Section 2.1.2.2 by Equations 2.19 and 2.21 respectively.

As per Equation 2.19, calculating μ_{LTD} required the multiplication of two terms, mean demand and mean lead time. Each of these was first considered on its own and then combined with the other to form the desired estimate of μ_{LTD}. It was not always necessary to consider the sequence of their calculation. However, the method chosen for the estimation of mean demand in this study required that the mean demand was calculated over the mean lead time and hence mean lead time was calculated first. Before any estimation took place the units of measure for both mean demand and lead time were defined. This ensured the logical multiplication of the two to form the estimate of μ_{LTD}. The unit of measurement for mean lead time was defined as days, therefore mean demand was measured in units per day.

The estimation of both the mean lead time and mean demand were achieved by means of different forecasting models. For the estimation of μ_{LT}, a simple moving average was applied to historical lead time observations.

Estimation of the mean demand followed a more complex forecasting solution, utilising automatic ARIMA forecasting models [Hyndman et al., 2019]. Automatic forecasting model selection is common practice in industry and formed the methodology basis for mean demand forecasting employed by Company A. These forecasts were not generated on each iteration of the Tactical management process as with the mean lead time estimates. Rather, a bulk forecast was generated on the first iteration for all required months. The bulk forecast included a twelve month period projected forecast for every month in the full demand data set. The twelve month period forecast generated for each month utilised all the demand history up until that specific month, and as such these bulk forecasts were equivalent to the forecasts which would have been obtained during the Tactical management process iterations.

Pre-calculating a bulk forecast reduced the time required to implement the Tactical management process. On each repetition of the Tactical management process the forecasts could simply be accessed from a predefined forecast file, as opposed to requiring recalculation. The purpose of generating a forecast for months not in the simulation period enabled calculations of historical mean squared error values, which were necessary for the estimation of σ_{LTD} (discussed later in this section).

The first step in generating the demand forecasts was to aggregate the historical daily demand data into monthly intervals. The forecast was carried out in monthly intervals to match the approach of Company A, which forecasted in monthly buckets (or bins) in order to reduce forecast data variability. A forecast was then generated at the start of each month in the historical data using all monthly demand observations prior to the start of the said month. These data were then used as input for the auto.arima function in R to generate the forecast [R Core Team, 2018]. The auto.arima function forms part of the forecast package in R.[1] The output from the auto.arima forecast was a twelve month projected monthly demand forecast. A twelve month forecast is common in practice as the long period ensures that a forecast is always available in case of unexpectedly long lead times. The process of generating a forecast described above was repeated for every month in the historical data. Combining all the forecasts formed the bulk forecast from which forecasts were accessed on each iteration of the Tactical management step.

The monthly demand forecasts relevant to the iteration of the Tactical management process were obtained from the forecast file. They were then divided by the number of days in the forecast month in order to obtain daily forecasts for each item. The daily forecast values were then summed over a period equal to that of the mean lead time estimate described above and hence formed the estimate of μ_{LTD}.

Although Company A did not operate on weekends and public holidays, it was decided to include daily forecasts for these days as described in the calculation for μ_{LTD} above. The reason being that both historical and forecast values of lead times included these non-business days. The Operational management process which executed the policy simulation therefore included simulations for non-business days. However, due to the zero demand occurrences on non-business days there was no potential of order suggestions occurring during the simulation. This was essential for a realistic simulation as order placements were not possible on weekends for Company A.

The second parameter estimate required for the LTD distribution was the standard deviation, σ_{LTD}. The calculation of this estimate made use of both the demand and lead time forecasts as described above. The equation for σ_{LTD} was given in Section 2.1.2.2 by Equation 2.21. Substituting the lead time and demand forecast values into the equation resulted in the function:

$$\begin{aligned} \sigma_{LTD} &= \sqrt{\mu_{LT} \times \sigma_D^2 + \mu_D^2 \times \sigma_{LT}^2} \\ &= \sqrt{\overline{x}_{LT} \times MSE + \overline{x}_D^2 \times s_{LT}^2}, \end{aligned} \tag{5.1}$$

where $\overline{x}_{LT}$ was the forecast for lead time (described above), s_{LT}^2 was the variance

[1]Details of the automated model selection process are not discussed in this study. However, information regarding this topic may be found in Hyndman et al. [2019].

of all lead time observations prior to the day of the simulation on which the current iteration of the Tactical management process was performed, $\overline{x}_D$ was the average daily demand calculated as the average of the daily demand forecasts taken over a period equal to $\overline{x}_{LT}$ as described above, starting from the current day of the simulation, and MSE was the measure of demand forecast error for the twelve months prior to the current simulation day. The equation used for the calculation of the MSE was $MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2$, where n was the number of periods, t was the forecast and demand period Y_t, was the actual observed demand at time t, and F_t was the forecasted value of demand for time t generated in the months zero, one or two prior to the observed demand. This was referred to as the forecast horizon. The relevant forecast horizon was selected for the MSE calculation based on the logic:

$$if \begin{cases} \overline{x}_D \leq 30 & Horizon = 0 \\ 30 < \overline{x}_D \leq 60 & Horizon = 1 \\ x_D > 60 & Horizon = 2. \end{cases}$$

Once both estimates for the mean LTD and variance of LTD were known, the reorder point was calculated by substituting these values into Equations 2.10 and 2.11 in Section 2.1.2.1 and determining the value for k from the specified CSL and the unit Normal tables.

5.1.2.2 Bootstrapping Approach

The two remaining approaches compared in this study offered nonparametric alternatives to the Normal approximation adopted in the previous section. These approaches adopted bootstrapping techniques.

The first of the two bootstrapping approaches was a variation of the Bookbinder and Lordahl [1989] reorder point calculation methodology. The second was a variation of the do Rego and de Mesquita [2015] method for finding the reorder point. Both were briefly introduced in Chapter 3. Although both methods resulted in an estimate of a population percentile of interest of the LTD distribution, they differed in their application of bootstrapping to attain this estimate.

The variation of the Bookbinder and Lordahl [1989] approach adopted in this study followed the steps shown in Table 5.1. One of the fundamental differences between

Table 5.1: Modified Bookbinder and Lordahl [1989] reorder point calculation

Step	Description
1	Obtain historical demand data.
2	Obtain historical lead time data.
3	Calculate historical LTD observations for each historical lead time. Demand within each lead time was summed to obtain the LTD.
4	Randomly sample with replacement from the historical LTD observations.
5	Apply the jittering process described in Table 5.2 to each sampled observation.
6	Repeat steps 4 and 5 until a sample the same size as the original sample has been generated.
7	Repeat steps 4 to 6 1000 times.
8	Calculate the percentile corresponding to the chosen CSL for each of the 1000 samples.
9	Calculate the average of the 1000 percentiles to form a single percentile estimate of the LTD sampling population.

the approach adopted in this study and that of Bookbinder and Lordahl [1989] was the data source. In the present study industry observed data were used in all three reorder point methodologies in order to formulate inventory control policy parameters and simulate results. Bookbinder and Lordahl [1989], in contrast used LTD observations randomly sampled from predefined distributions. Additionally, in the present study a jittering process based on the one proposed by do Rego and de Mesquita [2015] was added to the LTD sampling process.

In keeping with Bookbinder and Lordahl [1989], the historical LTD observations were calculated rather than keeping the lead time and demand observations separate. This process took place in step 3 of the procedure outlined in Table 5.1. Each historical LTD observation was calculated by summing all demand occurrences over a particular lead time period. Therefore all demand which occurred between when an order was placed for stock (purchase order) and the delivery of the item was included. This was carried out for every historical lead time observation. By calculating the LTD in such a way, the variability of both lead time and demand were incorporated into a single observation. As with Bookbinder and Lordahl [1989], random samples were drawn from the resulting LTD values (step 4, Table

Table 5.2: Jittering procedure [do Rego and de Mesquita, 2015]

Step	Description
1	Randomly sample a LTD or demand (dependent on method) quantity from historical data (X^*).
2	Randomly generate a single value from the standard Normal distribution ($z \sim Normal(0, 1)$).
3	Calculate jittered value = $integer(0.5 + X^* + z\sqrt{X^*})$.
4	If jittered value ≤ 0, then jittered value $= 0$.

5.1).

The jittering process in step 5 of Table 5.1 represented a further modification to the Bookbinder and Lordahl [1989] method. The jittering process applied (see Table 5.2) was an addition suggested first by Willemain et al. [2004] and subsequently adjusted by do Rego and de Mesquita [2015]. The introduction of this process made the bootstrapping technique more realistic, as instead of limiting the sampled values to those which had previously been observed, the possibility of similar but not exactly the same values was catered for.

It was in steps 4-9 of Table 5.1 that the bootstrapping procedure was used to generate an estimate of the LTD population percentile of interest. This was done by randomly sampling, with replacement, from the observed LTD dataset to form multiple bootstrap samples each of the same size as the original sample. The percentile of interest was then calculated for each of said samples and combined to form a sampling distribution of the LTD percentile of interest. Taking the expected value of this sampling distribution resulted in a single estimate of the true LTD population percentile of interest. As with the Normal approximation, the percentile of interest corresponded to the CSL, as defined in step 1 of the Overall simulation process flow in Figure 5.1.

The population percentile was used as the estimate for the reorder point. The choice of measure was made clear by the definition for CSL. The expression for the CSL was given by $CSL = P[X \leq s]$, where s was the reorder point and X the LTD. This was equivalent to finding the distribution percentile at which a percentage, equal to the CSL, of the LTD observations was found below this point.

The procedure for the second bootstrapping approach adopted in this study, the do Rego and de Mesquita [2015] adaptation, followed the steps detailed in Table 5.3. As with the Bookbinder and Lordahl [1989] model, bootstrapping was again used to generate the distribution of a measure relating to LTD. The fundamental difference between the two was the particular measure for which a distribution was generated. do Rego and de Mesquita [2015] utilised bootstrapping to generate a sampling distribution for the LTD, whilst Bookbinder and Lordahl [1989] employed bootstrapping to generate a distribution of a LTD population percentile. The do Rego and de Mesquita [2015] method also separately sampled the lead time, demand and inter demand arrival times in building the LTD periods, where inter demand arrival times were simply the number of days between subsequent demand periods. Whereas the Bookbinder and Lordahl [1989] method used the demand sequences and lead times as they occurred to calculate LTD values, and then sampled from those observed LTD values.

There were three key modifications made to the do Rego and de Mesquita [2015] methodology in this study. These included the choice of: lead time distribution, the service level measure, and as a result, the reorder point calculation. The modification to the lead time distribution was simple, in that instead of making use of a Triangle distribution as in do Rego and de Mesquita [2015], the actual data was used for drawing the random samples (described in step 3 of Table 5.3). The service level measure modification had a slightly larger impact on the method applied in this study. The original do Rego and de Mesquita [2015] method adopted the target fill rate as the service level measure whilst, as previously mentioned, this study made use of the CSL. Both metrics are commonly used in practice and research. Due to the simple interpretation and application of the CSL this metric was a natural choice for this study. It also allowed for comparisons between the three methods examined in this study.

As a result of the service level modification, the technique used to calculate the reorder point from the generated LTD distribution required adjustment. This simply required that the percentile of the generated LTD distribution be calculated as the estimate for the reorder point. As with the Bookbinder and Lordahl [1989] method, the percentile calculated in this do Rego and de Mesquita [2015] adaptation was dependent on the CSL defined in step 1 of the Overall simulation process flow shown in Figure 5.1.

Each of these modifications was accounted for in the steps listed in Table 5.3. This procedure constructed multiple LTD observations through randomly sampling

Table 5.3: Bootstrapping procedure [do Rego and de Mesquita, 2015]

Step	Description
1	Obtain historical demand data (including demand size and inter demand arrival times).
2	Obtain historical lead time data.
3	Randomly sample a lead time from the historical lead time data.
4	Randomly sample a demand quantity from the historical demand data. Apply jittering process.
5	Randomly sample inter-arrival time from historical data.
6	Increase time horizon by this interval (thus resulting in the time of the next demand occurrence).
7	If time horizon is equal to or lower than the lead time sample from step 2 then return to step 4. Otherwise, sum jittered demand during the lead time and obtain a single LTD value.
8	Repeat steps 3-7 2000 times to obtain a LTD sampling distribution.
9	Calculate the percentile of the LTD sample corresponding to the CSL.

from historical data in order to form a sampling distribution for LTD. The sampling distribution was then used to estimate the reorder point.

Once the data had been obtained (as per steps 1 and 2 in Table 5.3), a lead time observation was randomly sampled from the observed lead times in step 3. This was followed by randomly sampling both a demand and inter demand arrival time observation from the observed demand data and applying the jittering process described in Table 5.2 to the sampled demand value. The sampling of both demand and inter demand arrival time was repeated with replacement as many times as necessary in order to simulate the total demand over the sampled lead time. It was thus necessary to determine when the sampled lead time had passed and the next LTD observation should begin. Therefore on each repetition, a measure of elapsed time was incremented by a single day as well as a value equal to the randomly sampled inter demand arrival time. At the point when the elapsed time was greater than or equal to the sampled lead time, this sub-process was terminated and all jittered demands were added to form a single LTD observation. This procedure was repeated two thousand times to form a bootstrap sample or sampling distribution of LTD. The only remaining step in this method was to calculate the percentile, specified by the CSL, of the bootstrap sample generated in step 8. The resulting percentile represented the reorder point estimate.

This concludes the overview of the methodologies for the implementation of each of the three reorder point calculations which were compared in this study. In the following section the calculation used for the estimation of the optimal order quantity is discussed, followed by the Operational management process.

5.1.3 Tactical Management - Order Quantity Calculation

The order quantity Q, and the reorder point were required as input for the Operational management process. The calculation of the order quantity was carried out after the first iteration of the Tactical management process, where parameters were initialised, as well as at every iteration of the Tactical management process. It should be noted that the three methods being compared in this study differed in their calculation of the reorder point, s, but all methods made use of the same order quantity, Q.

A lack of information on Company A's ordering and holding costs meant that the standard EOQ formula given in Equation 2.7 could not be used for the calculation of this quantity. Instead, an alternative employed by Company A at the time of this research was adopted. Although this approach did not use the EOQ itself, the optimal replenishment cycle derived from the EOQ was utilised. The optimal replenishment cycle for the EOQ was defined as $T = \frac{Q}{D}$ in Equation 2.8 in Section 2.1.1.1, where T was the replenishment cycle, D was the average annual demand for an item, and Q the economic order quantity, or in this case referred to as the optimal order quantity. Based on the experience and expertise of Company A a replenishment cycle of one week or $\frac{1}{52}$ years was regarded as optimal. Given that both T and D were known, a simple manipulation of Equation 2.8 resulted in the desired optimal order quantity.

5.1.4 Operational Management - Inventory Control Policy

In order to compare the three reorder point methods discussed in Section 5.1.2, a simulation of their effect on daily stock levels was performed. This formed the Operational management process shown previously in Figure 5.1.

A specific inventory policy defining the method by which stock was ordered during the simulation was required. Section 2.1 introduced numerous continuous and period review policies commonly found in practice. The policy selected for the purposes of this simulation was an extension of the continuous review policy (s, Q) discussed in Section 2.1.2.1. The policy, also adopted by do Rego and de Mesquita [2015], was denoted by (s, nQ). As previously defined, s was the reorder point, Q was the order quantity and in this particular policy, n was an integer value which increased the order quantity by the minimum integer multiple which raised the IP above the reorder point. Similar to the (s, Q) policy described in Section 2.1.2.1 the (s, nQ) policy employed in this study continuously monitored the IP, which decreased at a variable rate as customer demand was filled. When the IP fell below the reorder point, s, an order of size nQ was placed. The (s, nQ)

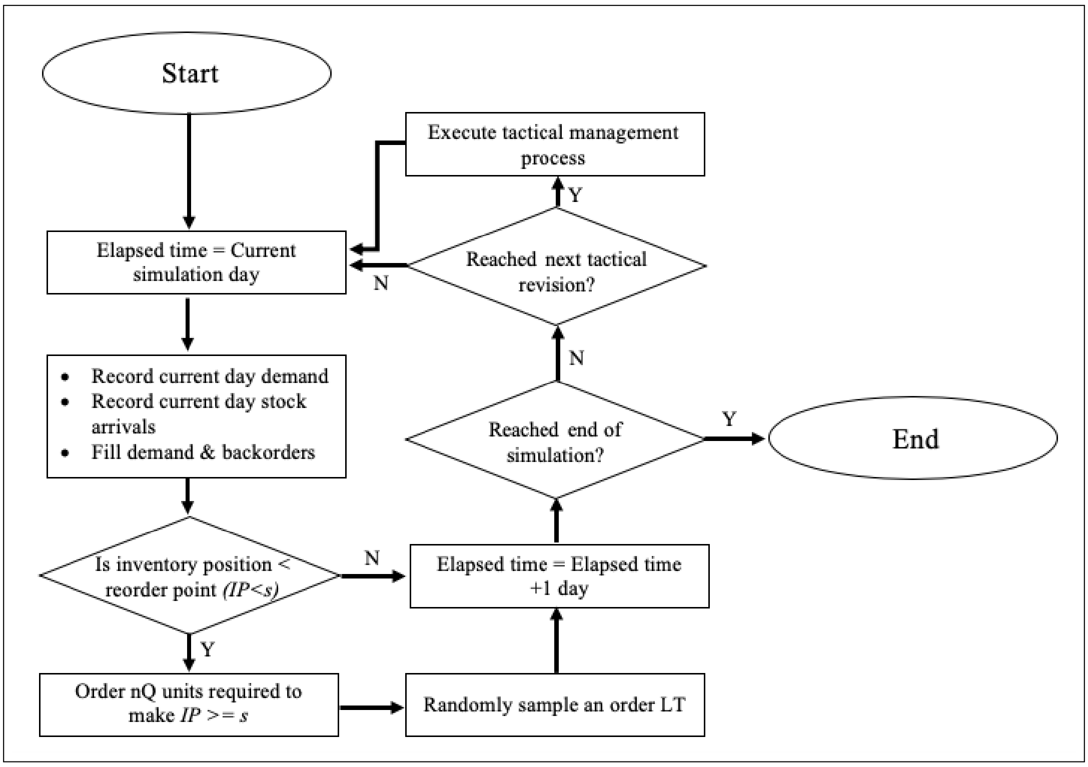

Figure 5.3: Operational simulation flow do Rego and de Mesquita [2015]

extension of the standard (s, Q) policy accounted for the occurrence of random demand sizes. In the event of a non unit sized replenishment triggering demand, occurring and causing the IP to drop below the reorder point, the (s, nQ) policy ensured that the IP always returned to a point which was greater than the reorder point [Federgruen and Zheng, 1992; Waters, 2003]. The adoption of this policy was supported by the fact that the industry data used in this study was subject to random demand sizes.

The process flow for the (s, nQ) policy simulation carried out for each of the policy versions is illustrated in Figure 5.3. Each policy version represented one of the three reorder point calculation approaches discussed in Section 5.1.2. Elapsed time, which contained the current date of the simulation, was continuously tracked and incremented on each passing day (iteration) of the simulation until such time that the end of the simulation period was reached (i.e. once elapsed time reached 31 December 2015). This formed the first step of the simulation flow shown in Figure 5.3. In the second step of the Operational simulation flow, each simulation day's actual demand observation and any outstanding back orders were subtracted from the IP while any stock arrivals were added. The closing IP was recorded at the end of every day. A decision step followed from this which determined whether an order was required or not. If $IP < s$, an order of size nQ was placed and a lead time was randomly sampled from historical data to determine when that order would arrive. Following this the elapsed time was incremented by one day. If $IP > s$ at the start of the simulation day, the ordering process was simply skipped and elapsed time incremented.

Once elapsed time had been incremented the decision of whether to update the s and Q parameters was taken. For the purposes of this study, these two parameters were revised monthly on the first of every month in the simulation period. At this point the policy simulation was temporarily suspended and the Tactical management revision process performed. The resulting updated policy parameters were then used in the next iteration of the Operational management process. This process continued until such time that the policy parameters required revision or the elapsed time had reached the end date of the simulation period.

5.2 Policy Comparison Metrics and Methods

This section describes the metrics and methods used in the comparison of results from the simulations of the three inventory policies described in Section 5.1.

The metrics chosen to represent the performance of each policy included: Realised Average Daily Stock on Hand (RSOH), Realised Total Holding Cost (RHC), Realised Fill Rate (RFR) and Realised Cycle Service Level (RCSL). These were chosen based on both data availability and similarities to metrics used by do Rego and de Mesquita [2015]. These metrics were calculated for each policy simulation.

The RSOH metric represented the average daily stock on hand for each item over the entire simulation period 1 January 2015 to 31 December 2015.

The RHC included all cost associated with holding inventory. Due to a lack of information on the subcomponents of this cost it is common practice to estimate the holding cost as a percentage of inventory value. There are numerous industry standards which exist for this estimation. In this study it was decided to calculate the RHC as 21.5 percent of the average daily inventory value. This choice was similar to that of do Rego and de Mesquita [2015].

The RHC was thus calculated by multiplying the average daily stock on hand for an item over the simulation period by its cost price and then multiplying by 21.5 percent. The cost price was assumed constant throughout the simulation period. This decision was based on the fact that the holding cost was evaluated over a relatively short period of time, and any inflationary effect on the cost price would have been minimal. In addition to this it was assumed that all items would have been affected by inflation to the same extent.

The two service level metrics utilised for method comparison in this study were the RFR and RCSL. Although fill rate was not used to set the inventory policies themselves, as with CSL, it is a common measure of policy performance in practice. The fill rate is defined as the proportion of demand satisfied by stock on hand in a replenishment cycle [Caplice, 2017]. The RFR was thus calculated by averaging weekly fill rates, given by demand quantity satisfied from stock on hand for a week divided by the total demand in that week. This resulted in an average weekly fill rate, referred to as the RFR.

The fourth and final metric used for simulation comparisons, RCSL, was determined as $1 - \frac{Stock\,out\,days}{Replenishment\,cycle\,length}$. As discussed in Section 5.1.3 the replenishment cycle length used in this study was one week i.e. seven days. The RCSL was calculated for each of the 52 replenishment cycles in the simulation period. The average of the 52 resulting CSL values was then calculated giving the average RCSL for each item over the full simulation period.

Each of the four metrics used in the simulation comparisons was first calculated at item level as described above. This was followed by aggregation to numerous levels, including a 1) high level aggregation, generating a single value to represent all items as well as 2) an aggregation of item results grouped into different segments based on either demand or lead time characteristics.

Three segmentation methods were adopted in this research and included: a demand based classification following the logic of Syntetos et al. [2005], a four class demand volume based segmentation (used at Company A), and a lead time variability based segmentation. Both demand based segmentations utilised demand observations for the period July 2014 to December 2014. The Syntetos et al. [2005] classification, implemented by do Rego and de Mesquita [2015], utilised both inter demand arrival time and CV of demand to define its classes, with cut-offs defined as in Figure 3.1. Further details of this method were discussed in Chapter 3.

For the demand volume segmentation, total demand over this period was summed for all items and three percentiles, namely the 25, 50 and 75th were calculated over all items. Each percentile represented the cutoff point for identifying the four groups of items. Group A was items with summed demand greater than the 75th percentile, group B comprised items with summed demand between the 50th and 75th percentiles, group C between the 25th and 50th percentiles and group D were those slow moving items below the 25th percentile.

The lead time based segmentation made use of lead times observed during the period 1 January 2014 to 31 December 2014. The CV of the lead times for each item was calculated during this period. They were then used to determine the 25, 50 and 75th percentiles of all item CV's to identify the cutoff points for four segments (also labelled A-D in this study). This segmentation was much the same as the demand volume segmentation, with only a change in the variable and descriptive statistic used to determine the segments. By incorporating these additional levels of comparison, it was possible to identify whether one method outperformed another for different types of items.

Finally, RCSL results for each method were compared and tested for statistically significant differences by either Quade's test or the Kruskal-Wallis test [Kruskal and Wallis, 1952; Quade, 1979]. Both tests suited the skew nature of the observed RCSL data that resulted from this research's simulations. It was for this reason that these tests were chosen over an Analysis of Variance (ANOVA) test (which assumes that the data are Normally distributed).

Quade's test supports data in the form of an unreplicated block design. This was the case when comparing the high level RCSL results as well as when comparing whether outlier treatment had an effect on each method's performance. In both these comparisons the dependent variable was RCSL, the independent variable was the method, and item code was the blocking variable. The null and alternative hypotheses of Quade's test were dependent on the similarity of the shape and spread of the dependent variable for each group under comparison. That is, when group distributions were symmetrical, the null hypothesis was that group medians were equal. However, when the condition of symmetry was not met, the null hypothesis was that the distribution of the observations for each group were equal. The alternative hypothesis in each case changed to the negation of the corresponding null hypothesis.

Quade's test simply determined whether a difference existed between any of the groups being compared. In order to determine between which pairs of groups differences existed post-hoc analysis was carried out using Quade's multiple comparison test [Quade, 1979]. This study made use of the R function quade.test to determine whether any significant differences existed between the distributions of observations for each group and the function posthoc.quade.test for the post-hoc analysis [R Core Team, 2018]. Descriptions and examples of these functions as well as the R packages from which they originate were all discussed in Mangiafico [2016].

The Kruskal-Wallis test is similar to the Mann-Whitney U test, but appropriate for one-way data with more than two groups [Kruskal and Wallis, 1952]. This test was therefore suitable for all segmented results that will be discussed in Chapter 6, where RCSL was the dependant variable and segment the independent variable with four groups. The null hypothesis for this test was that the groups were sampled from populations with identical distributions or similarly that the group distributions were identical. To determine the pairwise significance of group RCSL distributions, Dunn's Kruskal-Wallis multiple comparison test, also referred to as the Dunn test, was performed [Dunn, 1964]. This study adopted the R functions kruskal.test to ascertain whether any of the groups distributions differed significantly from the others and, dunnTest for post-hoc analysis, both of which originated from the FSA R package [Ogle et al., 2019].

This concludes the description of the methodology followed in this study. This methodology was represented by three key processes which included: the Overall, Tactical management and Operational management process. Together, these described the steps taken to calculate the relevant inventory policy parameters and simulate their effect using the industry data provided by Company A. In the following chapter, the results of each simulation are compared and discussed using the methods described in Section 5.2 of this chapter.

Chapter 6

Results and Discussion

This chapter introduces, compares and discusses the results for each of the simulations performed using one of the three reorder point calculation methods detailed in the previous chapter. These were variations of the bootstrapping methods presented by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] as well as the Normal approximation method described by Waters [2003]. These methods were referred to as B&L, R&M and Normal methods respectively in the remainder of this chapter. Four metrics were used for the purpose of measuring performance of these simulations. These metrics were: RSOH, RHC, RFR and RCSL - the calculations of which were presented in Section 5.2. Each metric was calculated over the full simulation period, January 2015 to December 2015, at multiple levels of item aggregation. Considering these metrics at multiple levels of aggregation enabled identification of the best reorder point calculation method for groups of items which exhibited similar demand or lead time properties. Although the large sample size used in this study increased the likelihood of significant results, for the sake of completeness, method results were compared by means of the Quade and Kruskal-Wallis tests for significance (see Section 5.2). These methods were chosen as ANOVA assumptions (specifically related to Normality and sphericity) were not met.

The following subsections include: a comparison and discussion of the performance metrics calculated for each of the methods at a high level (all items), for segments of items grouped by both demand and lead time characteristics, and finally a discussion of the results produced from data with outliers replaced by median demand values. Additionally, the results obtained from the analyses prompted the repetition of the simulations for all the methods at a range of CSLs, namely

85, 87.5, 90, 92.5, 95, 97.5 and 99 percent with additional CSLs of 70, 75 and 80 percent for the Normal method. These CSLs were hereafter referred to as the Implemented Cycle Service Level (ICSL). The Target Cycle Service level (TCSL) set by Company A remained constant at 95 percent throughout the discussions.

6.1 High Level Performance Across All Items

Initially performance of the three methods was assessed at a high level, across all items. The high level policy comparison metrics seen in Table 6.1 represent the results from the three policies compared in this study, where each policy had an ICSL of 95 percent. The first notable result visible in the mean RCSL column in Table 6.1 was the inability of any methods to achieve the TCSL of 95 percent set by Company A. The Normal method did however come close, with a mean RCSL of 94 percent. The fact that neither bootstrapping method met the TCSL was not surprising, as results reported by do Rego and de Mesquita [2015] also found the realised service level measure to be negatively biased when compared to the target service level. Teunter and Duncan [2009] and Syntetos et al. [2015] also both noted a difference between target and realised CSLs. Syntetos et al. [2015] stated that this was due to errors in estimating the LTD distribution. That is, estimations were below the TCSL.

A slightly more promising result for the bootstrapping methods, in terms of RCSL, was evident from the median RCSL metrics. The median RCSLs for both bootstrapping methods were closer to the 95 percent TCSL than the mean values, differing by a maximum of 8 percent from the TCSL. The large differences between mean and median RCSL suggested a skewed distribution of the metric across all items. This was confirmed by means of the boxplots shown in Figure 6.1. The five number summaries used to generate these boxplots were used to further investigate the spread of the observations for each method. Simply comparing the difference between the first and third quartiles (Q_1, Q_3) for each method confirmed the presence of the visually differing spread amongst the methods. There was approximately 9 percent difference between the Q_1 and Q_3 RCSL for the Normal method, 24 percent for the B&L method and 38 percent for the R&M method. This also suggested differing levels of variability in the observed RCSLs of each method. The Normal method produced the most consistent RCSLs, B&L the second most consistent and R&M the least.

Table 6.1: High level method comparison

Method	RSOH	RHC	RCSL (mean)	RCSL (median)
R&M	13	7721253	77	87
B&L	19	9552651	84	89
Normal	38	17284260	94	100

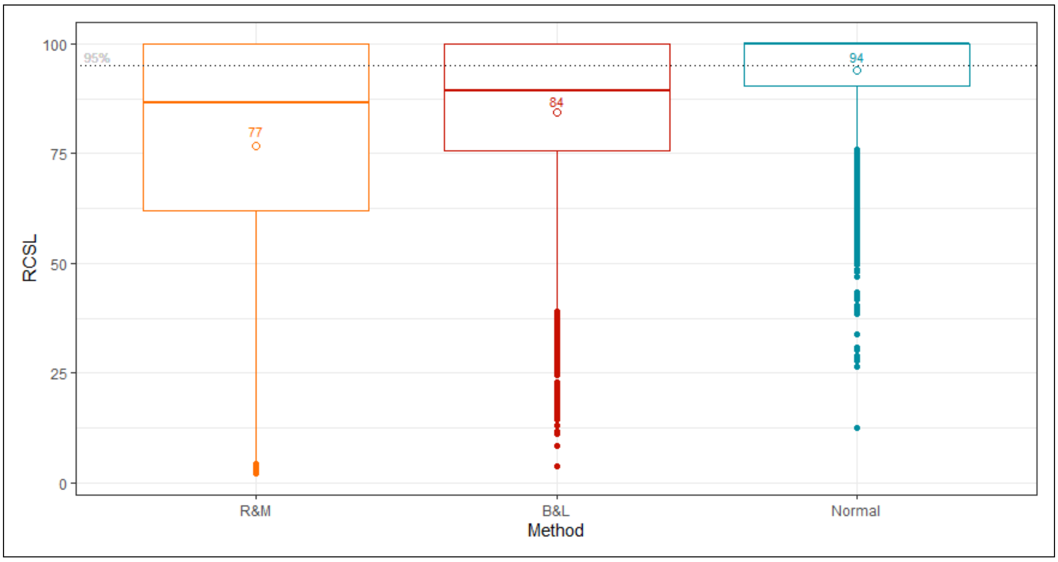

Figure 6.1: Boxplot of RCSL resulting from the three simulated inventory policies

RCSL distributions were compared and tested for statistical differences as a first step towards determining method performance and superiority and in order to determine whether the resulting differences in RCSL were in fact significant. As previously mentioned, the skewed distribution of the RCSL metric posed a problem to identifying significant differences between methods using a traditional ANOVA approach and hence Quade's test was used. Unlike the ANOVA, which compares group means, Quade's test compares medians, or in the case where spread and shape of the groups being compared are not similar, group distributions. This test supported data in the form of an unreplicated block design. This was true in the case of the high level results, where the dependent variable was RCSL, the independent variable method, and item code the blocking variable. The hypotheses were as follows:

$$H_0 : \; All\, methods'\; RCSL\, distributions\, are\, equal$$
$$H_a \; : At\, least\, one\, method's\, RCSL\, distribution\, is\, not\, equal$$

The results of Quade's test (see Appendix A, Table A.19) led to a rejection of H_0 and it was concluded that at least one methods' RCSL distributions was significantly different from the others, with a 95 percent level of confidence. Post-hoc analysis using Quade's multiple comparison test (see Appendix A, Table A.20) confirmed that in fact all three methods' RCSL distributions were significantly different from one another. This conclusion suggested that the three methods did indeed generate different results from each other. Finding a significant difference between method RCSL distributions was not surprising given the large sample size used in the simulations. The focus therefore shifted to the practical significance of these differences, which is discussed later in this section.

The Normal method clearly outperformed the others in terms of RCSL, however, this came at a cost of higher RHC and RSOH. This was most obvious for the RHCs as seen in Table 6.1 where the Normal method resulted in almost 125 percent more RHCs than the R&M method and over 80 percent more RHCs than the B&L method. The extent of the superior RCSL results of the Normal method were further investigated by means of identifying the common stocking behaviour present in the results of each method. As previously stated, the TCSL was set at 95 percent. It was considered unlikely that a particular method would achieve exactly 95 percent, making it difficult to determine exactly how many items reached, under or over-shot, the TCSL. For this reason RCSLs were split by what will be referred to as an "acceptable" range. This range included all items with a mean RCSL satisfying the inequality $94\% < RCSL < 96\%$. As CSL was concerned with the probability of stock out, items with $RCSL \leq 94\%$ were considered to be understocked, those in the acceptable range had the desired stock level, and those with $RCSL \geq 96\%$ were overstocked. The percentage of total items found in each of these ranges are shown in Table 6.2.

Table 6.2: Percentage of items in relevant RCSL range for each simulated method

Method	Understocked ($RCSL \leq 94$)	Desired stock ($94 < RCSL < 96$)	Overstocked ($RCSL \geq 96$)
R&M	63	4	33
B&L	60	5	35
Normal	32	6	62

From Table 6.2 it was obvious that the high RCSLs achieved by the Normal method came at the cost of overstocking 62 percent of items included in the study. This also explained the extremely high RHC seen in the results for the Normal method in Table 6.1. As Company A valued costs as well as CSL, this result reduced the initial signs of superiority of the Normal method when considering only RCSL. Furthermore, Table 6.2 also revealed a very similar performance for the methods when considering the percentage of items which achieved the desired stock range. The 4 to 6 percent of items falling into this range suggested poor performance amongst all methods in achieving the TCSL. At this stage, identifying the best method would require the company in question to identify whether holding cost or CSL was more important to the business and its strategy. If Company A wished to achieve a balance between understocking, which could cause customer losses, and overstocking which resulted in higher holding costs, then B&L may be a marginally better method. However, if promoting high customer satisfaction and retention, irrespective of higher inventory costs, was more in line with the business strategy, then the Normal method would be most appropriate.

Table 6.3: Realised fill rate and related measures comparison

Method	RFR (mean)	RFR (median)	Unfilled demand
R&M	77	88	20194
B&L	85	92	13373
Normal	94	100	4585

In an attempt to provided further insight into what significant differences between the three methods would mean in a more practical sense, the RFRs were calculated for each method's simulation results. This metric was chosen for its somewhat more informative and intuitive nature. The method of calculation for this metric was described in detail in Section 5.2. Fill rate can be measured in a number of marginally different ways when aggregating over a number of items and periods. In this study the metric represented on average what portion of weekly demand was met by stock on hand for all items. The mean and median RFRs as well as the unfilled demand for each method are detailed in Table 6.3. The unfilled demand column was calculated as the average weekly unfilled demand summed over all items. Despite the differing calculations of RFR and RCSL the two were obviously similar when comparing results in Tables 6.1 and 6.3. As with the TCSL, the target fill rate was set at 95 percent by Company A. An intuitive rational for the highly skewed distributions of both metrics across all items was that the methods performed better for groups of items with certain characteristics. This is discussed in the Sections 6.2 and 6.3.

A basic interpretation of the practical impact of the observed RFR for each method follows. For all items, the mean RFR of 77 percent observed for the R&M method resulted in approximately 20194 units of demand which could not be filled from stock on hand. The B&L method resulted in about two thirds of this amount, 13373 units, not being filled and just over a third of this amount, 4585 units, were not filled under the Normal method. This clearly illustrated the practical significance to the customer of seemingly minor differences between method results. That is, although methods only differed in their RFR by a maximum of 17 percent, the average unfilled weekly demand was almost five times larger, for the two extreme methods, R&M and Normal. To illustrate the potential impact on customers, consider the following example. If the average demand per customer was 1 unit per week and there were 100 customers, the resulting RFRs showed that approximately six customers would not have their demand filled per week for the Normal method, 15 customers for the B&L method and 23 for the R&M method. Logically, the larger the number of unfilled demand, the higher the number of unsatisfied customers which could in turn lead to a significant revenue loss. The practical results thus supported the previously reported presence of statistically significant differences between the three method RCSL results. Furthermore, based on the service level alone the practical results lend support to the use of the Normal method.

Thus far, only the results of simulations executed with an ICSL of 95 percent have been discussed. These results alluded to a positive relationship between the mean RCSL, RSOH and RHC between methods. That is, methods with higher mean RCSL also had higher RHC and RSOH. As previously discussed this result was predictable as it is logical that when keeping demand consistent, holding additional stock on hand, and hence increasing holding costs would result in an increase in RCSL. It thus remained to be determined whether this relationship was true for different ICSLs. Running the simulations at additional ICSL allowed for the curves of the relationship between mean RCSL and mean RHC across different ICSLs which resulted from each method, to be compared. These simulations also facilitated discussion around whether an increase in the ICSLs could result in the bootstrapping methods achieving the TCSL of 95 percent without increasing RHCs above those of the Normal method. As previously mentioned, policies were simulated for multiple ICSL including 85, 87.5, 90, 92.5, 95, 97.5 and 99 percent with the resulting values for RCSL, RHC and RSOH presented in Figures 6.2 and 6.3.

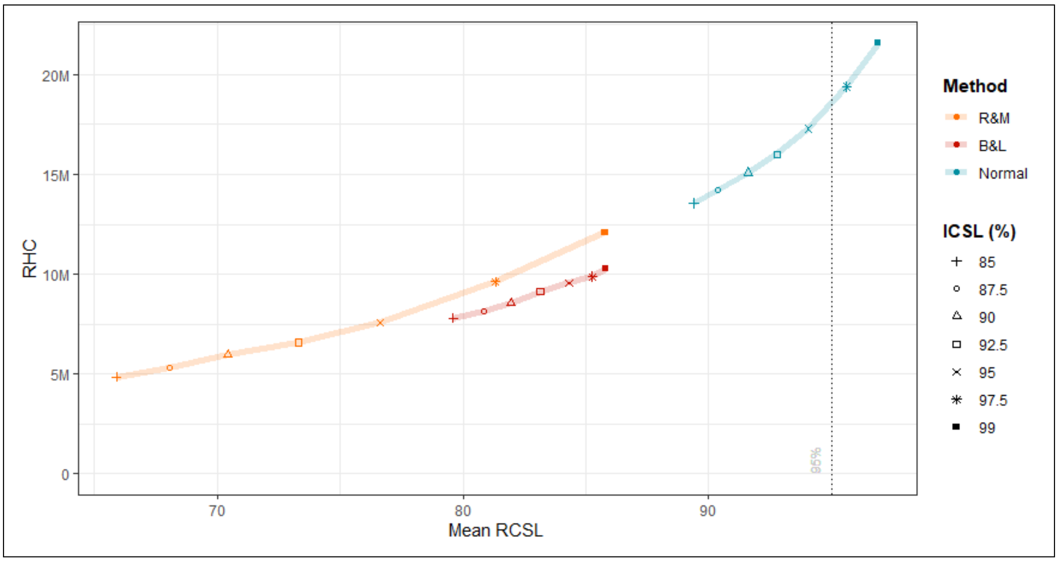

Figure 6.2: Mean RHC versus mean RCSL results for multiple ICSLs

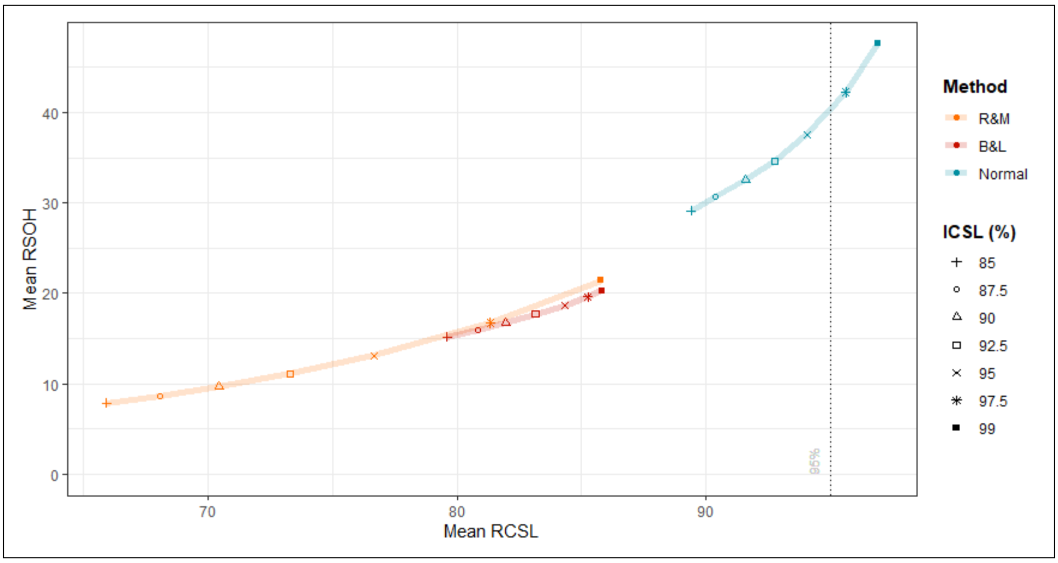

Figure 6.3: Mean RSOH versus mean RCSL at multiple ICSLs

The first notable result from the comparison of RHC and mean RCSL for each method (Figure 6.2) was that the Normal method again resulted in the highest mean RCSL and the highest RHC across all ICSLs. Both the mean RCSL and RHC obtained by the Normal method were not matched by either of the bootstrapping methods.

The presence of RCSLs higher than the ICSLs implied that the Normal method overstocked on all ICSLs below 95 percent, with the mean RCSL not dropping below 89 percent, despite the ICSL being set as low as 85 percent. For ICSLs equal to or above 95 percent this method resulted in understocking and hence mean RCSLs lower than the ICSLs. This suggested that the Normal method overestimated tail probabilities of the LTD distribution up to 95 percent but underestimated the more extreme tail probabilities above 95 percent. Both bootstrapping methods understocked for all ICSLs, with the RCSL not exceeding 90 percent even for the maximum ICSL of 99 percent. This showed that severe underestimation of the full range of tail probabilities of the LTD distributions occurred with the bootstrapping methods and did not provide for enough stock to cover the observed LTD to the specified service level. A possible reason for this was that the bootstrapping sampling procedure only resampled from the observed data with minor departures introduced by the jittering procedure, and consequently may not have been able to estimate extreme percentiles accurately. This is in comparison to a parametric approach, which can extrapolate beyond the observed data. This effect would have been especially pronounced in items with a limited number of distinct demand observations. This poor result for the bootstrapping methods further supported the use of the Normal method in terms of achieving the TCSL.

The R&M method showed the lowest RCSLs of all the methods and only reached comparable RCSLs to the B&L method at ICSLs of 97.5 and 99 percent. In order to reach similar RCSLs to the B&L method, the R&M method required increased stock holding, indicating that it would be a more expensive approach without achieving the benefit of additional service levels.

Also notable was the presence of diminishing returns experienced by both the R&M and Normal methods. This was particularly obvious for the Normal method which exhibited an increasing curve gradient towards the higher ICSLs. That is, with each incremental increase in ICSL the mean RCSL of the Normal method experienced only small increases (between 1 and 2 percent), whilst mean RHC became progressively larger with each increment in ICSL. As a result the Normal method showed the largest range of RHC values. This sensitivity of the Normal method to changes in the ICSL (especially at high ICSL values) may have negative consequences in practice. Inventory managers often increase the CSL arbitrarily in a bid to improve customer service, without considering the resultant non-linear increase in inventory holding costs.

The R&M method displayed a much more gradual increase in the curve gradient, making the presence of diminishing returns less visibly obvious. This was as a result of the much larger differences between mean RCSL for each increment of ICSL. However, upon closer inspection it was noted that the benefit of the large increases in mean RCSLs for each increase in ICSL became less favourable for ICSL above 95 percent. This was due to the increasing value of the RHCs. The B&L method did not exhibit the same pattern of increasing gradient of RHC versus mean RCSL as the other two methods, and consequently had a much smaller range of RHC values.

It was interesting to note the smaller range of mean RCSLs achieved by both the B&L and Normal methods. The B&L method realised a difference of 6 percent between the minimum and maximum mean RCSLs resulting from the ICSLs of 85 and 99 percent respectively. Similarly, the Normal method realised a difference of 7 percent between these two extreme points. The R&M on the other hand had a much larger range of mean RCSLs equating to approximately 20 percent. The larger range of mean RCSLs of the R&M method may have been due to the way in which the R&M bootstrap sample was created – this method allowed for more variation in the LTD values used to generate the LTD distribution and this may have resulted in a larger range of mean RCSL values.

Similar results and patterns seen in the comparison of RHC and mean RCSL were also evident in the relationship between mean RSOH and mean RCSL (Figure 6.3). However, when comparing the two graphics it was noted that unlike the RHCs/mean RCSL curves, which had a notable gap between the R&M and B&L methods, the mean RSOH/mean RCSL curves for the two methods were almost equivalent. This may have been due to the R&M method recommending a larger number of higher cost items, which tended to be slower moving and would result in similar mean stock on hand levels but higher RHCs for the R&M method and lower RHCs for the B&L method. This is investigated further in later sections which discuss results for items segmented by demand characteristics.

The key findings from this comparison were thus:

- The Normal method both over and under-achieved ICSLs, but overall resulted in far higher RCSLs and RHC than the bootstrapping methods.
- Neither bootstrapping method surpassed 90 percent mean RCSL.

- The B&M method achieved similar RCSLs to the R&M method for lower RHC.
- The Normal method showed sharp increases in RHC as ICSL increased, and yielded the biggest range of RHC overall.
- The range of mean RCSLs for both the Normal and B&L methods were small, at 6 and 7 percent respectively, while the R&M method had a much larger range of 20 percent.
- For the same range of mean RCSL values, the RSOH was consistently similar for the B&L and R&M methods. However, the RHCs were lower for the B&L method.

This concludes the discussion of the high level results (across all items) for each method. In the sections which follow, items were segmented into groups to determine whether the methods performed differently for each group.

6.2 Segmented Demand Level

In order to provide a more thorough investigation of the simulation results, items which formed part of the study were segmented into groups based on their demand characteristics. Two methods were adopted for this purpose, the details of which were discussed in Section 5.2. Both methods segmented items into four groups based on demand history characteristics. The Syntetos et al. [2005] method of segmentation was carried out first and the subsequent results calculated for each of the four groups are presented below.

Initial results from this segmentation, presented in Figure 6.4, summarised the distribution of items among the four segments. The most commonly occurring items were those defined as "Slow". These items had demand with a CV^2 less than or equal to 0.49 and an average inter demand arrival time greater than 1.32 days. The "Slow" segment thus represented items with low demand variability and large gaps between demands. These demand characteristics are prevalent in the spare parts industry. This fact was supported by the distribution of the items in the do Rego and de Mesquita [2015] research, which also made use of automotive data. The remaining 37 percent of the items used for the present study were spread between the "Smooth", "Erratic" and "Lumpy" segments, with a slightly larger portion found in the "Smooth" and "Lumpy" segments.

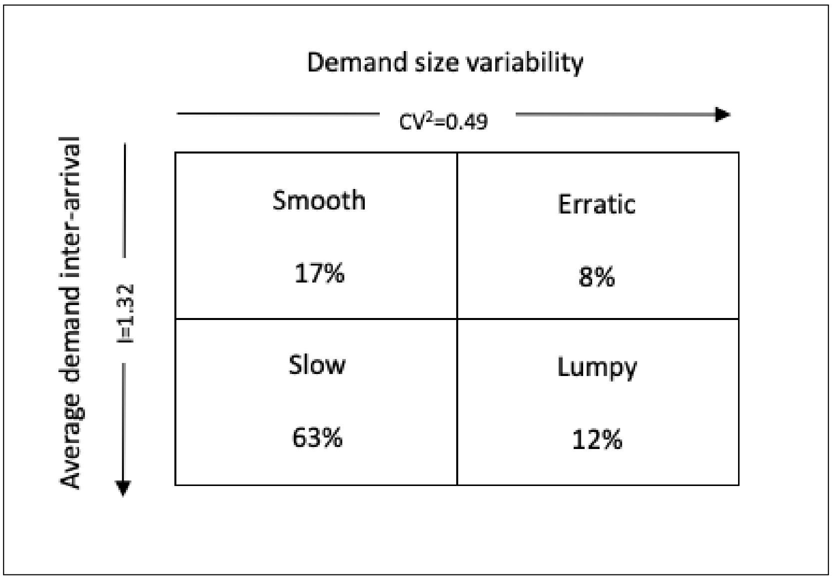

Figure 6.4: Syntetos et al. [2005] item segmentation – percentage of items in each segment

The purpose of this segmentation was to determine whether the three inventory policy methods performed differently for each of the demand segments. As a first step towards identifying the presence of these differences, the mean RCSL results for each method, with an ICSL of 95 percent, were plotted for the four segments. RCSL results were aggregated to a mean value for each method and segment in Figure 6.5. For Figure 6.6 the item level RCSLs were summarised in the form of a boxplot for each method and segment. These two figures illustrated that differences between the RCSLs of each method were noticeable amongst segments. The R&M method in particular had one segment which presented large differences between the mean RCSL it achieved and those of the remaining three segments. This was the "Slow" segment and was the best performing for the R&M method. This segment performed above the mean RCSL for the method, achieving a mean RCSL of 84 percent. This supports the potential reason identified in the high level results (Section 6.1), for the difference between the RSOH/RCSL and RHC/RCSL curves for the B&L and R&M methods - that the R&M may have recommended a larger number of higher cost slow moving items.

Segment RCSL distributions were tested for significant differences for each method using the Kruskal-Wallis test and post-hoc analysis was done using the Dunn test. The Kruskal-Wallis test is appropriate for one-way data with more than two groups. This was the case for all segmented results discussed in the remainder of

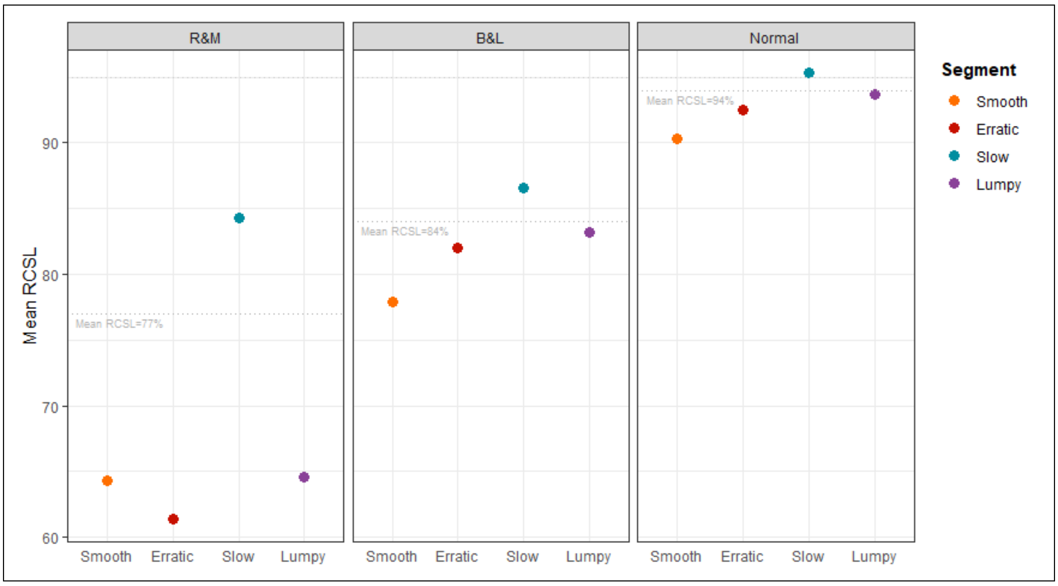

Figure 6.5: Mean RCSL achieved for each Syntetos et al. [2005] segment and method at 95 percent ICSL

this section, where RCSL was the dependant variable and segment the independent variable with four groups. Further details on these tests were discussed in Section 5.2. The hypotheses were as follows:

$$H_0: \; RCSL\, distributions\, are\, equal\, for\, all\, segments$$
$$H_a: At\, least\, one\, RCSL\, distribution\, is\, not\, equal\, to\, those\, of\, the\, other\, segments$$

As alluded to by the boxplots for the R&M method in Figure 6.6 the results of these tests (see Appendix A, Tables A.1 & A.2) concluded that all segments RCSL distributions were significantly different to the “Slow” segment. One further significant difference was found between the “Erratic” and “Lumpy” segments.

For the B&L and Normal methods, the “Slow” segment did not show an obvious visual difference in distribution as was the case for the R&M method (see Figure 6.6). However, both methods’ “Smooth” segment RCSL distributions appeared to be visually different from the remaining three segments. Once again making use of the Kruskal-Wallis and Dunn tests, significant differences between distributions were found between all segments apart from the “Erratic” and “Lumpy” segments for both the B&L and Normal methods (see Appendix A, Tables A.3 - A.6).

The key conclusion from these results was that the three methods did in fact perform differently, in terms of RCSL, depending on the item demand characteristics.

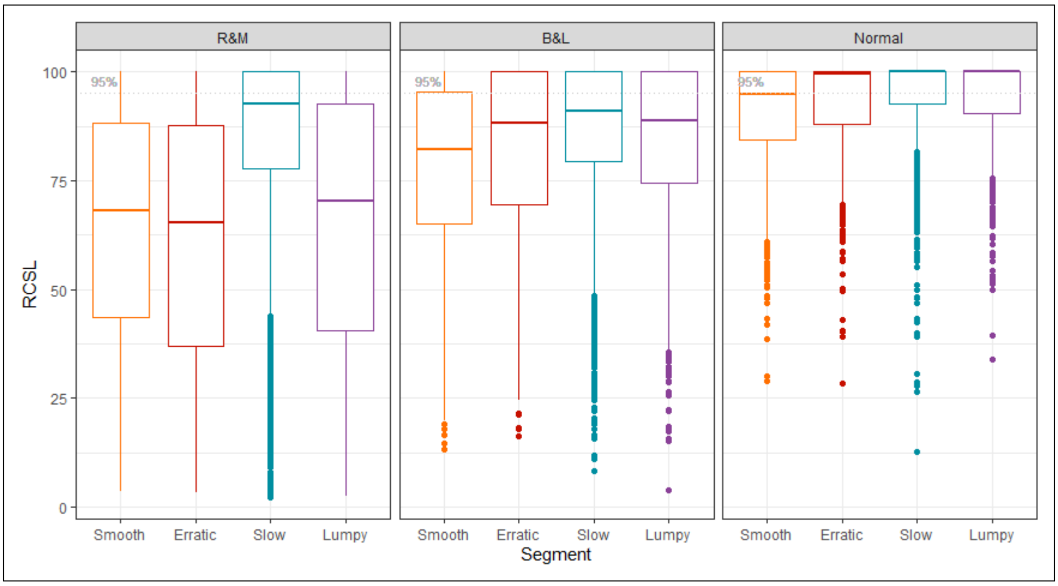

Figure 6.6: Boxplots for RCSLs achieved for each method and Syntetos et al. [2005] segment at a 95 percent ICSL

Regarding RCSL, the R&M method's best results were for "Slow" items. However, the mean RCSL still did not reach the TCSL of 95 percent, likely due to a long tail in the LTD data. The B&L method performed best for "Erratic", "Lumpy" and "Slow" segment items. As with the R&M method, the "Slow" segment items under the B&L method achieved the highest mean RCSL. The Normal method performed well for all segments despite the lower mean and median RCSLs achieved by the "Smooth" segment. The results from Figure 6.6 also suggested that the Normal method overstocked on items with either high demand variability or high inter demand arrival times.

At this point it was unclear whether changes in the ICSL used for each method would result in an outcome where additional method-segment combinations achieved the desired TCSL of 95 percent. It was also necessary to determine whether a bootstrapping method could produce an acceptable RCSL for a lower inventory cost than the Normal method. For these purposes, the resultant RCSLs for each method using a range of ICSL (85, 87.5, 90, 92.5, 95, 97.5 and 99) were graphed against the corresponding RHCs (Figure 6.7). As with the high level results this also provided an opportunity to examine the relationship between RHC and RCSL for each method-segment combination. The resulting figures for RHC versus mean RFR and RSOH versus mean RCSL were similar to those in Figure 6.7 and were not discussed further here (see Appendix B, Figures B.1 & B.2).

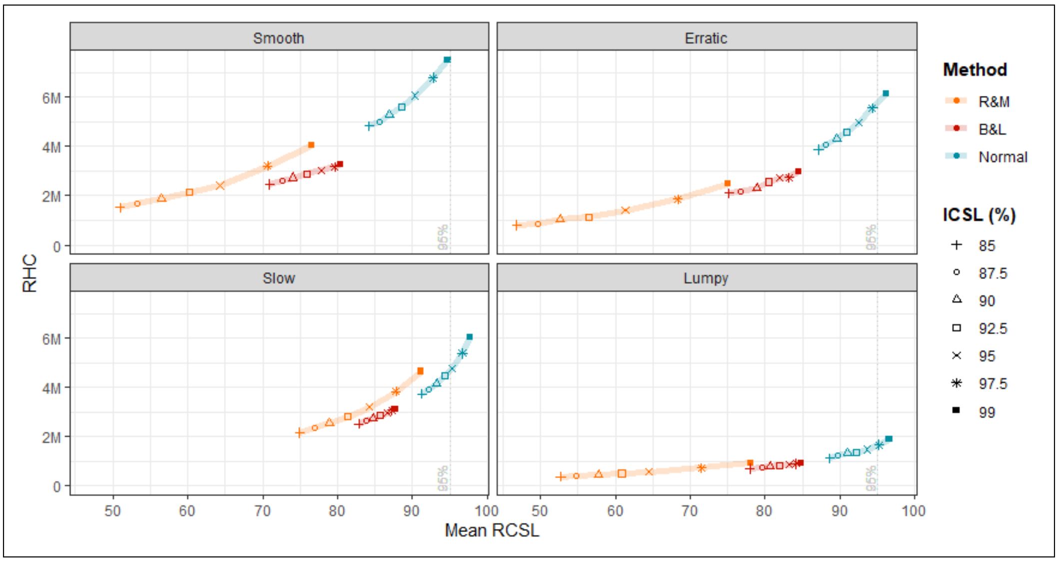

Figure 6.7: RHC versus RCSL for each Syntetos et al. [2005] segment and method combination

In keeping with the high level results seen in Figure 6.2, the segmented results seen in Figure 6.7 indicated that the Normal method was the only one to achieve the TCSL of 95 percent, irrespective of segment. The Normal method provided higher RCSL than the bootstrapping methods for all ICSLs and segments. The B&L method achieved higher RCSL than the R&M method for each corresponding ICSL and segment, with the exception of the "Slow" segment.

Across all segments, the RCSLs achieved by the R&M method were too low for it to be considered a viable option. The exception to this was the "Slow" segment where a RCSL of above 90 percent was achieved. However, the Normal method was able to achieve this and higher RCSLs at a lower RHC than the R&M method. For the remaining segments, the R&M method showed that the highest ICSL (99 percent) resulted in more RHC and a lower RCSL than results obtained from the B&L method. Consequently, there did not appear to be justification for the use of the R&M method in any of the Syntetos et al. [2005] segments.

In general, the superior RCSL of the Normal method was accompanied by higher RHC values. The exception to this was the "Slow" segment where the Normal method achieved better RCSLs for lower RHCs than the R&M method. Extrapolating the Normal method curve to lower RCSLs for this segment suggested that this method would continue to achieve lower RHCs than the R&M method and result in similar RHCs to the B&L method. The Normal method thus appeared to be the best approach for the "Slow" segment, given its lower RHCs and lower degree of complexity.

For the “Erratic” and “Smooth” segments, the maximum mean RCSLs achieved by the B&L method were 84 and 81 percent respectively. The trend and shape of the points plotted for the Normal method suggested that at RCSLs equal to the maxima achieved by the B&L method, the Normal method could result in a higher RHC for both these segments. Therefore, if Company A were to place a higher priority on minimising RHCs and reduce their TCSL to 84 and 81 percent for the “Erratic” and “Smooth” segments respectively, the B&L method could potentially outperform the Normal method in terms of RHC for these item segments. This observation was supported through additional simulations at lower ICSLs, the results of which are discussed later in this section. However, if customer service was a more important consideration, the Normal method remained the more appropriate choice.

For the “Lumpy” segment, extrapolating the Normal method curve to lower ICSLs suggested similar results to those of the B&L method. This was not investigated further in the absence of strong visual evidence to suggest the B&L method might provide better results at lower RCSLs. Therefore the Normal method was concluded to be most appropriate for this segment.

In their research, do Rego and de Mesquita [2015] recommended which method best achieved the TCSL for the lowest RHC for each item segment. Given that in the current study the only method to reach the TCSL of 95 percent was the Normal method, following the logic of do Rego and de Mesquita [2015], the recommendation would be that all item segments utilised the Normal method. However, considering the results discussed above, a possible alternative would be to use the B&L method for items with “Erratic” or possibly even “Smooth” demand characteristics, especially if Company A were willing to drop their TCSL.

In addition to the Syntetos et al. [2005] segmentation, one based solely on demand volume was performed on the data. The methodology for the calculation of and constraints for each segment was discussed in Section 5.2. This segmentation was included for two reasons, firstly, to address the demand volume characteristic not addressed in the Syntetos et al. [2005] segmentation, and secondly, because this was the segmentation technique employed by Company A. The resulting segments each contained approximately 25 percent of the item pool under study. Those items with the highest demand volumes were termed “A” and the lowest “D”. The performance metrics at an ICSL of 95 percent were summarised for all items in each of the demand volume segments with the resulting metrics presented in Table 6.4. In addition to these tabulated results, Figure 6.8 illustrates the distribution of the RCSLs for each method and segment in the form of boxplots.

As expected, methods with higher mean RCSLs were accompanied by higher RHCs for each segment (Table 6.4). The exception to this however was the "C" segment for which the B&L method required RHCs approximately 15000 lower to provide a two percent higher mean RCSL than the R&M method. The relationship between RCSL and RHC for the demand volume segments will be discussed in more detail later in this section.

Table 6.4: Performance metrics for each method and demand volume segment combination at an ICSL of 95 percent

Method	Demand volume segment	RSOH	RHC	RCSL (mean)	RCSL (median)	RFR
R&M	A	37	3807378	57	58	57
	B	6	1460685	72	80	73
	C	5	1196662	85	92	86
	D	4	1256528	91	97	92
B&L	A	58	5716058	79	83	79
	B	9	1676978	84	89	84
	C	5	1181300	87	91	88
	D	3	978315	88	92	89
Normal	A	124	11069244	91	96	91
	B	14	2814005	93	99	94
	C	8	1782805	95	100	96
	D	5	1618206	96	100	97

Results for all methods, as seen in Table 6.4, suggested the presence of an inversely proportional relationship between demand volume and mean RCSL. That is, as demand volume decreased towards the "D" segment, the mean RCSL increased. This relationship was a lot less noticeable for the B&L and Normal methods, both of which appeared to be less affected by different demand volume characteristics.

The Kruskal-Wallis and Dunn tests for significance and post-hoc analysis were performed to determine whether the above mentioned differences in the RCSL

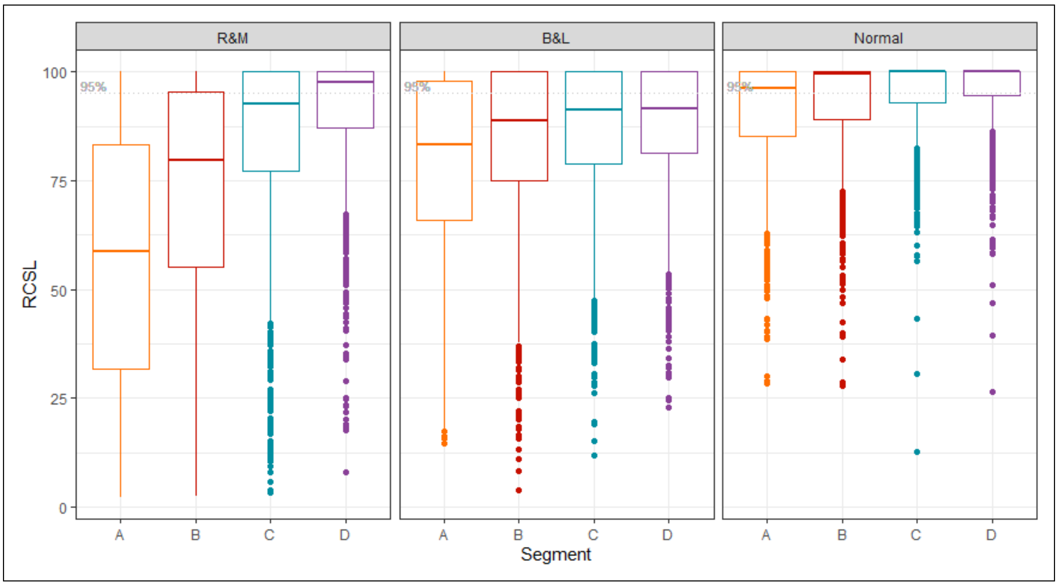

Figure 6.8: Boxplots for RCSLs achieved for each method and demand volume segment at 95 percent ICSL

distributions were in fact statistically significant. The hypotheses considered for the tests conducted for each method were as follows:

$$H_0: \text{ All demand volume segment } RCSL \text{ distributions are the same}$$
$$H_a: \text{ At least one demand volume segment } RCSL \text{ distribution is not the same}$$

From the results of these tests, all segments' RCSL distributions were concluded to be significantly different from each other for both the R&M and B&L methods at a 95 percent level of confidence (see Appendix A, Tables A.7 - A.10). For the Normal method all demand volume segments except for "C" and "D" were found to be significantly different at a 95 percent level of confidence (see Appendix A, Tables A.11 & A.12). As with the Syntetos et al. [2005] segmentation, the relationship between mean RCSL and RHC for each segment and method provided valuable insight into the superior performance areas of each method. In order to investigate this, the simulations were repeated at a range of ICSLs (85, 87.5, 90, 92.5, 95, 97.5 and 99). Figure 6.9 illustrates the resulting mean RCSL and RHC for each method and demand volume segment. Figures B.4 and B.3 in Appendix B present the graphical output for RHC versus mean RFR and mean RSOH versus mean RCSL respectively, which resulted in the same conclusions as those drawn from Figure 6.9.

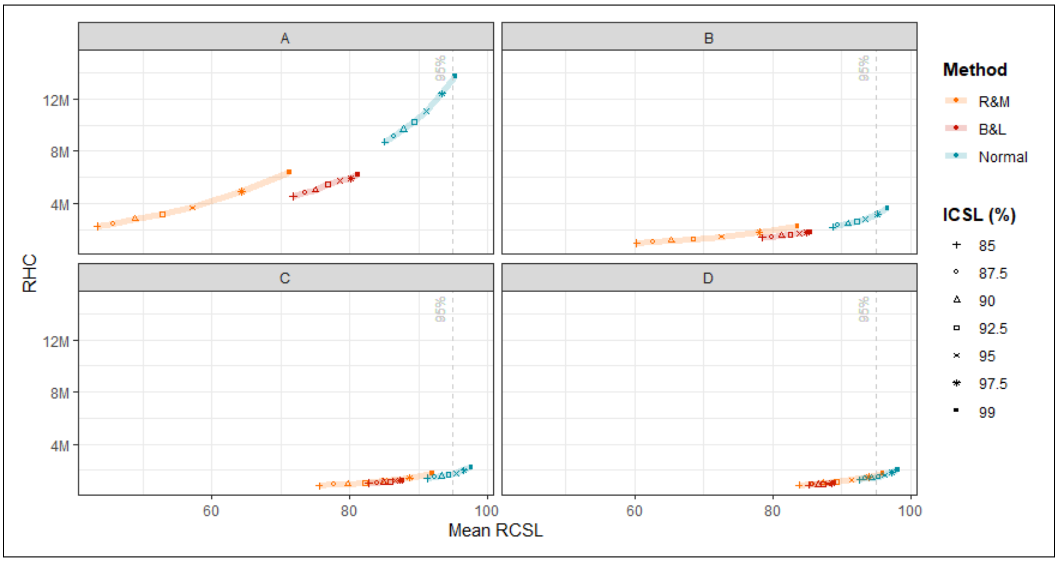

Figure 6.9: RHC versus mean RCSL for each method and demand volume segment combination at multiple ICSLs

The first obvious take away from Figure 6.9 was the high RHCs for the "A" segment items, irrespective of method. This result was expected to be as high-volume items inherently require larger volumes of stock on hand and hence increase holding costs. "A" segment items exhibited similar results to those of the "Smooth" segment from the Syntetos et al. [2005] segmentation. The conclusion for items with high demand volumes would thus be to use the Normal method, if RCSL was the key priority for Company A, or if a slightly lower RCSL was acceptable, to then consider the use of the B&L method as this may result in lower RHCs. This is investigated further later in this section.

The Normal method was the only approach to reach the TCSL for "B" segment items, with the next closest being the B&L method with a maximum RCSL of 85 percent. Here, the lower RCSL of the B&L method did not yield as visually notable a drop in RHC as the "A" segment. The Normal method was also the only method to reach the TCSL for "C" segment items, although the R&M method did come close to the TCSL, with a maximum mean RCSL of 92 percent (achieved when the ICSL was set to 99 percent). The 92 percent mean RCSL achieved by the R&M method was however accompanied by RHCs higher than the RHCs for the same and higher RCSLs of the Normal method. The B&L method also did not show a visually notable drop in RHC compared to the Normal method for the "C" segment. The Normal method would therefore be recommended for items which fall into the "B" and "C" demand volume segments.

The “D” segment items presented an interesting result for the R&M method. For this segment the R&M method surpassed the TCSL of 95 percent. This was the first time either of the bootstrapping methods were able to reach or surpass 95 percent mean RCSL. However, the R&M method did require a slightly higher RHC than the Normal method to achieve this. The B&L method exhibited the poorest performance in terms of RCSL and did not show visual evidence of a significant drop in RHC at the lower RCSL. The recommendation for this segment would thus be to use the least complex Normal method.

As noted previously, results of the B&L and Normal method at the high level, and for selected segments, suggested that at lower RCSLs the B&L method would result in lower RHCs and hence be best at that RCSL. This conclusion was based on visual extrapolation of the simulation results for each ICSL seen in Figures 6.2, 6.7 and 6.9. Three further simulations of the Normal method were conducted at ICSLs of 70, 75 and 80 percent in an attempt to be more precise in the extrapolation of results and lend support to these conclusions. The points for RHC and mean RCSL resulting from each of these simulations were added to the high level results in Figure 6.2, the “Smooth” segment results in Figure 6.7 and the “A” segment results in Figure 6.9, with the resulting graphs shown in Figure 6.10.

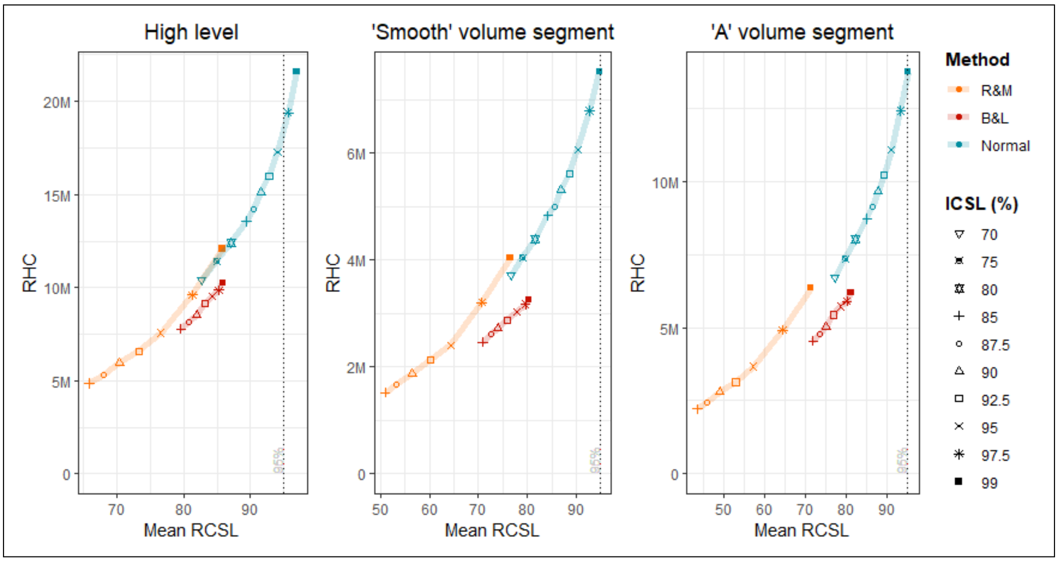

Figure 6.10: High level, “Smooth” and “A” segment graphs of mean RCSL and RHC with additional ICSLs 70, 75 and 80 percent

These results clearly indicated that at a lower mean RCSL, below 86 percent for high level results and below 81 percent for “Smooth” and “A” segment items, the

Normal method resulted in higher RHCs than those of the B&L method for the same RCSL. These results therefore supported the earlier conclusion that the B&L method was superior in terms of RHC if a reduced CSL for "A" or "Smooth" items was acceptable. The B&L method also provided the smallest range in RHC. As noted previously this offered a practical advantage as, in practice, TCSLs are often increased without quantifying the effect on RHC.

This concludes the analysis and discussion of the demand based segmentations of the simulation results for the three methods under investigation. The section that follows investigated the simulation results which were segmented based on lead time characteristics, as opposed to the demand characteristics focused on in this section.

6.3 Segmented Lead Time Level

As a further step in determining the best method, the effect of lead time variability on simulation outcomes was investigated. For the purpose of this investigation the lead time CV for each item was calculated over the period 12 months prior to the simulation start date. In much the same way as in the previous section, where demand volume was used to identify four item segments A-D, the lead time CV for each item was used to identify four item segments for this comparison, also labelled A-D. Each segment consisted of approximately 25 percent of the items. "A" items had the largest lead time variability and "D" items the lowest.

The aggregated results for each method simulation, with an ICSL of 95 percent, were tabulated for each of the four segments in Table 6.5. Boxplots for the item level RCSLs achieved for each method-segment combination, with an ICSL of 95 percent, were illustrated in Figure 6.11.

The results from Table 6.5 and Figure 6.11 showed relative consistency in RCSL across lead time CV segments A-D. As expected, RHCs were highest for, "A" segment items and lowest for "C" and "D" segment items.

To determine whether segment RCSL distributions differed significantly, each method was tested using the Kruskal-Wallis test and Dunn test for post-hoc analysis. As with the segmented demand results in the previous section, the data for the lead time CV segmented results were suited to the Kruskal-Wallis and Dunn

Table 6.5: Performance metric comparison for each lead time CV segment

Method	Lead time CV segment	RSOH	RHC	RCSL (mean)	RCSL (median)	RFR
R&M	A	21	3691447	80	91	80
	B	11	1600789	74	84	75
	C	10	1189089	76	85	77
	D	10	1239918	77	87	77
B&L	A	27	4788681	85	90	85
	B	17	1953127	83	88	84
	C	15	1429582	84	89	86
	D	15	1381261	85	89	85
Normal	A	69	9748767	96	100	96
	B	35	3413788	94	99	94
	C	23	2077512	93	99	94
	D	23	2044193	93	99	93

tests. Hypotheses were as follows:

$$H_0: \text{All lead time CV segment RCSL distributions are the same}$$
$$H_a: \text{At least one lead time CV segment RCSL distribution is not the same}$$

The results from these tests (see Appendix A, Tables A.13 - A.18) supported the following conclusions:

- Considering the R&M method alone, the RCSL distribution for lead time CV segment "A" differed from the other segment distributions. The "A" segment also produced the highest mean RCSL for the R&M method. The only other significant difference for this method was found to exist between the "B" and "D" segment RCSL distributions.

- The results from the B&L method analysis indicated that the "B" segment RCSL distribution was significantly different from the other segments. This was the only significant difference found between the four segment distributions. Also, although significant differences were concluded to exist between the "B" segment and the other segments, the Dunn test p-values were close to the chosen level of significance value of 0.05. This suggested that distribution differences were not highly significant and hence lead time variability had a limited influence on the results for the B&L method.

- The "A" segment RCSL distribution was concluded to be significantly different from all three other segment distributions for the Normal method. This was clearly visible in Figure 6.11 where the "A" segment showed a much smaller inter quartile range than the other three segments. The "A" segment also produced the highest mean RCSL for the Normal method. In fact the mean RCSL exceeded the TCSL of 95 percent.

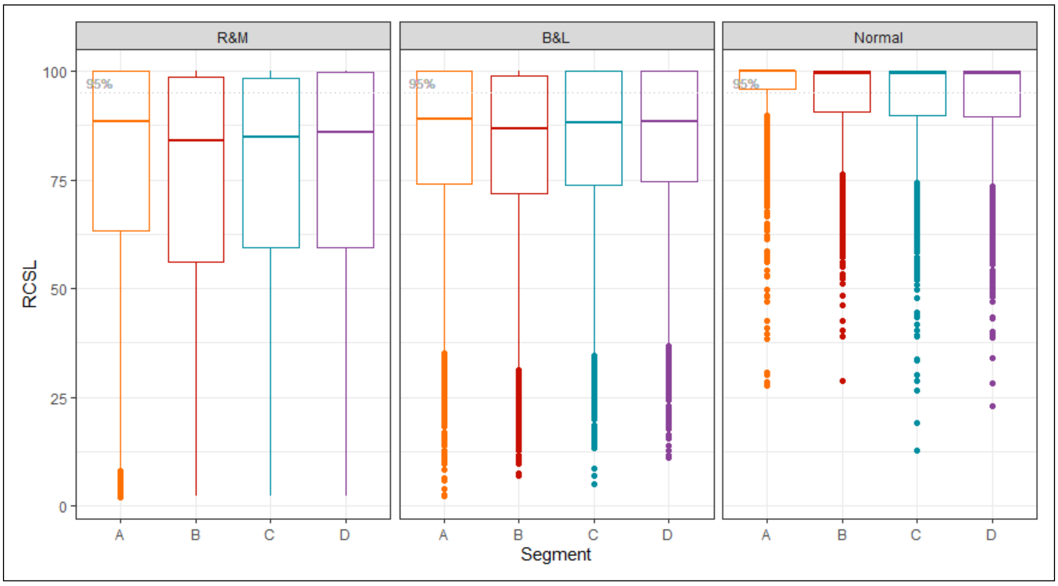

Figure 6.11: Boxplots for lead time CV segmented RCSL results of each method at 95 percent ICSL

As previously reported, the Normal method was the only one to reach a TCSL of 95 percent irrespective of lead time CV segment. In order to investigate the relationship between RCSL and RHC, and to establish whether the bootstrapping methods could achieve acceptable RCSLs in any of the lead time CV segments, the simulations were repeated for each method for a range of ICSLs (85, 87.5, 90, 92.5, 95, 97.5 and 99 percent). The results are shown in Figure 6.12 with additional graphical output seen in Appendix B, Figures B.5 & B.6 for RSOH and RFR versus RCSL which lead to the same conclusions as those drawn from Figure 6.12.

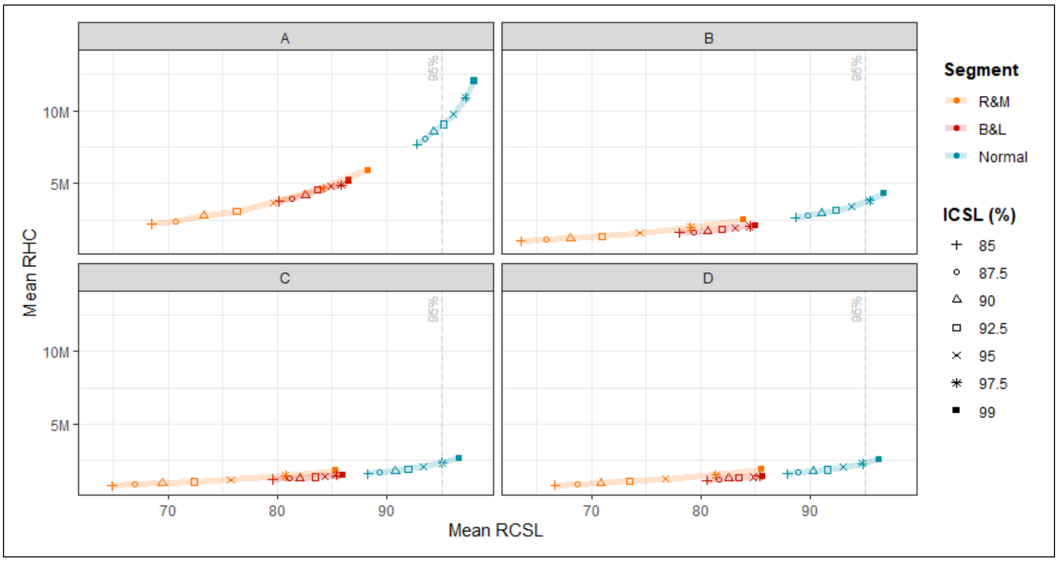

Figure 6.12: RHC versus mean RCSL for each method and lead time CV segment for multiple ICSLs

When compared to the maximum RCSL points achieved for the B&L and R&M methods (Figure 6.12), the Normal method achieved much better RCSLs at only slightly higher RHCs for segments B-D. This clearly indicated the superiority of the Normal method for all three of these segments (B-D). The "A" segment results again suggested that the Normal method was best in terms of RCSL. However, as expected the higher RCSLs were accompanied by higher RHCs. Unlike the results in both the high level and segmented demand sections, extrapolation from the Normal method curve did not suggest a large difference between the Normal method RHCs and those of the bootstrapping methods at a lower RCSL. Therefore, even at a lower TCSL it would not be recommended to use either of the bootstrapping methods, as these would only add complexity to the reorder point calculation with no corresponding reduction in RHCs.

Despite neither of the bootstrapping methods being recommended for use on any of the lead time CV segments, it was interesting to note that the R&M method outperformed the B&L method in terms of RCSL at a 99 percent ICSL for the "A" segment items. This was the only lead time CV segment where this occurred. A possible reason for this, which is unrelated to the lead time CV, was the demand characteristics of the items found in lead time CV segment "A". Upon further investigation it was found that 57 percent of the items in the "A" lead time CV segment were also classified as "Slow" items based on their demand characteristics. It was possible that the demand characteristics were in fact responsible for this behaviour. It was thus not surprising the R&M method achieved higher RCSLs than the B&L method at the highest ICSL, as this was also observed for the “Slow” segment.

This concludes the investigation into results obtained when items were segmented by lead time characteristics. In the section that follows the effect of outliers was investigated.

6.4 Outliers Included or Replaced

In the research conducted by do Rego and de Mesquita [2015] outliers were identified as a possible cause of the poor performance in the simulations. In order to assess their influence in the current study, identification and replacing of outliers in the training data was used to determine the relevant reorder points. The outlier detection and replacement method, described in Chapter 4, resulted in an overall decrease in total demand of 22 percent when compared to the original training data. In addition to this, an item level analysis revealed that 33 percent of items experienced a decrease in total demand, whilst the remaining 67 percent were unaffected. Automotive demand is inherently low volume and slow moving, leading to a large portion of zero demand occurrences. This in turn reduced the likelihood of small demand occurrences being identified as outliers. It was therefore not surprising to find that only high demand occurrences in the training data were identified as outliers. These outliers were replaced with the lower median value, resulting in the decrease seen in the total demand.

Simulations of the three methods were carried out using data with outliers replaced to determine inventory policy parameters and were compared to the results obtained on the original data (with no outlier replacement, as seen in Section 6.1). The relevant performance metrics were summarised in Table 6.6. The simulations made use of an ICSL of 95 percent. In addition to the tabulated results, Figures 6.13 and 6.14 illustrate simulation results, and focused on both the relationship between RHC and RCSL as well as the distribution of RCSL for each method and dataset combination.

The results from Table 6.6 and Figure 6.13 exposed the negative impact of the replacement of outliers, in terms of RCSL, on all three methods. That is, all methods achieved lower RCSLs in simulations carried out using the demand data with outliers replaced. It was expected that RCSL would decrease slightly as a result of the decrease in demand data, resulting from the correction of outliers, which would have resulted in decreased reorder points. It is logical that using lower reorder points on the same test data would result in lower RCSLs.

Table 6.6: Performance metric comparison for simulations using demand data with and without outliers replaced

Method	Outliers Replaced (Yes/No)	RSOH	RHC	RCSL (mean)	RCSL (median)	RFR
R&M	Yes	9	6042579	74	85	75
	No	13	7721253	77	87	77
B&L	Yes	15	7241366	81	87	82
	No	19	9552651	84	89	85
Normal	Yes	26	12265959	92	99	92
	No	38	17284260	94	100	94

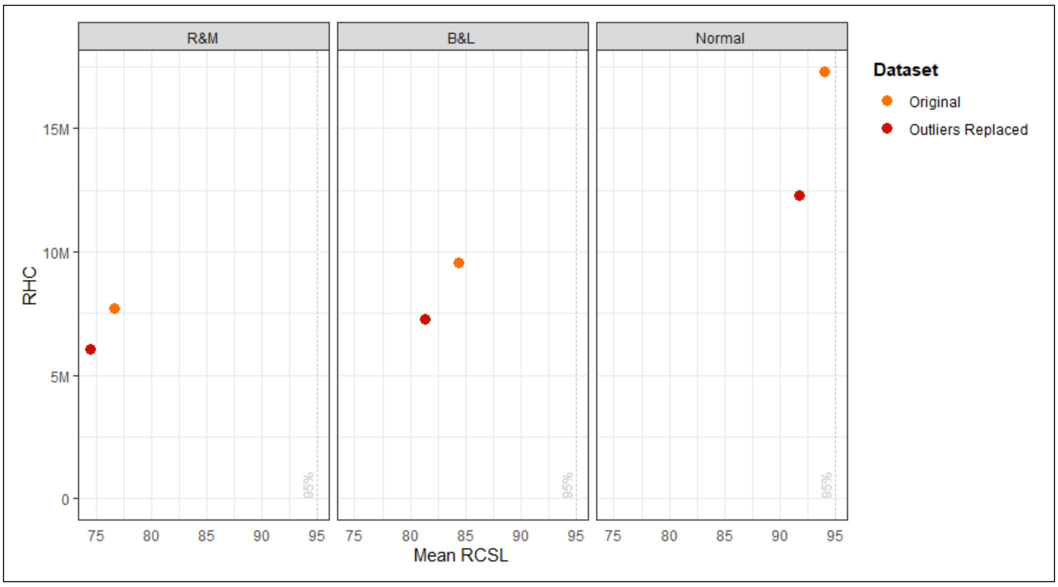

Figure 6.13: RHC and mean RCSL comparison between methods simulated using demand data including and replacing outliers at 95 percent ICSL

However, all three method simulations showed improved RHCs when carried out using demand data with replaced outliers. The difference between the RHC observations for each method was largest for the Normal method at over 5 million while the R&M method showed the smallest difference at about 1.67 million. In addition to the larger decrease in RHC, the Normal method also resulted in the smallest decrease in RCSL. This suggested that all results discussed in the previous sections would likely improve in terms of RHC with the correction of outliers, and that RCSL results would likely decrease slightly. Quades test was applied

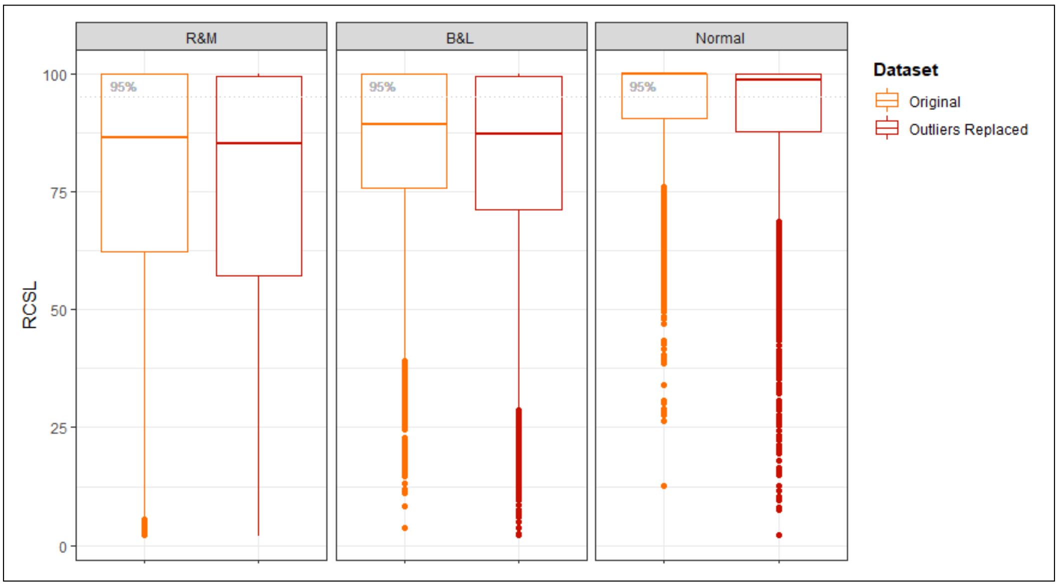

Figure 6.14: Boxplots of RCSLs for methods simulated using demand data including and replacing outliers at 95 percent ICSL

in order to confirm the validity of conclusions drawn from Table 6.6 and Figure 6.13. It tested for significant differences between the RCSL distributions of each simulation conducted, using the data with and without outliers replaced for each method. Similar to the high level results, data for this test was in the form of an unreplicated block design. The dependent variable was RCSL, the independent variable the dataset indicator, and item code the blocking variable. The hypotheses for the significance tests were as follows:

$$H_0 : \textit{The distribution of RCSLs for a given method are equal for the simulations using demand data with and without outliers replaced}$$

$$H_a : \textit{RCSL distributions are not equal}$$

From the results of this test (see Appendix A, Tables A.21 & A.22) it was concluded that RCSL distributions for all methods were significantly different from each other when comparing the method simulation which used the data with outliers replaced to the method simulation which used the original data set. This did not support the suggestion made by do Rego and de Mesquita [2015] that replacing outliers may improve the performance of the bootstrapping methods as the Normal method was less affected in terms of RCSL and had much better improvements in RHC.

This concludes the analysis and discussion of results for the 27 simulations conducted as part of this research. Concluding remarks on the results of each of the inventory policy methodologies simulated are provided in Chapter 7.

Chapter 7

Conclusion

The purpose of this study was to determine whether variations of the bootstrapping methods for reorder point calculation, originally presented by Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015], could match or outperform the standard industry approach (Normal approximation).

Methods were compared by means of an empirical study on data provided by a South African automotive spare parts business referred to as Company A. Several inventory policy simulations, for the period January 2015 to December 2015, were carried out for this purpose and each simulation summarised in the form of a number of inventory metrics. The primary metrics included RCSL, RHC, RFR and RSOH. The resulting metrics were discussed and presented in Chapter 6. In an attempt to provide a thorough investigation of the simulation results, the analysis and discussion first approached results from a high level perspective (i.e. over all items) and then for different item segments. The item segments were defined by different item demand and lead time characteristics.

The decision to adopt these particular reorder point calculation methods was based on three key factors. Firstly, each method was required to account for variable lead time. This was essential given the nature of the spare parts data utilised for the simulations. Both the Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] methods were two of only a few bootstrapping methods to consider lead time as a random variable. The Normal method also allowed for lead time to be modelled as random. Secondly, given the common perception that parametric methods were restrictive on the LTD distribution, both parametric and nonparametric approaches to the reorder point calculation were included for comparison. Given increasing computational power, nonparametric methods such as bootstrapping have become an increasingly viable option for practical implementation. Finally, at least one method was required to represent a standard industry approach to the reorder point calculation, which provided the base method for the comparison. The Normal approximation method was selected for this purpose. This particular method was chosen for its popularity in both practice and theory. Both Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] made use of the Normal approximation method as a comparison for their proposed methods, thus further supporting the inclusion of this method in the present study. In addition to this, the Normal approximation method was employed by Company A, from which the data used in the simulations were provided.

The method proposed by do Rego and de Mesquita [2015] built upon those proposed by Zhou and Viswanathan [2011] and Willemain et al. [2004]. The method implemented an extension for lead time as a random variable as well as an alternative approach to generating the inter demand arrival time sequence. Two further adjustments were made to the original do Rego and de Mesquita [2015] bootstrapping methodology for the purposes of the present study. The first concerned the service level measure, target fill rate, which was changed to CSL. The decision to use CSL was taken as it corresponded to the chosen service level measure for Company A and allowed the opportunity to determine whether the R&M method behaved differently under the influence of an alternative service level measure. The second adjustment to the original do Rego and de Mesquita [2015] bootstrapping methodology was that lead times were sampled from observed data rather than from a fitted triangle distribution.

The reorder point calculation method presented by Bookbinder and Lordahl [1989] provided the second nonparametric method compared in this research. This method was one of the first bootstrapping techniques for the reorder point calculation presented in literature. The method showed some positive results in the original study when LTD was generated from a "non-standard" distribution such as Bimodal. The study made use of simulated demand and lead time data. The key differences between the Bookbinder and Lordahl [1989] and do Rego and de Mesquita [2015] methods were the samples from which the LTD population distributions were generated. The Bookbinder and Lordahl [1989] technique sampled from the sample of LTD observations whereas the do Rego and de Mesquita [2015] method sampled from the demand, lead time and inter demand arrival time samples separately in order to generate LTD samples. The lack of literature on the performance comparison of these two methods when applied to industry data supported the inclusion of both in the current research.

The first key finding observed in both the high level and the segmented simulation results of this research was that, of the three methods under comparison, the Normal method had the best performance in terms of RCSL. The research of Porras and Dekker [2008], Zhou and Viswanathan [2011] and do Rego and de Mesquita [2015] supported this conclusion with all authors suggesting that the implemented parametric methods were superior in terms of service level. However, results from this study also found that the superiority of the Normal method, in terms of RCSL, was accompanied by high RHCs. This finding was supported by Zhou and Viswanathan [2011] who specifically noted the presence of higher total costs accompanying the superior realised service levels achieved by the parametric methods when using a simulated dataset. The high RHCs found in this study provided an opportunity for the B&L method to outperform the Normal method in terms of RHC, under certain conditions. More specifically, the high level results showed that, at RCSLs below 86 percent, the B&L method resulted in significantly lower RHCs when compared to the Normal method. do Rego and de Mesquita [2015] implicitly weighted the importance of the service level metric higher than costs in their comparison of reorder point calculation methods. That is, the authors favoured methods which resulted in higher realised service levels and only considered the realised cost when two or more methods under comparison had surpassed the target service level. Using this approach, despite the lower RHCs observed for the B&L method, the Normal method would be the obvious choice as it was the only method to consistently achieve the TCSL.

It is a common occurrence in empirical research that realised service level measures do not match the target values [Teunter and Duncan, 2009; do Rego and de Mesquita, 2015; Syntetos et al., 2015]. It was thus not surprising that simulation results in the present research exhibited this trait. However, the problem was much more severe in both bootstrapping methods, which exhibited RCSL values 11 to 18 percent lower than the target. Bootstrapping methods were only able to generate bootstrap samples which differed marginally from the observed data through the jittering procedure. This may have contributed to their inability to estimate the tail percentiles of the LTD distribution corresponding to high TCSLs. This may have been particularly pronounced for items with a limited number of demand values.

The R&M method, for the most part, showed an even greater difference between TCSL and RCSL than the B&L method. The R&M method was therefore found to be inferior to both the Normal and B&L methods. This conclusion followed from the generally poor RCSL and high RHC results of this method, which meant that it was unable to outperform the other methods in any of the simulation results. The poor results shown for the R&M method under the CSL metric suggested that it may be more suited to the target fill rate metric as presented by do Rego and de Mesquita [2015], as in their study the method showed promising results for some item segments. There may also have been a dependence between the demand quantities and the inter demand arrival times, which the B&L method was better able to capture by calculating LTD using demand quantities and their adjacent inter demand arrival times. The R&M method would not have captured this dependence as it sampled separately from demand quantity and inter demand arrival time. Intuitively, there may be dependence between the demand quantity and inter demand arrival times, as during periods of high demand it would be expected that both customer order quantity and customer order frequency would increase.

Conclusions based on the segmented results as defined by Syntetos et al. [2005] showed similarities to those of do Rego and de Mesquita [2015] in terms of identifying segments where a parametric approach was most appropriate, but differed in identifying segments where the R&M bootstrapping method was most appropriate. do Rego and de Mesquita [2015] recommended that parametric distributions be used for modelling LTD for items belonging to the "Erratic" and "Slow" segments with the exception of the "Slow" segment at a high target fill rate for which the bootstrapping approach was recommended. As previously mentioned, these conclusions were based on both the realised service level and costs observed for each method. The present research also found the parametric method (Normal approximation) to be superior for these two segments but did not find the R&M method to be superior for the "Slow" segment at any TCSL. However, it was noted that the "Slow" segment showed the most promising results of the four segments in terms of RCSL when considering the R&M method on its own. For the remaining two segments, do Rego and de Mesquita [2015] suggested that the bootstrapping method was superior when implemented for "Lumpy" items at all target fill rates and "Smooth" items with target fill rates below 90 percent. This was not found to be true in the simulation results of the R&M method in the present study. However, the B&L method did, to some extent, exhibit this behaviour for "Smooth" items. That is, at RCSLs below 81 percent, the B&L method was able to achieve lower RHCs for the same RCSL as the Normal method. The B&L method was thus preferable at these lower RCSLs for both high level and "Smooth" segment results. Although extrapolation suggested that the B&L method was able to achieve similar RHCs at RCSL below 85 for "Lumpy" items, the methods inability to achieve RCSL above 85 and its added complexity lead to the Normal method being chosen as the preferred method for the "Lumpy" segment items.

The parametric approaches recommended by do Rego and de Mesquita [2015] above utilised different distributions (such as Gamma and Negative Binomial) to the Normal approach investigated and recommended in the current study. It is interesting that the segments identified as most appropriate for parametric methods by do Rego and de Mesquita [2015] did not change with the different parametric method utilised in this study.

Conclusions based on the segmentation defined by demand volume (as used by Company A) were similar to those drawn from the Syntetos et al. [2015] segmentation. That is, the B&L method resulted in lower RHCs than the Normal method at lower service levels for the high volume items belonging to the "A" segment. For the three remaining segments ("B", "C" and "D"), the Normal method was found to be superior to the bootstrapping methods, in terms of RCSL, with the exception of the "D" segment which showed very similar results for the R&M and Normal methods (see Figure 6.9). This was an interesting result for the R&M method as it was the first time that either of the bootstrapping methods was able to reach the TCSL of 95 percent. However, the R&M method resulted in higher RHCs than the Normal method to achieve this.

Results were also compared for item segments defined by lead time variability. These results showed that lead time variability did not have a notable affect on method recommendations and was hence not considered in determining the superior reorder point calculation method.

In addition to the simulations which made use of the original LTD data from Company A, three further simulations were conducted using demand data with outliers replaced. Outliers were replaced by median demand values for each item and were identified by means of the MAD method described in Chapter 4. do Rego and de Mesquita [2015] expected that the exclusion of outliers from the training data would improve their method's results and hence the reason for the additional simulations. However, the results from this current research did not support their hypothesis. In fact the particular outlier treatment used in this research resulted in improved results for the Normal method which exhibited lower RHCs for a small decrease in RCSL, which was still higher than those of the bootstrapping methods. The improvement in RHCs obtained through outlier detection and replacement was a promising result and applying outlier correction techniques to other parametric approaches proposed in literature is an area of potential future research.

Parametric methods such as the Normal method require a forecast of mean demand. An automated ARIMA process was used for this purpose in the current research. The ARIMA models were able to take into account both seasonality and trend in the data. The bootstrapping methods did not take seasonality or trend into account. While automotive spare parts data are not generally seasonal to the extent of, for example, retail data, the superior performance of the Normal method with ARIMA demand forecasts may indicate some seasonality in the data. Additionally, with the selection of data to ensure items were in the mature stage of their lifecycle, it was not originally believed that trend would have a great impact on the items under investigation. However, as indicated by the superior results of the Normal method, this may not be the case. An investigation into the presence of trend and seasonality in the data is an area of future research. A possible difference in seasonality and trend attributes of the dataset may have contributed to the difference in results obtained between this study and that of do Rego and de Mesquita [2015]. Another potential area for future research would be to adapt the bootstrapping methods to take into account trend and seasonality.

While the Normal method was the only method evaluated in this study to achieve acceptably high service levels, a drawback to this method was the sensitivity of RHCs to minor changes in TCSL, especially at higher service level values. This factor should be taken into consideration by practitioners when adjusting service levels upwards. The B&L method was found to be the least sensitive to changes in TCSL.

This study was able to thoroughly compare the results of the three chosen reorder point calculation methods, thereby satisfying the objective of the research. The most valuable practical outcomes of this research were that the Normal method performed the best in terms of RCSL, while the B&L method had some cost saving advantages at lower RCSLs. The R&M method did not exhibit any notable advantage over either of the other methods. Before concluding on performance of this method, it is recommended that further research be conducted, focusing on the effect of a change in service level metric on the performance of the R&M method. The effect of trend and seasonality on the results should also be assessed.

Although the further research suggested would be beneficial based on the findings of this research, practitioners could, in a business where customers are willing to accept lower RCSLs, potentially experience cost savings with the implementation of the B&L method. However, in a business where high RCSLs were essential to customer satisfaction, the Normal method would be more suitable. When considering method complexity, both bootstrapping methods presented more computationally intensive and complex options for the reorder point calculation. Thus despite the potential cost savings of the B&L method at lower RCSLs, practitioners should implement this method with caution and only under the close advisement of an experienced statistician with the ability to adjust the methodology to suit the business objectives.

Bibliography

M. Babai and A. Syntetos. An empirical comparison of parametric and nonparametric forecasting methods for intermittent demand items. In *Proceedings of the British Operational Research Conference (OR49), Edinburgh*, 2007.

J. Bookbinder and A. Lordahl. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions*, 21(4):302–312, 1989.

C. Caplice. CTL.SC1x Supply chain fundamentals. MOOC offered by Massachusetts Institute of Technology, 2017. URL `https://www.edx.org/course/supply-chain-fundamentals-3`.

S. Chopra, G. Reinhardt, and M. Dada. The effect of lead time uncertainty on safety stocks. *Decision Sciences*, 35(1):1–24, 2004.

B. R. Cobb. Mixture distributions for modelling demand during lead time. *Journal of the Operational Research Society*, 64(2):217–228, 2013.

B. R. Cobb, A. W. Johnson, R. Rumí, and A. Salmerón. Accurate lead time demand modeling and optimal inventory policies in continuous review systems. *International Journal of Production Economics*, 163:124–136, 2015.

J. D. Croston. Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 23(3):289–303, 1972.

L. Das. *The impact of bimodal distribution in ocean transportation transit time on logistics costs: An empirical and theoretical analysis*. PhD thesis, Massachusetts Institute of Technology, 2013.

J. R. do Rego and M. A. de Mesquita. Demand forecasting and inventory control: A simulation study on automotive spare parts. *International Journal of Production Economics*, 161:1–16, 2015.

O. J. Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.

B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, pages 1–26, 1979.

R. Ehrhardt. The power approximation for computing (s, S) inventory policies. *Management Science*, 25(8):777–786, 1979.

S. Emmett and D. Granville. *Excellence in inventory management: How to minimise costs and maximise service*. Cambridge Academic, 2007.

G. D. Eppen and R. K. Martin. Determining safety stock in the presence of stochastic lead time and demand. *Management Science*, 34(11):1380–1390, 1988.

D. Erlenkotter. Ford Whitman Harris and the economic order quantity model. *Operations Research*, 38(6):937–946, 1990.

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

C. K. Gudum. *Managing variability in a supply chain: An inventory control perspective*. PhD thesis, Copenhagen Business School, 2002.

G. Hadley and T. Whitin. *Analysis of inventory systems*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

Z. Hua, B. Zhang, J. Yang, and D. Tan. A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society*, 58(1):52–61, 2007.

R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. forecast: Forecasting functions for time series and linear models, 2019. URL `http://pkg.robjhyndman.com/forecast`. R package version 8.7.

M. Krever, S. Wunderink, R. Dekker, and B. Schorr. Inventory control based on advanced probability theory, an application. *European Journal of Operational Research*, 162(2):342–358, 2005.

W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

R. S. Kumar and A. Goswami. A continuous review production–inventory system in fuzzy random environment: Minmax distribution free procedure. *Computers & Industrial Engineering*, 79:65–75, 2015.

C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

S. Mangiafico. *Summary and Analysis of Extension Program Evaluation in R, version 1.18.1.* 2016. URL `rcompanion.org/handbook/`.

I. Moon and G. Gallego. Distribution free procedures for some inventory models. *Journal of the Operational Research Society*, 45(6):651–658, 1994.

P. Narayan and J. Subramanian. *Inventory Management-principles and Practices.* Excel Books India, 2009.

G. Nenes, S. Panagiotidou, and G. Tagaras. Inventory management of multiple items with irregular demand: A case study. *European Journal of Operational Research*, 205(2):313–324, 2010.

D. H. Ogle, P. Wheeler, and A. Dinno. *FSA: Fisheries Stock Analysis*, 2019. URL `https://github.com/droglenc/FSA`. R package version 0.8.24.

E. Porras and R. Dekker. An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *European Journal of Operational Research*, 184(1):101–132, 2008.

D. Quade. Using weighted rankings in the analysis of complete blocks with additive block effects. *Journal of the American Statistical Association*, 74(367):680–683, 1979.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, 2018. URL `https://www.R-project.org/`.

A. J. Ruiz-Torres and F. Mahmoodi. Safety stock determination based on parametric lead time and demand information. *International Journal of Production Research*, 48(10):2841–2857, 2010.

E. Silver and R. Peterson. *Decision systems for inventory management and production planning*. Wiley, 1985.

A. A. Syntetos, J. E. Boylan, and J. Croston. On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5):495–503, 2005.

A. A. Syntetos, M. Z. Babai, and E. S. Gardner Jr. Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research*, 68(8):1746–1752, 2015.

R. H. Teunter and L. Duncan. Forecasting intermittent demand: A comparative study. *Journal of the Operational Research Society*, 60(3):321–329, 2009.

J. W. Tukey. Box-and-whisker plots. *Exploratory data analysis*, pages 39–43, 1977.

J. E. Tyworth and L. O'Neill. Robustness of the normal approximation of lead-time demand in a distribution setting. *Naval Research Logistics (NRL)*, 44(2): 165–186, 1997.

B. Vernimmen, W. Dullaert, and S. Engelen. Schedule unreliability in liner shipping: Origins and consequences for the hinterland supply chain. *Maritime Economics & Logistics*, 9(3):193–213, 2007.

C. Waters. *Inventory Control and Management*. Wiley, 2003.

T. R. Willemain, C. N. Smart, and H. F. Schwarz. A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20(3):375–387, 2004.

R. Wilson. *A scientific routine for stock control*. Harvard Univ., 1934.

C. Zhou and S. Viswanathan. Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems. *International Journal of Production Economics*, 133(1):481–485, 2011.

Appendix A

Statistical Significance Test Results

A.1 Kruskal-Wallis and Dunn Test Results

Table A.1: Results of Kruskal-Wallis test for significant differences between Syntetos et al. [2005] segment distributions for the R&M method

Kruskal-Wallis chi-squared	df	p-value
1088.2	3	$<$ 2.2e-16

Table A.2: Results of the Dunn test comparing Syntetos et al. [2005] segment distributions for the R&M method

Comparison	Z	P unadj.	P adj.
Erratic-Lumpy	-2.415387	1.571848e-02	2.357773e-02
Erratic-Slow	-19.351722	1.971445e-83	3.942889e-83
Lumpy-Slow	-19.929913	2.239522e-88	6.718567e-88
Erratic-Smooth	-0.756193	4.495335e-01	4.495335e-01
Lumpy-Smooth	2.075411	3.794848e-02	4.553817e-02
Slow-Smooth	25.590101	1.966186e-144	1.179712e-143

Table A.3: Results of Kruskal-Wallis test for significant differences between Syntetos et al. [2005] segment distributions for the B&L method

Kruskal-Wallis chi-squared	df	p-value
204.87	3	$<$ 2.2e-16

Table A.4: Results of the Dunn test comparing Syntetos et al. [2005] segment distributions for the B&L method

Comparison	Z	P unadj.	P adj.
Erratic-Lumpy	-1.156733	2.473816e-01	2.473816e-01
Erratic-Slow	-4.172986	3.006335e-05	4.509502e-05
Lumpy-Slow	-3.366855	7.603075e-04	9.123689e-04
Erratic-Smooth	5.315869	1.061496e-07	2.122992e-07
Lumpy-Smooth	7.533836	4.927088e-14	1.478126e-13
Slow-Smooth	14.161590	1.583710e-45	9.502258e-45

Table A.5: Results of Kruskal-Wallis test for significant differences between Syntetos et al. [2005] segment distributions for the Normal method

Kruskal-Wallis chi-squared	df	p-value
267.96	3	$<$ 2.2e-16

Table A.6: Results of the Dunn test comparing Syntetos et al. [2005] segment distributions for the Normal method

Comparison	Z	P unadj.	P adj.
Erratic-Lumpy	-1.506968	1.318188e-01	1.318188e-01
Erratic-Slow	-5.329523	9.847114e-08	1.477067e-07
Lumpy-Slow	-4.256573	2.075844e-05	2.491012e-05
Erratic-Smooth	5.520315	3.383925e-08	6.767850e-08
Lumpy-Smooth	8.196549	2.473860e-16	7.421579e-16
Slow-Smooth	16.084779	3.262149e-58	1.957289e-57

Table A.7: Results of Kruskal-Wallis test for significant differences between demand volume segment distributions for the R&M method

Kruskal-Wallis chi-squared	df	p-value
1819.6	3	$<$ 2.2e-16

Table A.8: Results of the Dunn test comparing demand volume segment distributions for the R&M method

Comparison	Z	P unadj.	P adj.
A - B	-14.558040	5.192564e-48	6.231076e-48
A - C	-29.546189	7.350709e-192	2.205213e-191
B - C	-15.232245	2.160445e-52	3.240667e-52
A - D	-39.759708	0.000000e+00	0.000000e+00
B - D	-25.027883	3.040114e-138	6.080227e-138
C - D	-9.186835	4.045702e-20	4.045702e-20

Table A.9: Results of Kruskal-Wallis test for significant differences between demand volume segment distributions for the B&L method

Kruskal-Wallis chi-squared	df	p-value
217.51	3	$<$ 2.2e-16

Table A.10: Results of the Dunn test comparing demand volume segment distributions for the B&L method

Comparison	Z	P unadj.	P adj.
A - B	-7.737605	1.013068e-14	2.026137e-14
A - C	-11.724976	9.492844e-32	2.847853e-31
B - C	-4.112480	3.914310e-05	4.697172e-05
A - D	-13.572821	5.804420e-42	3.482652e-41
B - D	-5.733521	9.836676e-09	1.475501e-08
C - D	-1.473072	1.407317e-01	1.407317e-01

Table A.11: Results of Kruskal-Wallis test for significant differences between demand volume segment distributions for the Normal method

Kruskal-Wallis chi-squared	df	p-value
275.13	3	$<$ 2.2e-16

Table A.12: Results of the Dunn test comparing demand volume segment distributions for the Normal method

Comparison	Z	P unadj.	P adj.
A - B	-6.576911	4.803231e-11	7.204847e-11
A - C	-12.653796	1.065975e-36	3.197924e-36
B - C	-6.186353	6.157206e-10	7.388647e-10
A - D	-15.149281	7.659977e-52	4.595986e-51
B - D	-8.490388	2.059477e-17	4.118954e-17
C - D	-2.083601	3.719645e-02	3.719645e-02

Table A.13: Results of Kruskal-Wallis test for significant differences between lead time CV segment distributions for the R&M method

Kruskal-Wallis chi-squared	df	p-value
46.373	3	4.726e-10

Table A.14: Results of the Dunn test comparing lead time CV segment distributions for the R&M method

Comparison	Z	P unadj.	P adj.
A - B	6.372502	1.859688e-10	1.115813e-09
A - C	5.263492	1.413444e-07	4.240331e-07
B - C	-1.109562	2.671877e-01	2.671877e-01
A - D	3.817330	1.349039e-04	2.698077e-04
B - D	-2.551949	1.071223e-02	1.606835e-02
C - D	-1.443168	1.489730e-01	1.787676e-01

Table A.15: Results of Kruskal-Wallis test for significant differences between lead time CV segment distributions for the B&L method

Kruskal-Wallis chi-squared	df	p-value
15.535	3	0.001412

Table A.16: Results of the Dunn test comparing lead time CV segment distributions for the B&L method

Comparison	Z	P unadj.	P adj.
A - B	3.7765150	0.000159038	0.000954228
A - C	1.4403903	0.149757011	0.224635517
B - C	-2.3360059	0.019490931	0.038981863
A - D	0.9180755	0.358579356	0.430295228
B - D	-2.8561721	0.004287827	0.012863480
C - D	-0.5214786	0.602033394	0.602033394

Table A.17: Results of Kruskal-Wallis test for significant differences between lead time CV segment distributions for the Normal method

Kruskal-Wallis chi-squared	df	p-value
183.85	3	$<$ 2.2e-16

Table A.18: Results of the Dunn test comparing lead time CV segment distributions for the Normal method

Comparison	Z	P unadj.	P adj.
A - B	10.6762058	1.315385e-26	2.630771e-26
A - C	11.0582208	2.000260e-28	6.000780e-28
B - C	0.3804938	7.035789e-01	8.442947e-01
A - D	11.4219798	3.247505e-30	1.948503e-29
B - D	0.7498405	4.533508e-01	6.800261e-01
C - D	0.3695983	7.116818e-01	7.116818e-01

A.2 Quade Test

Table A.19: Results of Quade test for significant differences between method RCSL distributions

Quade F	df	p-value
4351.8	2	$<$ 2.2e-16

Table A.20: Results of the post-hoc Quade test for significant differences between RCSL distributions

	R&M	B&L
B&L	$<$ 2e-16	-
Normal	$<$ 2e-16	$<$ 2e-16

Table A.21: Results of Quade test for significant differences between RCSL distributions for methods implemented with & without outlier treatment

Quade F	df	p-value
180.69	1	$<$ 2.2e-16

Table A.22: Results of the post-hoc Quade test for significant differences between RCSL distributions for methods implemented with & without outlier treatment

	Original data
Outliers replaced data	$<$ 2.2e-16

Appendix B

Additional Graphical Output

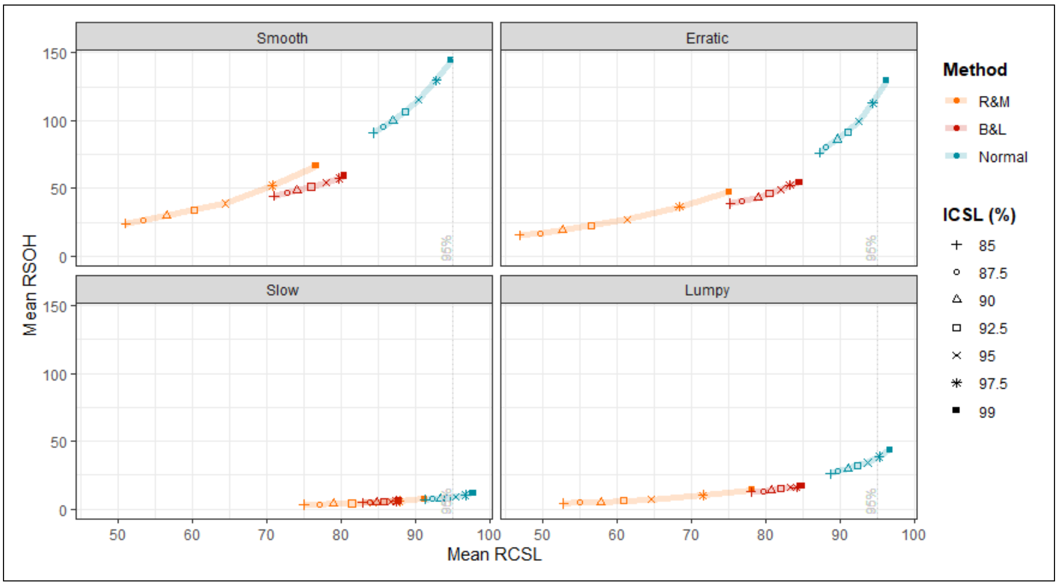

Figure B.1: Mean RSOH versus mean RCSL for each Syntetos et al. [2005] segment and method combination

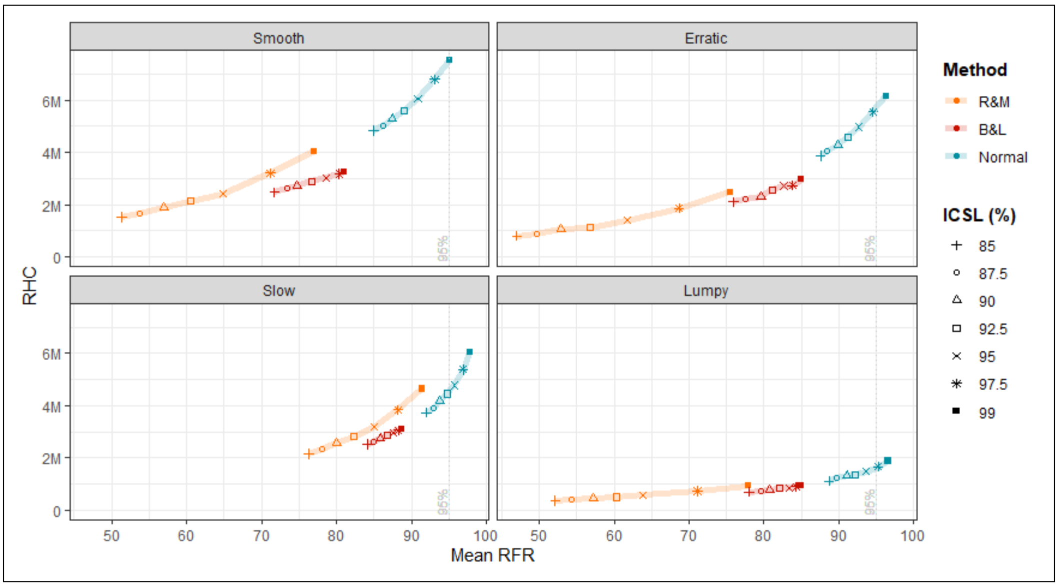

Figure B.2: RHC versus mean RFR for each Syntetos et al. [2005] segment and method combination

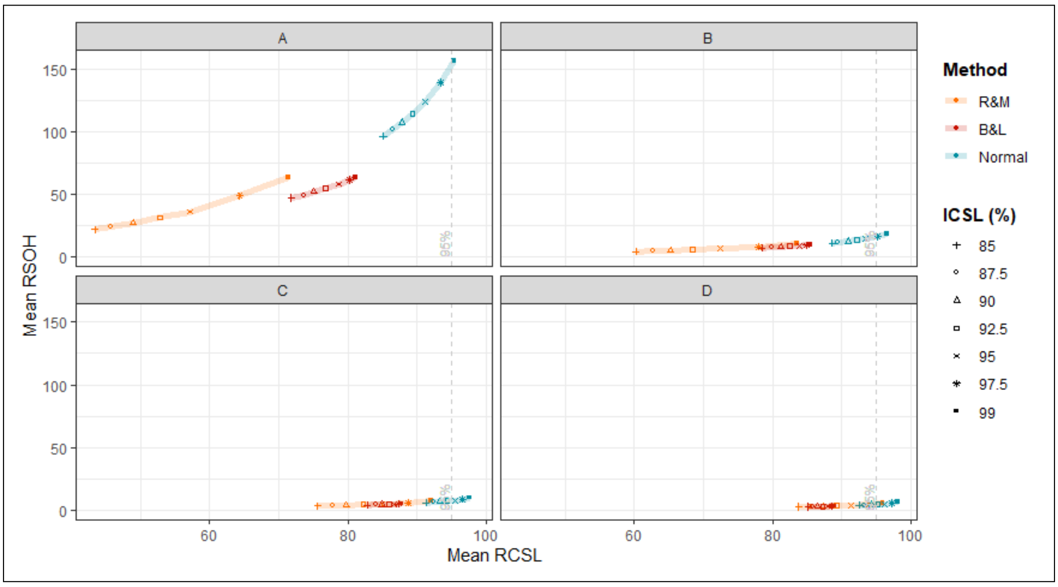

Figure B.3: Mean RSOH versus mean RCSL for each demand volume segment and method combination

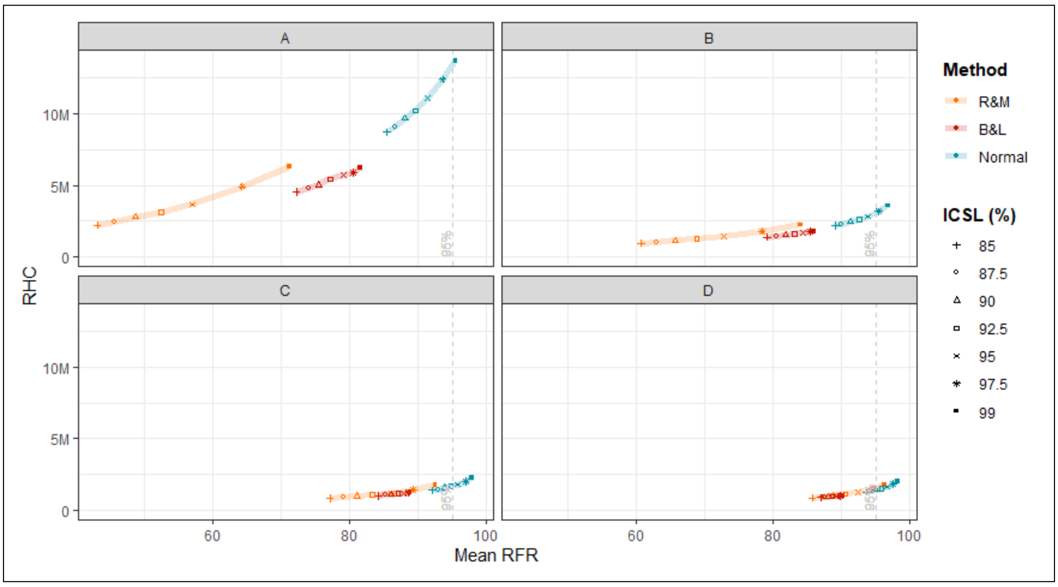

Figure B.4: RHC versus mean RFR for each demand volume segment and method combination

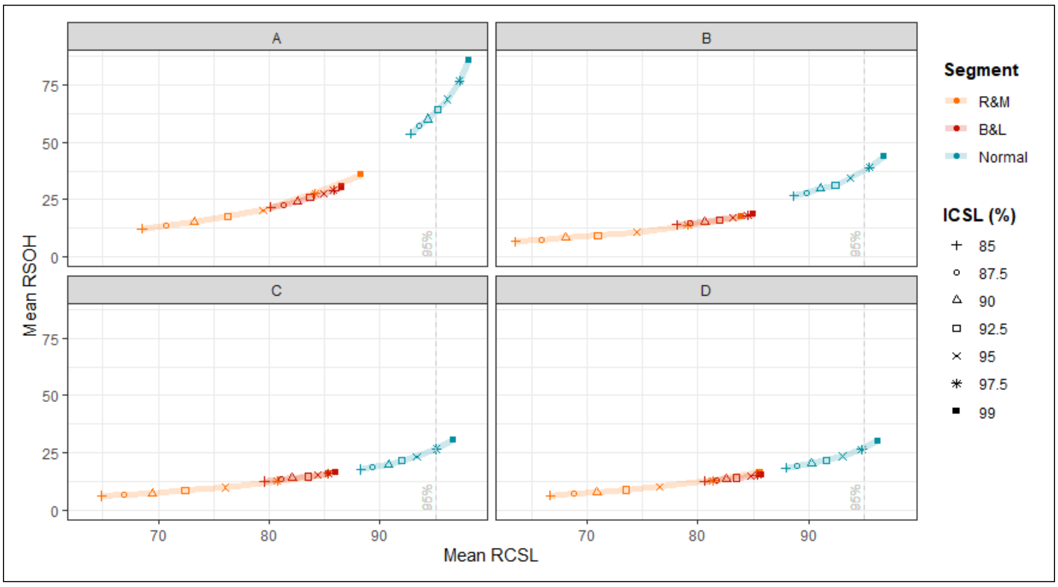

Figure B.5: Mean RSOH versus mean RCSL for each lead time CV segment and method combination

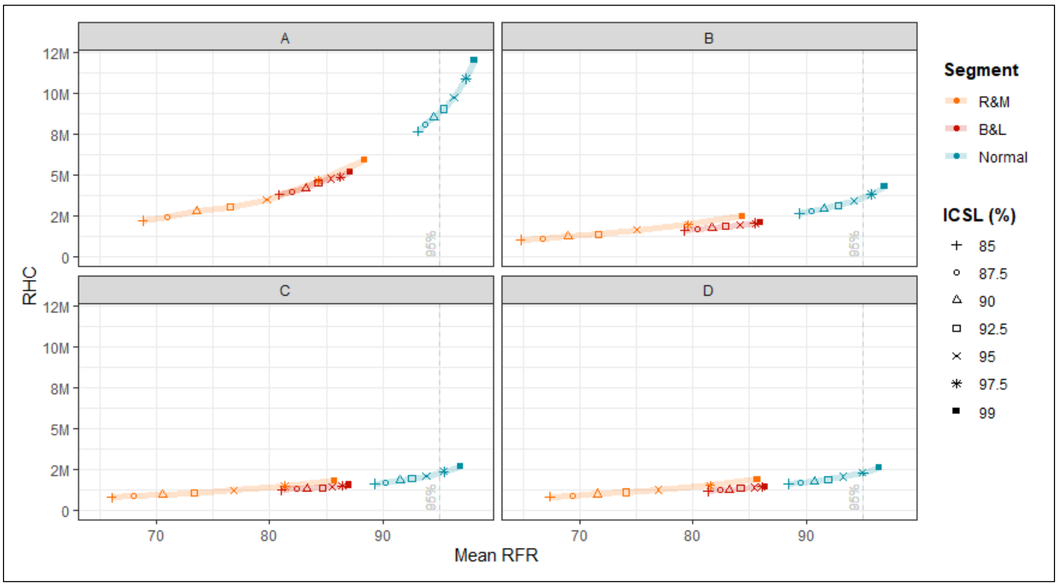

Figure B.6: RHC versus mean RFR for each lead time CV segment and method combination