



Estimation of expected stockout delays in divergent multi-echelon inventory systems using simulation

David Svanholm and Jacob Boissonnault

Department of Production Management
Lund University LTH

Supervisor: Peter Berling

Examiner: Johan Marklund

MIOM01/MIOM2 - Degree Project in Production Management

Thesis date: May, 2026

Acknowledgments

We would like to extend our thanks to Peter Berling for being our supervisor for this thesis. You have been a great source of knowledge, guidance, and help. We would also like to thank Johan Marklund for giving access to a multi-echelon ExtendSim model, assisting with its modifications, and giving limited access to ongoing research. Finally, we would like to thank our families for supporting us along the way.

I (Jacob Boissonnault) dedicate this thesis to my son, Henry, as a representation of all the hard work I strive to do for him.

Abstract

Stockout delay is an important component of lead time calculations in inventory control literature. This thesis presents a deep dive into the topic through a literature review, which is accompanied by a numerical study. This numerical study provides a comparison of available models along with a simulation model that looks at two different order fulfillment types: complete and partial. Based on the findings, we present an analysis of the results that focus on the accuracy of the models and how different variables/parameters interact. We present future areas of research that were identified as gaps from the thesis along with examining ongoing research.

Keywords: stockout delay, delay, multi-echelon, complete delivery, partial delivery, order fulfillment, simulation, lead time

Contents

1	Introduction	5
2	Methodology	7
2.1	Literature Review Approach	7
2.2	Simulation Study Approach	7
3	Literature Review	8
3.1	Lead Time and Stockout Delay in Literature	8
3.2	Gaps in Literature	9
3.3	Overview of Available models	10
3.3.1	General Inventory Control Theory	10
3.3.2	The METRIC Approach	12
3.3.3	Axsäter Normal Approximation Model	15
3.3.4	Kiesmüller & de Kok Mixed-Erlang Distribution Model	17
3.3.5	Berling & Farvid Lead Time Model	20
3.3.6	Grob & Bley Wait Time Approximation Model	24
4	Simulation Study and Model Overview	26
4.1	Simulation Model	26
4.2	Model Verification	27
4.3	Experimental Design	28
5	Results and Analysis	31
5.1	Results	31
5.2	Analysis	32
5.2.1	Model comparison and Ranking	34
6	Future research	37
6.1	Ongoing Research	37
7	Conclusion	38
	References	39
A	Model Code for the Tested Models	41
B	ExtendSim Model Description	51
C	Complete Result Tables and Result Calculations	60

1 Introduction

Multi-echelon systems are a key component in inventory control theory. Research in this area has produced many different types of models over the years to try and capture how these systems behave and how to optimize certain parameters. An important component for any model is the lead time. The lead time consists of two components: transportation time and delays due to shortages upstream. This thesis will focus on the delay component, which is a policy dependent stochastic variable. More specifically, the main problem that lead time effects is the service levels within the multi-echelon system. When the lead time is stochastic, it is difficult to exactly calculate the effects on a given service level. The stochastic lead time has two primary causes: external and internal. External causes are a broad category and mostly fall outside the control of the system. Transportation time is sometimes used to represent this. An internal cause is represented as the previously mentioned stockout delay. The stochastic nature of the lead time stems from this unpredictability. This is why developing an estimate for this variable is so important. The estimate of the delay gives any model a way to represent this otherwise difficult variable to calculate. Figure 1 is an example of a multi-echelon system.

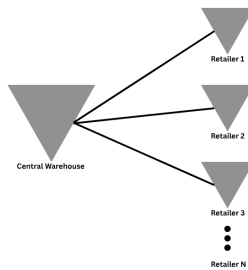


Figure 1: A Multi-Echelon Supply Chain System

In practice, it seems rather straightforward observe the lead time manually through data collection. Taking into account cost, this approach would be expensive and labor intensive, which makes it impractical for larger systems. This is an unnecessary cost when mathematical models exist that can assist in predicting the lead time and determining a policy to reduce it. Moreover, mathematical models allows you to predict how a system behaves and functions, providing valuable information that can be used to achieve certain goals or optimization. When the lead time is viewed in a theoretical sense, the lead time is represented by a stochastic variable due to variability of something called delay. For this thesis, when we mention delay, we are referring to stockout delay. A definition of this varies from each paper, but in this thesis, stockout delay refers to a delay from the central warehouse caused by having insufficient stock to cover an order from a retailer, causing a backorder. The value of the delay represents the time interval from when the order is placed to when it is delivered. This also raises some problems based on the order fulfillment policy. A complete order fulfillment requires the order to be shipped in full to a retailer while a partial fulfillment policy will ship an order with whatever stock on hand is available. This raises questions on how to define the interval in the case of partial fulfillment. For this thesis, we define complete order fulfillment as the time interval it takes for an order to be received at the central warehouse, filled, then received by the retailer. There are two different ways to define partial order fulfillment. The first way, which we will use in this thesis, is based on a unit basis. Therefore, the time interval is the same as in complete order fulfillment, but each order is broken down into individual units. The second measure is similar to the complete order fulfillment in that it measures the entire interval together until it is delivered in full. The difference here is that some of the delivery may be already delivered, but the time internally is measuring until the last item in that delivery is delivered.

It is important to recognize why delay is important. All multi-echelon inventory models factor in delay with an approximation because an exact calculation is not feasible. This raises important questions about the trade-off between the accuracy of these approximations and their computational tractability. It also raises the question of what stockout delay does to the entire multi-echelon system. What variables have the most impact on delay? How much does delay influence fill rate or service level? How does customer demand tie into this delay? These are all questions that make studying stockout delay interesting and important.

This thesis primarily focuses on this stockout delay and how it is calculated. This stochastic variable has been noted by a few paper as either extremely difficult or outright impossible due to the variability (Berling and Farvid [2014](#) & Sherbrooke [1968](#)). Our goal is conduct a numerical study to test various approximations found through our literature review along with a simulation numerical study that compares both complete and partial order fulfillment. Our main research questions to investigate are:

How do different system variables affect the stockout delay under complete and partial order fulfillment?

Which of the existing models available in research perform the best?

These questions look to observe and analyze a fundamental component to lead time calculations and add to the limited pool of dedicated stockout delay research.

2 Methodology

In this methodology section, we outline our approaches and the reasoning behind our decisions. This applies both to the literature review chapter and the simulation study chapter. Due to this thesis falling under operations research, we structure our overall approach to Hillier and Lieberman (2015). They outline it in 7 steps: "1: Defining the Problem and Gathering Data, 2: Formulating a Mathematical Model, 3: Deriving Solutions from the Model, 4: Testing the Model, 5: Preparing to apply the model, 6: Implementation, and 7: Conclusions (Hillier and Lieberman 2015, pp. 7–22)." These steps provided an outline for our approach for this thesis.

While we follow these steps as an outline, our approach diverges in a few steps. The first step is followed similarly. We have defined our research question and will conduct a literature review to gather models that we wish to test and examine. This problem, as stated in the introduction chapter, is to examine and determine lead time delays. The second and third step diverges in this thesis. We are not developing a mathematical model, rather evaluating already existing mathematical models. We rely on earlier models and a simulation model to gather information about stockout delays. Our contribution is providing a comparison point that lacks in existing research and examines an area lacking in stockout delay research, partial order fulfillment. The available literature is a component of a larger model, while we dedicate this thesis only to stockout delay. For the Fourth step, we do a numerical study that tests the models stockout delay calculations found within the literature review. Steps 5 and 6 are not included in this thesis due to this thesis being primarily rooted in examining earlier theory. However, we will present further avenues for future research and a ranking of the models examined based on a variety of factors. The final step is fairly self-explainable, but all conclusions from the numerical study will be examined and analyzed.

2.1 Literature Review Approach

Our general approach to this literature review is to use the "citation pearl growing" search strategy. See Hawkins and Wagers (1982) for a detailed breakdown of this method. This involves using a key paper to forward and back search it to gather a wide range of similar and related articles. This process can continue to branch out to collect related papers from the forward or back search. Berling and Farvid (2014) was our key paper in our case. It is the foundation to this thesis and provided a well cited trove of papers to conduct this literature review. This method was chosen due to the limited number of literature on stockout delays.

2.2 Simulation Study Approach

The approach of the simulation study is heavily inspired by Grob and Bley (2018). They use a 3-factor experimental design to evaluate both their own model and three other similar models. The important thing to note here is that they compared based on the concept of delay and its variations. Two of the models follow Tempelmeier (1983) and his waiting time theory to calculate delay. The other two calculate delay with expected backorders. Detailed explanations are given in Chapter 4. Our main motivation for this choice is the comparability between the models. Our planned study includes all four of the models Grob and Bley (2018) use. The second motivation is that the study conducted can be perceived as unfair. Grob and Bley (2018) use a stochastic transportation time to conduct their study. However, Grob and Bley (2018) model is the only model that uses a stochastic transportation time in its calculations. So, in our case, choosing to use a fixed transportation time provides a clearer picture as the majority of the models (4 out of 5) use this in their calculations.

3 Literature Review

Multi-echelon systems hold an important role and definition in Inventory Management literature. The echelons within these systems have a delay associated with it, which will or may be caused by stockouts at the central warehouse. Most literature focuses on the optimization of certain parameters in their models, which tends to focus on an optimal Reorder point R . See Axsäter (2006) for a compilation of different multi-echelon optimization methods and types. The condition or focus on lead time from the central warehouse to the retailer is an overlooked portion in these models due to the difficulty of its computation. To overcome this, models will approximate the delay and the resulting lead time. As stated previously, our goal is to examine the stockout delay that can cause this lead time to be stochastic. Berling and Farvid (2014) examined this in great detail and provides the inspiration to this thesis and application to the models developed in Berling and J. Marklund (2013) and Berling and J. Marklund (2014). This literature review focuses specifically on examining the different contributions to approximating stockout delay in sections 3.1 and 3.2, then providing a deep explanation of five approaches that are investigated in the simulation study.

3.1 Lead Time and Stockout Delay in Literature

From a historical perspective, the common approach for approximating lead time and stockout delay stems from the METRIC model (Sherbrooke 1968). The METRIC model replaces the stochastic lead time with a deterministic lead time that is assumed the same for all retailers, which has the effect of decreasing lead time for all retailers when stock at the central warehouse is increased (Sherbrooke 1968). The METRIC model accounts for delay by defining the delay as the average delay per demand. This means that the model calculates the delay as “the total expected system delay over any time period is simply the expected number of units on which delay is being incurred at a random point in time multiplied by the length of the time period (Sherbrooke 1968, p. 133).” At the time of writing this thesis, the METRIC model has been cited approximately 1639 times, which includes these highlights: Deuermeyer and Schwarz (1981), Andersson, Axsäter, and J. Marklund (1998), Andersson and J. Marklund (2000), Cho and Parlar (1991), Berling, Johansson, and J. Marklund (2023), Johansson et al. (2020), Axsäter (2006), Berling and J. Marklund (2013), Berling and J. Marklund (2014), Lee and Billington (1993), De Kok et al. (2018), Axsäter (1990), and Fleischmann, Van Nunen, and Gräve (2003). This model laid the groundwork for multi-echelon inventory systems and remains a viable option for companies due to its simplicity. It was originally developed for the United States Military to allocate stock for spare parts for repairs. However, it can and has been adapted to a standard multi-echelon structure of a central warehouse and multiple retailers, as in the evidence stated above. The next important development to note comes from Deuermeyer and Schwarz (1981). The authors use Little’s Law to derive an estimate of the expected values of delay. The model directly builds on the METRIC model as it provides an approximation for the deterministic lead time. The main issue with this model is variance is left out of their calculations and analysis. This model will be examined further in section 3.3 of this chapter.

Tempelmeier (1983) builds on an interesting concept for delay by representing the stockout duration with its relation to customer wait time. This motive is to show that the lead time is not constant. Tempelmeier is also noteworthy for not using Little’s Law to calculate the expected backorders. The paper also states that lead time is a discrete random variable with a probability distribution. An earlier paper covers a similar approach, but with an $(s-1, S)$ inventory system (Kruse 1980). The final paper in the historical range comes from Svoronos and Zipkin (1988). They calculate the estimated expected values of delay along with their variance. The paper states the importance of variance in the calculations in direct comparison to Deuermeyer and Schwarz (1981). Svoronos and Zipkin (1988) paper provides some key developments that can be seen as extensions to their model. Firstly, their approach can replace the deterministic delays with stochastic delays. This is an important development as it shows there are more accurate approaches to consider. Secondly, the identical retailer assumption can be changed to allow non-identical retailers. Although this will essentially require separate calculations for each retailer, it could reflect more realistic structures. It is important to note that this assumes that the batch sizes, Q , remain consistent. They note that the option is open to examine different batch sizes and by extension partial deliveries, but admit that it is more difficult and requires analysis.

Andersson and J. Marklund (2000) add to delay calculations and the resulting lead time calculations in their normal approximation method, which also builds on the earlier METRIC model approach. They replace the stochastic lead time variables with averages, similar to earlier work. The model can account for partial orders and diverges by using Little’s Law. The explanation is best said by them: “... the parameter Q can be interpreted as an average batch size, which takes into account both the size and frequency of the retailer orders arriving at the warehouse (Andersson and J. Marklund 2000, p. 497).” This means that reserve units will only occur when parts of a retailer’s order is backordered. The importance of this for delay is that it should be calculated upstream at the central warehouse, not at the retailer. This model will be examined further in section 3.3 of this chapter. Axsäter (2003) derives the delay at the central warehouse by incorporating a stochastic inventory level at each retailer. The average number of backorders therefore incorporates both the stochastic delay and stochastic inventory levels in a normal approximation. This approximation includes the mean and variance. This model will be examined further in section 3.3 of this chapter. Kiesmüller et al. (2004) produces an approximation method for stockout delay using the mean and variance with the assumption that the aggregate demand is a mixed-Erlang distribution. A key difference is that they pull on the earlier work of Tempelmeier (1983) mentioned earlier and do not use Little’s Law in the calculation of their approximation. This model will be examined further in section 3.3 of this chapter. Although related to the stockout delay, Chatfield, Hayya, and Cook (2013) provides some interesting findings in their Beer Game inspired model. They use a gamma distribution for lead times and concluded three key points: high lead time variance leads to higher stockout rates and more amplification, safety stock calculations can include lead variance and mean, and stockout propagation occurs upstream. The last key paper in this area comes from Berling and Farvid (2014). They suggest replacing the stochastic demand process with its rate instead. This is a key development as its results suggest an improvement over other existing methods. This model will be examined further in section 3.3 of this chapter.

A few papers have covered further topics related to delay after 2014. Stenius et al. (2016) covers an exact method for the exact waiting-time behavior that is caused by stockouts. The other component related to stockouts is the ship consolidation concept, seen in Johansson et al. (2020). De Kok et al. (2018) conducted a literature review examining the different inventory control models. Most papers they reviewed assume delay is constant, a connection to earlier METRIC type approaches. A few other points such as exact models favor exponential distributions for lead time and the major upside to using constant delay is the assurance of non-overlapping orders (traceability). The final paper, Grob and Bley (2018) propose a negative binomial distribution to approximate the aggregate lead time demand at the central warehouse. They then compute the wait time in accordance with Kiesmüller et al. (2004). The paper claims the negative binomial distribution does not face numerical instability and negative variance, a noted issue in Kiesmüller et al. (2004). Grob and Bley (2018) will be examined in greater detail in section 3.3 of this chapter. It is also important to note a far more recent paper that effectively incorporates stockout delay using real data. Berling, Johansson, and J. Marklund (2023) uses a METRIC type approximation technique. However, the paper incorporates this in an omni/multi-channel distribution system, a more complex system that includes e-commerce orders on top of a multi echelon system.

3.2 Gaps in Literature

Stockout delay in multi-echelon inventory control literature faces some significant short falls. Dedicated papers investigating primarily stockout delay are very limited. Considering that these calculations are known to be difficult and time consuming, more research could still be invested into developing approximations to attempt different methods. Many papers include sections dedicated to delay calculations, but tend to favor METRIC-type approximations. The delay faces gaps in the type of order fulfillment. Complete order policies were used in all but one of the papers examined. Partial orders are important to consider due to their importance in spare part management. This is a key multi-echelon application in a practical sense. More simulations and research should investigate this area. Delay also faces a lack of standardized terminology and definitions. Future authors should consider addressing this issue. Finally, there are a few minor gaps to mention. Continuous (R, Q) inventory policies are the dominant system in the literature. Other types are represented, but in lower volume. This includes periodic review systems. These are less important to the overall gaps identified, but should still be considered.

3.3 Overview of Available models

We have identified five different models to examine in greater detail to conduct a simulation study. We are specifically looking at their stockout delay calculations and how we can apply them into our generic model. Each of these calculations can be applied independently of the overall model calculations as they are a subset of it. The general model we intend to use comes from Berling and J. Marklund (2013) and Berling and J. Marklund (2014). The five models chosen include: the METRIC approach Deuermeyer & Schwarz's and Andersson & Marklund's approaches, the Axsäter normal approximation model, the Kiesmüller & de Kok mixed-erlang distribution model, Berling and Farvid lead time model, and the Grob and Bley wait time approximation model. Note that each model uses slightly different notation. For consistency between each model, we will define a global set of variables to use throughout the rest of the thesis where it is applicable. Though some variables are hard to generalize and will have to stay as they are.

3.3.1 General Inventory Control Theory

In inventory control a divergent distribution system is often modeled as one warehouse that supplies a non-specified number of retailers. This would make it a two-echelon system though some models are defined for an arbitrary number of stages. Though any distribution system could be made arbitrarily large, the scope of the outside world to the model is often limited to a warehouse supplier that is assumed to always deliver on time with a constant lead time thus making it feasible to control the system.

The motivation for attempting to model a complex real-world system is usually to optimize for cost, either using shortages cost or applying a service level target such as fill rate. Fill rate is the demand that can be satisfied immediately from stock on hand and in the case of demand that arrives one unit at a time as in Poisson demand it can be defined as in (1). If the cost is minimized, the customer service is still regarded indirectly through the introduction of a back-order cost incurred by the retailer for missing sales and delivering poor customer service. The optimization procedure starts at the warehouse with deciding its reorder point R_0 which then will have an effect on the stock-out delay that the retailers will experience. Too high a reorder point and unnecessary holding cost will accrue at the warehouse. On the other hand, a low reorder point will result in larger and/or more delays due to stock-outs at the warehouse. Something that the retailers will have to counteract with larger holdings of stock and thereby increasing their costs. There lies the optimization problem and the opportunity to make operations more efficient and profitable.

$$S_2 = S_3 = P(IL > 0) = \sum_{j=1}^{R+Q} P(IL = j) = \sum_{j=1}^{R+Q} \frac{1}{Q} \sum_{k=\max(R+1,j)}^{R+Q} P(D(L) = k - j) \quad (1)$$

General notation

N	Number of retailers
L_0	Constant lead time from external supplier to central warehouse
L_i	Lead time for retailer i
l_i	Constant transportation time from central warehouse to retailer i
λ_i	Demand rate at retailer i
IL	Inventory level (stock on hand – backorders)
IP	Inventory position (stock on hand + units ordered – backorders)
μ_i	Mean demand rate during warehouse lead time for retailer i
0	Subscript representing the warehouse
R	Subscript representing generic retailer
B_0	Backorders at the warehouse
R_i	Reorder point for installation i ($i=0,1,2,\dots,N$)
Q_i	Batch size for installation i ($i=0,1,2,\dots,N$)
<i>Subscript 0 denotes the warehouse; $i \in \{1, \dots, N\}$ denotes retailer i.</i>	

The properties of the model that are needed to calculate the inventory levels and thus influencing the cost

and service levels are for the most part already defined in the setup. This would include the reorder points and the order quantities. However, the waiting time experienced by the retailer is determined inside the system due to the interactions between the warehouse and the retailers. A common way to model the retailer lead time L_i is to replace the delay with its mean as in 2. This is done in the METRIC model, an influential early inventory control model developed for spare parts replenishments.

$$\bar{L}_i = l_i + E[W] \quad (2)$$

It makes use of Little's law which is described in (3). This law holds for any stable system, meaning that the replenishment rate must be less than or equal to the demand rate.

$$L = \lambda \cdot W \quad (3)$$

The approximation in (4) works well for METRIC though it uses an $(s - 1, S)$ reorder policy and the results are not as good when retailers are allowed to choose their own batch size. In that case the waiting times for retailers with larger batch sizes are underestimated. Little's law is used as one step in the derivation of the mean delay in Axsäter 2003.

$$E[W] = \frac{E[BO_0]}{\sum_{i=1}^N \lambda_i} \quad (4)$$

$$P(D_i(L_0) = j), \quad j \in \{0, 1, \dots, j_{\max}\} \quad (5)$$

In many models it is of good use to find the demand at the warehouse during L_0 . This is done through a series of steps. The distribution of the number of orders placed by retailer i during L_0 in (6) is in the next equation (19) remade to represent the unit demand from retailer i .

$$P(N_i = n) = \sum_{k=R_i}^{R_i+Q_i} P(IP = k) \cdot P(k + (n - 1)Q_i \leq D_i(L_0) < k + nQ_i), \quad n > 0 \quad (6)$$

where N_i is the number of orders placed by retailer i during the warehouse lead time L_0 .

$$P(S_1 = j) = P\left(N_1 = \frac{j}{Q_1}\right), \quad j \in \{0, Q_1, 2Q_1, \dots\} \quad (7)$$

Through successive convolutions of the retailers demand distribution in (8) the unit demand during lead time at the warehouse (9) can be calculated.

$$P(S_k = d) = \sum_{j=0}^d P(S_{k-1} = j) \cdot P\left(N_k = \frac{d-j}{Q_k}\right), \quad k = 2, 3, \dots, N \quad (8)$$

$$D_0(L_0) = S_N \quad (9)$$

The inventory level is necessary to calculate holding costs, fill rates and backorders (10). To compute the backorders the definition in (11) is used.

$$P(IL_0(t + L) = j) = \frac{1}{Q_0} \sum_{k=\max(R_0+1, j)}^{R_0+Q_0} P(D_0(L_0) = k - j), \quad j \leq R_0 + Q_0 \quad (10)$$

$$E[BO] = E[(-IL)^+] = \sum_{j=-\infty}^0 (-j) \cdot P(IL = j) = \sum_{j=1}^{\infty} j \cdot P(IL = -j) \quad (11)$$

In the following, $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal probability density function and cumulative distribution function, respectively. The first and second order normal loss functions are defined as.

$$G(x) = \int_x^{\infty} (v - x) \varphi(v) dv = \varphi(x) - x(1 - \Phi(x)) \quad (12)$$

$$H(x) = \int_x^{\infty} G(v) dv = \frac{1}{2} [(x^2 + 1)(1 - \Phi(x)) - x\varphi(x)] \quad (13)$$

3.3.2 The METRIC Approach

The METRIC model (Sherbrooke 1968) does not provide an approximation for stockout delay. It makes an assumption the lead time is the same for all retailers and that the lead time is deterministic. Although it is assumed to be deterministic, there are different ways to calculate this including our topic, stockout delay. This is a component of the derived value of expected lead time. What makes the METRIC approach distinct is the use of expected backorders to calculate the expected lead time. We will examine two improvements to this particular approach and how it could affect the results of simulation. We will examine Deuermeyer and Schwarz (1981) first and then Andersson and J. Marklund (2000).

Deuermeyer & Schwarz Approach

The Deuermeyer and Schwarz (1981) model focuses on estimating the expected service level in a multi-echelon system. This model is an approximation of an earlier exact model. It is a (R, Q) inventory system with one central warehouse and multiple identical retailers. The following are assumptions of a fixed/constant value: lot sizes, reorder point, warehouse lead time, and nominal retailer lead times (if no delay). The demand for each retailer is Poisson and identical. With these general details, we now will look at the stochastic delay calculations, which is an adaptation of the METRIC approach by Sherbrooke (1968).

Deuermeyer and Schwarz (1981) state that their adapted approach (see Hadley and Whitin 1963) is a combination of the approximation of the warehouse demand and the deterministic approximation to the retailer effective lead times. This approach uses a normal distribution to approximate the warehouse demand during the lead time. First, we need to define the relevant unique notation used by Deuermeyer and Schwarz. Note that these definitions are how the authors wrote them with minor notational changes.

D_i	expected delay from the warehouse to retailer j due to inadequate warehouse stock, $j = 1, \dots, N$
$\mathcal{L}_i$	expected effective lead time to retailer j ($\mathcal{L}_j = L_j + D_j$)
T_i	expected delay at location j , $j = W, 1, \dots, N$
F_i	expected fill rate at location j , $j = W, 1, \dots, N$
μ_0	expected warehouse lead time demand
σ_0^2	variance of the warehouse lead time demand

They define stockout delay as: “the warehouse may run out of stock and be forced to backorder retailer orders while waiting for its own order(s) to be received. This causes a delay before the order can be shipped back to the retailer. The effect is to increase retailer lead time by a random amount (Deuermeyer and Schwarz 1981, p. 175).” To represent this in an equation for the expected lead time that includes this delay, they state:

$$\mathcal{L}_i = L_i + D \quad (14)$$

where

$$D = E[d] = T_0(Q_0, R_0; \mu_0, \sigma_0) = \lambda_0^{-1} B_0(Q_0, R_0; \mu_0, \sigma_0^2) \quad (15)$$

In the case of this model, D can be viewed as the average length of time that a random order placed at the central warehouse will wait. In more plain terms, it is the expected delay at the central warehouse. This makes the expected lead time deterministic and fixed in this case. To determine D, the authors compute T_0 . The basis of computing T_0 comes from dividing the expected backorders by λ_0 . The backorder calculation is represented by:

$$B_0(Q_0, R_0; \mu_0, \sigma_0^2) = \begin{cases} \frac{1}{Q_0} [\hat{\beta}(R_0) - \hat{\beta}(R_0 + Q_0)], & R_0 \geq 0, \\ \frac{1}{Q_0} [\hat{\beta}(0) - \hat{\beta}(R_0 + Q_0)] + \frac{R_0}{Q_0} \left[\frac{R_0}{2} - \hat{\alpha}(0) \right], & -Q_0 \leq R_0 < 0, \end{cases} \quad (16)$$

where

$$\hat{\alpha}(v) = \sigma_0 \phi \left(\frac{v - \mu_0}{\sigma_0} \right) - (v - \mu_0) \Phi \left(\frac{v - \mu_0}{\sigma_0} \right) \quad (17)$$

$$\begin{aligned} \hat{\beta}(v) = & \frac{1}{2} [\sigma_0^2 + (v - \mu_0)^2] \Phi \left(\frac{v - \mu_0}{\sigma_0} \right) - \frac{\sigma_0}{2} (v - \mu_0) \\ & \times \phi \left(\frac{v - \mu_0}{\sigma_0} \right) \end{aligned} \quad (18)$$

The expected backorders from the central warehouse are assumed to be derived from Little's Law. The specific functions $\hat{\alpha}(v)$ and $\hat{\beta}(v)$ uses two different types of normal distributions ϕ and Φ to calculate them. Refer to Axsäter (2006) or Hadley and Whitin (1963) for detailed explanations of these functions.

The mean and standard deviation for the central warehouse seen in these equations are derived by the authors for this approach. For the complete proofs, refer to Deuermeier and Schwarz (1981). As previously mentioned, these are approximates from a normal distribution. The mean is the expected value of the warehouse lead time while the standard deviation is the variance of the warehouse lead time. Equation (19) and (20) represent this.

$$\mu_0 = E[N(L_0)] = \frac{\lambda_0 L_0}{Q_i} + \frac{1 - Q_i}{2Q_i} \quad (19)$$

and

$$\sigma_0^2 = \text{var}[N(L_0)] = \frac{\lambda_0 L_0}{Q_i^2} \quad (20)$$

In the case of this model, D can be viewed as the average length of time that a random order placed at the central warehouse will wait. This makes the expected lead time deterministic and fixed in this case. This model specifically adheres to complete orders with no partial fulfilment, while it is unknown if the model will work for partial fulfilment. The model performs poorly in cases with extremely low safety stocks. There is also an interesting observation that the authors noted. When the fill rate decreases, the warehouse delay's variance increases. The distribution of this delay will differ from the mean whenever the warehouse has higher fill rate levels.

Andersson & Marklund Approach

Andersson and J. Marklund (2000)'s model works within a decentralized (R, Q) inventory system and looks

to optimize through a cost evaluation technique. The authors' approach follows the METRIC approach (Sherbrooke 1968), but differs from Deuermeyer and Schwarz (1981).

The general model follows a central warehouse with non-identical retailers. The following assumptions are made: transportation time is constant, outside supplier transport time is constant, complete deliveries, and all order quantities are fixed. It is important to note that the reorder point is fixed to units of Q . The model also uses non-identical retailers, a key difference from Deuermeyer and Schwarz (1981). The stockout delay defined in Andersson and J. Marklund (2000) simply states it is the additional time from stockouts at the central warehouse. This paper uses different notation, so we will define the relevant unique notation below. Note that this is how the authors denote the variables.

q_i the batch size used by retailer i , expressed in units of Q
 $\bar{L}_i$ expected lead time for an order to arrive at retailer i ($\bar{L}_i = E[L_i]$)

This model provides their calculations for the expected lead time, which incorporates both mean delivery delay to stockouts at the warehouse and the mean delivery delay due to reserved units waiting to be shipped. The first term to address is the expected backorders at the central warehouse, B_0 . This is derived the same way as Deuermeyer and Schwarz (1981). The expected backorders at the central warehouse is the key component for the METRIC approach as well as the expected lead time in this approach. The important aspect that comes from this is the mean delivery delay due to stockouts at the warehouse, seen below.

$$\bar{\Delta}_1 = E[B_0]Q/\mu_0 \quad (21)$$

The μ_0 variable that is listed above is defined by the authors as the demand intensity at the warehouse expressed in number of units. Compare this to Deuermeyer and Schwarz (1981) version of in equation (19). Andersson and J. Marklund (2000) use a summation over all the means from each retailer while, Deuermeyer and Schwarz (1981) incorporate both the retailer demand rate, batch size, and fixed lead time. Andersson and J. Marklund (2000)'s approach is much simpler, as seen below.

$$\mu_0 = \sum_{i=1}^N \mu_i \quad (22)$$

The mean delivery delay due to reserved units waiting to be shipped is the next term needed to define.

$$\bar{\Delta}_2 = E[B_0^r]Q/\mu_0 \quad (23)$$

The expected backorders from reserved units waiting is stated to not have an exact expression for it. The authors use an approximation that uses the size and frequency of retailer orders. This is represented as equation (24).

$$E[B_0^r] = E\left[\left\lceil \frac{B_0}{\bar{Q}} \right\rceil \bar{Q}\right] - E[B_0]. \quad (24)$$

where

$$\bar{Q} = \sum_{i=1}^n \frac{\mu_i}{Q_i \lambda_0} \cdot Q_i = \sum_{i=1}^n \frac{\mu_i}{\lambda_0}, \quad (25)$$

$$\lambda_0 = \sum_{i=1}^n \frac{\mu_i}{Q_i} \quad (26)$$

λ_0 is the demand intensity at the warehouse in the number of retailer orders. It is directly used by the authors to calculate the size and frequency of retailer orders, $\bar{Q}$. Now, the authors continue by showing the expected lead time calculations.

$$\bar{L}_i = \bar{\Delta}_1 + \bar{\Delta}_1 \left(\frac{\mu_i(Q_i - Q) - (\eta_1/N)}{\eta_1} \right) + \bar{\Delta}_2 \frac{\mu_i(Q_i - Q)}{\eta_1} + l_i. \quad (27)$$

where

$$\eta_1 = \sum_{i=1}^N \mu_i(Q_i - Q), \quad (28)$$

The term $\mu_i(Q_i - Q)$ is defined as “use the retailer’s relative deviation from the average retailer size (Andersson and J. Marklund 2000, p. 497).” The important distinction between this expected lead time and Deuermeyer and Schwarz (1981) version is the possibility of non-identical retailers. The Andersson and J. Marklund (2000) model allows for both identical and non-identical retailers, a distinct advantage. In this case, each retailer must be calculated separately. The authors state this expected lead time turns exact when there are identical retailers. The final part of interest is that this model allows for partial order fulfillment, but does not provide the necessary adjustments for it.

3.3.3 Axsäter Normal Approximation Model

In Axsäter (2003), the holding and shortage cost is minimized within a two-echelon distribution inventory system with a central warehouse and a number of retailers. The retailers all face heterogeneous compound Poisson demand and have an independent batch ordering quantity Q_i and reorder point R_i .

What is new in this paper is that it presents a novel approximative method based on normal approximations of both the retailer demand and the demand at the warehouse from the retailers. A benefit with this approach is that it is efficient even for larger systems. Below is the notation used in the paper relevant for the presentation of the estimation of stock delay experienced by the retailers.

- q = largest common factor of $Q_0, Q_1, \dots, Q_N$,
- I_i = stochastic inventory level at installation i ($i = 0, 1, \dots, N$),
- δ = stochastic delay at the warehouse for retailer orders,
- μ_δ = mean of δ ,
- $\mu_i(L_0)$ = average demand per time unit at retailer i ,
- $\text{Var}_i(L_0)$ = variance of demand at retailer i during the warehouse lead time,
- Var_δ = variance of δ .

$$\begin{aligned} \mu^W(L_0) &= \sum_{i=1}^N \mu_i^W(L_0), \\ \text{Var}^W(L_0) &= \sum_{i=1}^N \text{Var}_i^W(L_0). \end{aligned}$$

In the lead up to the estimation of the lead time demand a few assumptions are made. A unit is considered to arrive at time t_d and at time t_a it has arrived at the warehouse. If the unit is available upon it being demanded then it follows that $t_a \leq t_d$. Similarly, if there is a stockout at the warehouse then $t_d < t_a$.

In the next step, the difference between the time notations is defined as $\Delta = t_a - t_d$. The stockout is ultimately approximated as a normal distribution with mean μ_δ and variance Var_δ . It is assumed that the inventory at the warehouse is uniformly distributed on the interval $[R_0 + q, R_0 + 2q, \dots, R_0 + Q_0]$ at an arbitrary time. For modeling purposes it is approximated as a continuous uniform distribution on $(R_0 + q, R_0 + Q_0)$.

With that setup the expected stockout level $E[I_0^-]$ is approximated as:

$$E[I_0^-] = \frac{1}{Q_0 - q} \int_{R_0+q}^{R_0+Q_0} \int_0^\infty (x - R_0 - y) \frac{1}{\sigma^W(L_0)} \phi\left(\frac{x - \mu^W(L_0)}{\sigma^W(L_0)}\right) dx dy \quad (29)$$

The demand during the lead time is assumed to be normally distributed with mean $\mu^W(L_0)$ and standard deviation $\sigma^W(L_0)$. The outer integral integrates over y , where $y + R_0$ is the inventory position at the warehouse when an order is placed, ranging from $R_0 + q$ to $R_0 + Q_0$. The inner integral is integrated with respect to x , where x represents the lead time demand. Its range captures instances where the lead time demand is greater than the inventory position at the time of ordering.

This expression is then reformulated as

$$E[I_0^-] = \frac{(\sigma^W(L_0))^2}{Q_0 - q} \left[H\left(\frac{R_0 + q - \mu^W(L_0)}{\sigma^W(L_0)}\right) - H\left(\frac{R_0 + Q_0 - \mu^W(L_0)}{\sigma^W(L_0)}\right) \right] \quad (30)$$

The expressions $G(x)$ and $H(x)$ are defined as in 12 and 13.

With the use of Little's law the mean stochastic delay can be determined as

$$\mu_\delta = \frac{E[I_0^-]}{\sum_{i=1}^N \mu_i} \quad (31)$$

To find Var_δ the probability of no stockout delay is determined

$$P(\delta = 0) = \frac{1}{Q_0 - q} \int_{R_0+q}^{R_0+Q_0} \Phi\left(\frac{R_0 + y - \mu^W(L_0)}{\sigma^W(L_0)}\right) dy \quad (32)$$

For $Q_0 = q$ this simplifies to

$$P(\delta = 0) = \Phi\left(\frac{R_0 + Q_0 - \mu^W(L_0)}{\sigma^W(L_0)}\right) \quad (33)$$

The next assumption is that Δ is normally distributed. This means

$$P(\Delta \leq 0) = \Phi\left(-\frac{\mu_\Delta}{\sigma_\Delta}\right) = P(\delta = 0) \quad (34)$$

This can be rearranged as:

$$\alpha = -\frac{\mu_\Delta}{\sigma_\Delta} = \Phi^{-1}(P(\delta = 0)) \quad (35)$$

With the newfound value α the standard deviation for Δ is determined

$$E((\Delta)^+) + \mu_\Delta = \int_0^\infty \left[1 - \Phi\left(\frac{x - \mu_\Delta}{\sigma_\Delta}\right) \right] dx = \sigma_\Delta G(\alpha) \implies \sigma_\Delta = \frac{\mu_\delta}{G(\alpha)} \quad (36)$$

The mean of Δ is then

$$\mu_{\Delta} = -\alpha\sigma_{\Delta} \quad (37)$$

With equation 31, 33, 36 and 37 it can then be shown that Var can be obtained.

$$\text{Var}_{\delta} = E(\delta^2) - (\mu_{\delta})^2 = \sigma_{\Delta}^2(1 - P(\delta = 0)) + \mu_{\Delta}\mu_{\delta} - \mu_{\delta}^2 \quad (38)$$

3.3.4 Kiesmüller & de Kok Mixed-Erlang Distribution Model

In Kiesmüller et al. (2004), a divergent inventory system with an arbitrary number of echelons is modeled. Within this paper the following notation is used.

M	Number of stockpoints
N_e	Number of echelons
$\mathcal{M}$	Set of all stockpoints $\{1, 2, \dots, M\}$
$\mathcal{S}_k$	The set of all immediate successors to installation k
$\mathcal{P}_k$	The set of all immediate predecessors to installation k
$\mathcal{E}_i$	Echelon i , where $\mathcal{E}_1$ is the most downstream customer facing echelon
$N_k(L)$	Number of customers arriving at k during an arbitrary replenishment cycle at $k \in \mathcal{M}$
$D_j(t)$	Stochastic demand variable at installation during time t
O_k	Replenishment order size of $k \in \mathcal{M}$
L_d^k	Transportation time to $k \in \mathcal{M}$
L	Lead time at $k \in \mathcal{M}$
W_k	Waiting time of $k \in \bigcup_{i=1}^{N_e} \mathcal{E}_i$
T_k	Inter-replenishment time for replenishment $k \in \mathcal{M}$
c_Y	Coefficient of variation of Y , $c_Y = \frac{\sigma(Y)}{E[Y]}$

The waiting time W_k is defined as the length of the interval between the placement of the order and the moment when the entire order O_k can be shipped. A mixed-Erlang distribution $\mathbf{E}_{k_1, k_2}(\mu_1, \mu_2), (p_1, p_2)$ is then applied to approximate the waiting time. The parameters can be obtained from the first two moments of W_k as in (39) and (40).

$$E[W_k] = \int_0^{\infty} P\{W_k > x\} dx \quad \forall k \in \bigcup_{i=1}^{N-1} \mathcal{E}_i \quad (39)$$

$$E[W_k^2] = 2 \int_0^{\infty} x P\{W_k > x\} dx \quad \forall k \in \bigcup_{i=1}^{N-1} \mathcal{E}_i \quad (40)$$

To find an expression for $P\{W_k > x\}$ three propositions are introduced by DeKok as seen in (41)

Proposition 1 *The distribution function of W_k ($k \in \bigcup_{i=1}^{N-1} \mathcal{E}_i$) conditioned on $L_j = l$ can be approximated for $0 \leq x \leq l$ and $j \in \mathcal{P}_k$.*

$$P\{W_k \leq x | L_j = l\} \approx 1 - \frac{1}{Q_j} (E[(D_j(l-x) + O_k - s_j)^+] - E[(D_j(l-x) + O_k - (s_j + Q_j))^+]) \quad (41)$$

Proposition 2 *The first moment of the waiting time is approximated for all $k \in \mathcal{M}$ by*

$$E[W_k] \approx E[L_j] \frac{E[(D_j(\hat{L}_j) + O_k - s_j)^+] - E[(D_j(\hat{L}_j) + O_k - (s_j + Q_j))^+]}{Q_j} \quad (42)$$

where $\hat{L}_j$ ($j \in \mathcal{P}_k$) is a random variable with distribution function

$$F_{\hat{L}_j}(y) = \frac{1}{E[L_j]} \int_0^y (1 - F_{L_j}(z)) dz \quad (43)$$

Note that $\hat{L}_j$ is the residual lifetime of L_j .

Proposition 3 The second moment of the waiting time is approximated for $k \in \mathcal{M}$ by

$$E[W_k^2] \approx E[L_j^2] \frac{E[(D_j(\tilde{L}_j) + O_k - s_j)^+] - E[(D_j(\tilde{L}_j) + O_k - (s_j + Q_j))^+]}{Q_j} \quad (44)$$

where $\tilde{L}_j$ ($j \in \mathcal{P}_k$) is a random variable with distribution function

$$F_{\tilde{L}_j}(y) = \frac{2}{E[L_j^2]} \int_0^y \int_x^\infty (z - x) dF_{L_j}(z) dx \quad (45)$$

To calculate the first two moments of W_k a series of evaluations are performed.

To start with, the first three moments of A_j , the inter-arrival time for orders at the warehouse. k here refers to the stockpoints facing the customers, i.e the retailers while A_k is the mean time between arrivals.

$$E[A_j] = \frac{1}{\sum_{k \in \mathcal{S}_j} \frac{1}{E[T_k]}} \quad (46)$$

$$E[T_k] = E[R_k] E[A_k] = Q_k \frac{1}{\lambda_k} \quad (47)$$

The replenishment rate, R_k , is the number of customers arriving during the replenishment cycle. For a retailer the mean of R_k is Q_k .

$$E[A_j^2] = 2E[A_j] \left(\prod_{k \in \mathcal{S}_j} \frac{1}{E[T_k]} \right) \int_0^\infty \left(\prod_{k \in \mathcal{S}_j} \int_x^\infty (1 - F_{T_k}(y)) dy \right) dx \quad (48)$$

Each inter-arrival time for a customer placing an order at installation is exponentially distributed with intensity λ_k . The time for Q_k arrivals is then the sum of Q_k X_i and that is known as an *Erlang*(Q_k, λ_k)-distribution.

$$X_i \sim \text{Exp}(\lambda_i), \quad T_k = \sum_{i=1}^{Q_k} X_i \sim \text{Erlang}(Q_k, \lambda_k) \quad (49)$$

$$1 - F_{T_k}(y) = \sum_{n=0}^{Q_k-1} \frac{(\lambda_k y)^n e^{-\lambda_k y}}{n!} \quad (50)$$

Next, the third moment of A_j is fitted according to a mixed-erlang distribution derived from the first two moments of A_j . The mixed-erlang distribution $E_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ has six parameters that are decided in two different ways depending on the value of c_X^2 .

Case 1: $c_X^2 < 1$ — mixture of two Erlangs with the same rate:

$$k_1 = \left\lfloor \frac{1}{c_X^2} \right\rfloor, \quad k_2 = k_1 + 1 \quad (51)$$

$$p_1 = \frac{1}{1 + c_X^2} \left(k_2 c_X^2 - \sqrt{k_2(1 + c_X^2) - k_2^2 c_X^2} \right), \quad p_2 = 1 - p_1 \quad (52)$$

$$\mu_1 = \mu_2 = \frac{k_2 - p_1}{E[X]} \quad (53)$$

Case 2: $c_X^2 \geq 1$ — mixture of two exponentials:

$$k_1 = k_2 = 1 \quad (54)$$

$$\mu_1 = \frac{2}{E[X]} \left(1 + \sqrt{\frac{c_X^2 - \frac{1}{2}}{c_X^2 + 1}} \right), \quad \mu_2 = \frac{4}{E[X]} - \mu_1 \quad (55)$$

$$p_1 = \frac{\mu_1(\mu_2 E[X] - 1)}{\mu_2 - \mu_1}, \quad p_2 = 1 - p_1 \quad (56)$$

The mixed-Erlang $E_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ has density:

$$f_X(x) = \sum_{i=1}^2 p_i \frac{\mu_i^{k_i} x^{k_i-1}}{(k_i-1)!} e^{-\mu_i x} \quad (57)$$

By integrating $x^3 \cdot f_X(x)$ the third moment in (58) is computed.

$$E[A_j^3] = p_1 \cdot \frac{k_1(k_1+1)(k_1+2)}{\mu_1^3} + p_2 \cdot \frac{k_2(k_2+1)(k_2+2)}{\mu_2^3} \quad (58)$$

The next preparatory step is to find the first and second moment of D_j , the order size of the aggregate demand process at the warehouse.

$$E[D_j] = E[A_j] \sum_{k \in \mathcal{S}_j} \frac{E[O_k]}{E[T_k]} \quad (59)$$

$$E[D_j^2] = [A_k] \sum_{k \in \mathcal{S}_j} \frac{E[O_k^2]}{E[T_k]} \quad (60)$$

Step 1 The first two moments of $\hat{L}_j$ and $\tilde{L}_j$ is computed according to ((61)) and ((62)).

$$E[\hat{L}_j] = \frac{E[L_j^2]}{2E[L_j]}, \quad E[\hat{L}_j^2] = \frac{E[L_j^3]}{3E[L_j]}, \quad j \in \mathcal{P}_k \quad (61)$$

$$E[\tilde{L}_j] = \frac{E[L_j^3]}{3E[L_j^2]}, \quad E[\tilde{L}_j^2] = \frac{E[L_j^4]}{6E[L_j^2]}, \quad j \in \mathcal{P}_k \quad (62)$$

Step 2 Then the first two moments of $N(\hat{L}_j)$ is computed as in ((63)) and ((64)).

$$E[N(\hat{L}_j)] = \frac{E[L_j]}{E[A_j]} + \frac{E[A_j^2]}{2E[A_j]^2} - 1, \quad j \in \mathcal{P}_k \quad (63)$$

$$E[N(\hat{L}_j)^2] = \frac{E[\hat{L}_j^2]}{E[A_j]^2} + E[\hat{L}_j] \left(\frac{2E[A_j^2]}{3E[A_j]^3} - \frac{3}{E[A_j]} \right) + \frac{3E[A_j^2]^2}{2E[A_j]^4} - \frac{2E[A_j^3]}{3E[A_j]^3} - \frac{3E[A_j^2]}{2E[A_j]^2} + 1, j \in \mathcal{P}_k \quad (64)$$

The first two moments of $N(\tilde{L}_j)$ is then obtained in the same manner.

Step 3 The first two moments of $D(\hat{L}_j)$ is computed as in (65) and (66).

$$E[D_j(\hat{L}_j)] = E[N(\hat{L}_j)]E[D_j], \quad j \in \mathcal{P}_k \quad (65)$$

$$E[D_j^2(\hat{L}_j)] = E[N(\hat{L}_j)\sigma^2(D_j) + E[N(\hat{L}_j)^2]E[D_j]^2], j \in \mathcal{P}_k \quad (66)$$

In the same way, the first two moments of $D(\hat{L}_j)$ can be computed.

Step 4 The parameters of a mixed-Erlang distribution with the same first two moments as $D_j(\tilde{L}_j) + O_k$ and $D_j(\hat{L}_j) + O_k$ is computed as in equations (51), (52), (53), (54), (55) and (56).

Step 5 The parameters obtained from step 4 are then used to find $E[D_j(\hat{L}_j + O_k - R_j)^+]$, $E[D(\hat{L}_j) + O_k - (R_j + Q_j)^+]$, $E[D_j(\tilde{L}_j + O_k - R_j)^+]$ and $E[D(\tilde{L}_j) + O_k - (R_j + Q_j)^+]$. To find them equation (67) is used where X is distributed according to a mixed-Erlang distribution.

$$\begin{aligned} E[(X - z)^+] &= \int_z^\infty (x - z) dF_X(x) \\ &= \sum_{i=1}^2 \int_z^\infty (x - z) \frac{\mu_i^{k_i} x^{k_i-1}}{(k_i - 1)!} e^{-\mu_i x} dx \\ &= \sum_{i=1}^2 \left(\frac{k_i}{\mu_i} \int_z^\infty \frac{\mu_i^{k_i} x^{k_i+1}}{k_i!} e^{-\mu_i x} dx - z \int_z^\infty \frac{\mu_i^{k_i} x^{k_i-1}}{(k_i - 1)!} e^{-\mu_i x} dx \right) \\ &= \sum_{i=1}^2 \left(\frac{k_i}{\mu_i} \cdot P(\text{Erlang}(k_i + 1, \mu_i) > z) - z \cdot P(\text{Erlang}(k_i, \mu_i) > z) \right) \\ &= \sum_{i=1}^2 \left(\frac{k_i}{\mu_i} \sum_{n=0}^{k_i} \frac{(\mu_i z)^n e^{-\mu_i z}}{n!} - z \sum_{n=0}^{k_i-1} \frac{(\mu_i z)^n e^{-\mu_i z}}{n!} \right) \end{aligned} \quad (67)$$

Step 6 $E[W_k]$ and $E[W_k^2]$ is then computed according to (42) and (44). The variance is then found through $Var(W_k) = E[W_k^2] - E[W_k]^2$.

3.3.5 Berling & Farvid Lead Time Model

Berling and Farvid (2014) model is an approach that specifically investigates the lead time approximations and attempts to derive a more accurate approximation while being computationally feasible. The overall approach to the delay in this paper focuses on replacing the stochastic lead time demand with its demand rate instead. The model uses the following assumptions: non-identical retailers, the retailer demand is independent over time, retailer demand is non-negative and stationary, central warehouse demand is only from replenishment based orders by retailers, constant lead time from the outside supplier, transportation time is constant, (R, Q) inventory system, first-come-first-served principal, and complete order fulfillment. The model's notation follows the same style as Andersson and J. Marklund (2000) and Axsäter (2003), but with a few unique variables that are defined below.

τ_p	the time the order of the batch is considered is placed by the retailer
τ_s	the time the batch is considered shipped
Δ_i	delay at the central warehouse for a considered batch; $\Delta_i = \tau_s - \tau_p$
$\xi_{0,i}$	demand at the central warehouse in the time period $[\tau_s - L_0, \tau_p]$ in excess of the considered batch of Q_i units placed by retailer i at time τ_p
$\mu_{0,i}$	expected demand rate at the central warehouse in period $[\tau_p, \tau_s - L_0]$ expressed in units per time unit of length L_0
$\delta_j(n)$	probability that a retailer j orders at most n batches in $[\tau_s - L_0, \tau_p]$
$g_j(u)$	probability that demand from retailer j at the central warehouse in $[\tau_s - L_0, \tau_p]$ is u
$\mu_{\xi_{0,i}}$	expected value of $\xi_{0,i}$
$\sigma_{\xi_{0,i}}$	standard deviation of $\xi_{0,i}$

The authors define the distribution for a retailer i 's lead time as equation ().

$$L_i = \Delta_i + l_i \quad (68)$$

The delay here is explicitly defined as the delay due to the shortages at the central warehouse, as stated in the variable definitions. The main goal of the authors was to develop a heuristic that successfully changes the stochastic demand process with its rate instead. This rate is assumed to have demand occur in a continuous way, but with an uncertain rate. To begin this process, the authors lay out three different cases that occur under the assumption that $\xi_{0,i}$ is continuous and when $\mu_{0,i}$ is constant. Figure 2 illustrates the general process and how the inventory position tracks over time.

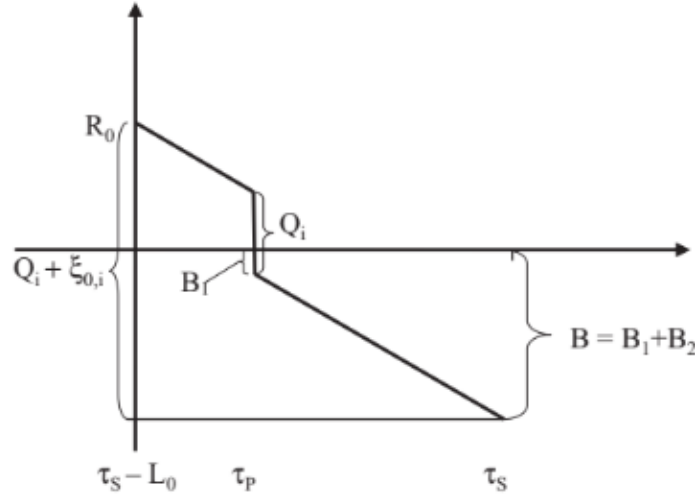


Figure 2: The outstanding order placed, $\tau_s - L_0$, and its evolution over the prevailing assumption (Berling and Farvid 2014, p. 179).

This figure (Berling and Farvid 2014, p. 179) is a general example of how an inventory position evolves over time at the central warehouse. When an order is placed by a retailer, Q_i , and it drops below the line seen above, a backorder is produced along with the first component of the delay, τ_p . The time it takes to ship is of course τ_s and the difference between the two is $\Delta_i = \tau_s - \tau_p$. Now, each of the three cases are shown below.

Case 1: $Q_i \leq R_0$

$$E[\Delta_i]_{R_0 \geq Q_i} = \frac{L_0}{Q_0} \sum_{x=R_0+1}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right) f_i(\xi_{0,i}) \right) \quad (69)$$

$$E[\Delta_i^2]_{R_0 \geq Q_i} = \frac{L_0^2}{Q_0} \sum_{x=R_0+1}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right)^2 f_i(\xi_{0,i}) \right), \quad (70)$$

Case 2: $0 \leq R_0 < Q_i$

$$E[\Delta_i]_{0 \leq R_0 < Q_i} = \frac{L_0}{Q_0} \left(Q_i - R_0 - 2 + \sum_{x=Q_i}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right) f_i(\xi_{0,i}) \right) \right) \quad (71)$$

$$E[\Delta_i^2]_{0 \leq R_0 < Q_i} = \frac{L_0^2}{Q_0} \left(Q_i - R_0 - 2 + \sum_{x=Q_i}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right)^2 f_i(\xi_{0,i}) \right) \right), \quad (72)$$

Case 3: $R_0 < 0$

$$E[\Delta_i]_{R_0 < 0} = \frac{L_0}{Q_0} \left(\sum_{y=R_0+1}^{-1} \left(1 + \frac{y-R_0}{\mu_{0,i}} \right) + Q_i + \sum_{x=Q_i}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right) f_i(\xi_{0,i}) \right) \right) \quad (73)$$

$$\begin{aligned} E[\Delta_i^2]_{R_0 < 0} = & \frac{L_0^2}{Q_0} \left(\sum_{y=R_0+1}^{-1} \left(1 + \frac{y-R_0}{\mu_{0,i}} \right)^2 + Q_i \right. \\ & \left. + \sum_{x=Q_i}^{R_0+Q_0} \left(\sum_{x=Q_i+1}^{\infty} \left(1 - \frac{x-Q_i}{\xi_{0,i}} \right)^2 f_i(\xi_{0,i}) \right) \right). \end{aligned} \quad (74)$$

The final part of the approach uses the methods of Andersson, Axsäter, and J. Marklund (1998). It involves the methods to obtaining the distribution function and the expected μ_0, i . The distribution function, $f_i(\xi_{0,i})$, is computed by compiling the probability distribution function of the number of batches ordered by every retailer. This is a process seen in Andersson, Axsäter, and J. Marklund (1998), but is distinct enough to be included here. First, Berling and Farvid (2014) show that the probability of a retailer, j , is ordering from the central warehouse at the most n batches over a period of time $[\tau_s - L_0, \tau_p]$. Equation (75) shows this while Equation (76) shows for a continuous demand.

$$\delta_j(n) = \sum_{z=1}^{Q_j} \frac{1}{Q_j} D_j(nQ_j + z, L_0) \quad (75)$$

$$\delta_j(n) = \int_0^{Q_j} \frac{1}{Q_j} D_j(nQ_j + z, L_0) dz \quad (76)$$

Next, the authors cover the probability that a retailer, j , orders exactly u units in equation (77). This is followed by equation (78), which is an approximation of retailer i 's contribution to $\xi_{0,i}$.

$$g_j(u) = \begin{cases} \delta_j(0) & \text{if } u = 0 \\ \delta_j(n) - \delta_j(n-1) & \text{if } u = n \cdot Q_j \text{ for } n > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (77)$$

$$\delta_i(n) = D_i((n+1)Q_i, L_0) \quad (78)$$

The authors explain that from this point, determining $\xi_{0,i}$ is computationally difficult and complex due to the pure number of possible combinations. The work around to this is applying a log-normal distribution function. Berling and Farvid (2014) explain this proof in their Appendix A. Refer to this for the proof. From this, the authors were able to find the following equations (79), (80), (81), (82) that estimate both the expected value and standard deviation of the delay.

$$\mu_{\xi_{0,i}} = \mu_{\xi_{0,i},1} + \mu_{\xi_{0,i},2} + \dots + \mu_{\xi_{0,i},N} \quad (79)$$

where

$$\mu_{\xi_{0,i},j} = \mu_j L_0 \quad \text{for } j \neq i \quad (80)$$

(Or an alternative/improvement to this equation is to use Equation (80))

$$\mu_{\xi_{0,i},i} = \max\{\mu_i L_0 - Q_i, 0\} \quad (81)$$

and

$$\mu_{\xi_{0,i},i} = \sum_{k=0}^{\infty} k Q_i g_i(k Q_i) \quad (82)$$

The variance of the demand at the central warehouse is equations (83) and (84).

$$\sigma_{\xi_{0,i}}^2 = \sigma_{\xi_{0,i},1}^2 + \sigma_{\xi_{0,i},2}^2 + \dots + \sigma_{\xi_{0,i},N}^2 \quad (83)$$

where

$$\sigma_{\xi_{0,i},j}^2 = \sum_{k=0}^{\infty} (\mu_{\xi_{0,i},j} - k Q_j)^2 g_j(k Q_j) \quad (84)$$

The authors pose a final step is to determine $g_j(k Q_j)$ between using Equation (75) and Equation (76) or Equation (78) based if the steady state or adjusted variance are needed.

For the expected, $\mu_{0,i}$ the authors compute this with Equation (85).

$$\mu_{0,i} = L_0 \left(\frac{\mu_{0,i}^i}{\tau} + \sum_{j \neq i} \mu_j \right) \quad (85)$$

There are few stipulations to calculating this. If the expected demand for a particular retailer i for the period $[\tau_p, \tau_s - L_0]$ is known along with the time length of the period $\tau = (\tau_s - L_0) - \tau_p$, the section of the equation related to the retailer contribution is straightforward. It is important to note that this is very similar to the previous equations[add EQ numbers later] for calculating μ_i . This equation can be viewed as an conversion to a rate as previously discussed. That is where the additional parts of the equation come from. To calculate

τ , there are a few different dependent variables. This is both the central warehouse inventory position and Q_i . The authors present the following estimate:

$$\tau \approx \frac{(-R_0/2)}{\sum_{j=1}^N \mu_j} \quad (86)$$

τ is the backbone of the calculations for the distribution function mentioned before.

3.3.6 Grob & Bley Wait Time Approximation Model

Grob and Bley (2018) are using a wait time approximation model, which calculates the expected lead time using the random variable of wait time. The general model is an (R, Q) order policy and attempts to optimize the multi-echelon system. The authors approach is to compare their stockout delay approximation compared to three other models, a similar approach to what we are doing here. The authors claim that their approach is an improvement on the Kiesmüller et al. (2004) approach. Grob and Bley (2018) use a negative binomial to approximate the lead time demand value.

The model has these following assumptions: complete deliveries, the wait time/lead of the warehouse i are independent random variables, unlimited stock at the supplier, and customer demand is a compound Poisson process. The authors define the stockout delay as “orders of local warehouses have to wait for an additional time until sufficient stock is available again (Grob and Bley 2018, p. 3).” The unique variables used are the following:

μ_i	expected customer demand at warehouse i during one time unit
σ_i	standard deviation of customer demand at warehouse i during one time unit
$W_i(R_0)$	wait time at the local warehouse i , which depends on the central reorder point
$L_i^{eff}(R_0)$	effective lead time of local warehouse i , defined as $L_i + W_i(R_0)$
$D_i(t)$	demand at the local warehouse i during a time period length t

This approach is an adaptation of the previously discussed Kiesmüller et al. (2004) approach. Grob and Bley (2018) calculate the expected wait time using the random variables $\hat{L}_0$ and $\tilde{L}_0$, which are defined in subsection 3.3.4. The following equations represent both the first then the second moment of wait time.

$$E[W_i] \approx \frac{E[L_0]}{Q_0} \left(E \left[\left(D_0(\hat{L}_0) + Q_i - R_0 \right)^+ \right] - E \left[\left(D_0(\hat{L}_0) + Q_0 - (R_0 + Q_0) \right)^+ \right] \right) \quad (87)$$

and

$$E[W_i^2] \approx \frac{E[L_0^2]}{Q_0} \left(E \left[\left(D_0(\tilde{L}_0) + Q_i - R_0 \right)^+ \right] - E \left[\left(D_0(\tilde{L}_0) + Q_i - (R_0 + Q_0) \right)^+ \right] \right) \quad (88)$$

There is a change of the extra random variable O_i to Q_i and Q_0 . This is due a large order quantity to the lead time demand that is present in their paper. This leads to the assumption that the actual order size is going to be always equal to the order quantity. The authors use the following equations to calculate the mean and variance of the lead time demand.

$$E[D_0(L_0)] = \sum_{i=1}^N \mu_i E[L_0] \quad (89)$$

$$\text{Var}[D_0(L_0)] = \sum_{i=1}^N \left[\sum_{k=0}^{\infty} (\mu_i E[L_0] - k Q_i)^2 s_i^{ord}(k) \right] \quad (90)$$

The s_i^{ord} value is the probability that exactly k subbatch orders of size Q_i are placed from depot i . This is represented as:

$$s_i^{ord}(k) = \begin{cases} \delta_i(0), & \text{if } k = 0, \\ \delta_i(k) - \delta_i(k-1), & \text{if } k > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (91)$$

δ_i is probability of depot i placing at most k orders with the condition that $L_0 = l$. This is a summation or integration over all possible times l , which the authors present next:

$$\delta_i(k) = \int_{l=0}^{\infty} \left(\sum_{x=1}^{Q_i} \frac{1}{Q_i} Pr(D_i(l) \leq kQ_i + x - 1) \cdot pdf_{L_0}(l) \right) dl \quad (92)$$

where pdf_{L_0} is the probability density function of L_0 .

The variance and mean are fitted to the negative binomial distribution, which is used as an alternative to the real distribution function. These previous equations culminate into a similar equation used in Kiesmüller et al. (2004), but fitted to the negative binomial distribution.

$$E[(X - z)^+] = E[X] - \sum_{x=0}^z x f_X(x) - z(1 - F_X(z)), \quad z \geq 0 \quad (93)$$

Equation (93) is used to calculate each expected value in equations A and B. The $f(x)$ and $Fx(z)$ functions are defined in section 3.3.4.

Grob and Bley (2018) claim that the approximation above is more stable. The main issue with the approximation comes from when $Q_i \ll Q_0$. The variance in this case might be a negative value, causing issues to the approximation. The authors also suggest a gamma distribution in cases when the variance is smaller than the mean. Equation 93 can be adapted to this distribution, but the authors do not show this.

4 Simulation Study and Model Overview

This chapter focuses on the detailed overview of the model we will use to conduct the study, the mathematical verification of the five models we wish to compare, and the experimental design of the numerical study. As stated in previously, the approach and design of this study is based on Grob and Bley (2018).

We consider a continuous (R, Q) multi-echelon system with 8 non-identical retailers. The customer demand is assumed to be Poisson. l_i is assumed to be constant with a value of 0. R_i has no effect on delay and is assigned arbitrary values. There will be two numerical studies: the first focuses on complete order fulfillment and the second focuses on partial order fulfillment. The order policy is First-Come-First-Served (FCFS).

4.1 Simulation Model

The simulation model used in this thesis was not developed by us, but by several individuals within the Production Management department at LTH. We have made slight modifications to fit the simulation study included in this thesis. The model is built within ExtendSim, a simulation software. In this section, we will provide an overview of how the model functions and its uses. Appendix B will provide an in-depth look into how to replicate the model.

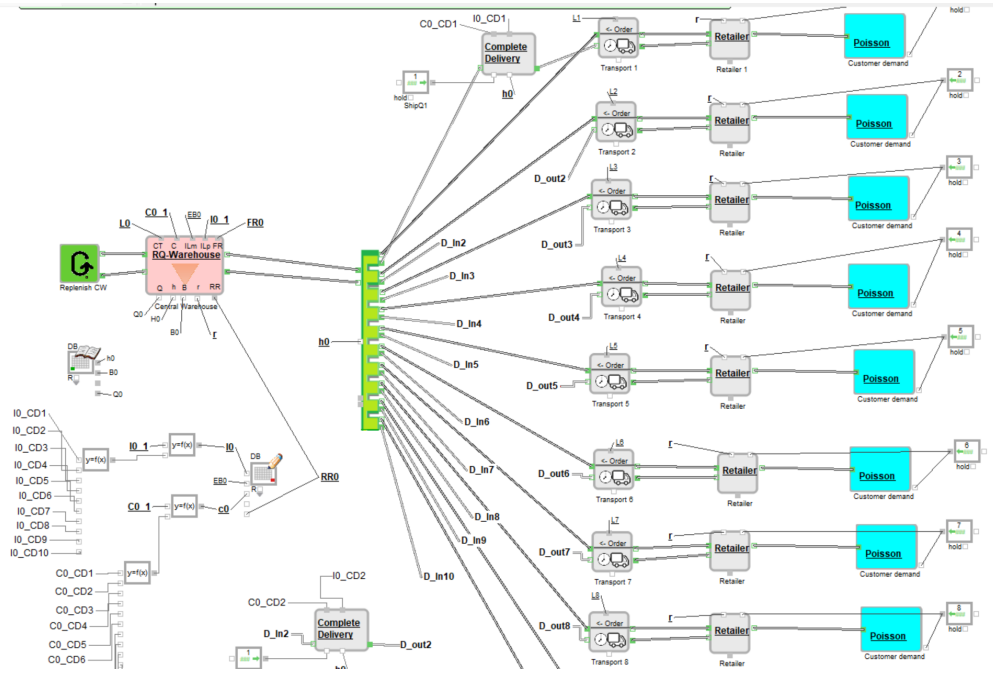


Figure 3: ExtendSim simulation model

Figure 3 is the simulation model in the current version of ExtendSim, ExtendSim 10. The model itself is a representation of a 10 retailer multi-echelon inventory system with an (R, Q) inventory policy. The customer demand allows for multiple types and is currently set to Poisson.

The model consists of 6 parts:

- **Customer Demand Block:** This is the bright blue block within the model. This model allows for different types of customer demand, but it is currently set to Poisson. When an order is generated, it is sent to the retailer block.
- **Retailer Block:** The gray retailer block receives this order and sends the batch request to the transport block. It has two major components. The first being that the retailer block satisfies customer demand. When stock reaches its reorder point, R_i , it sends the batch request to the transportation block. The

block also manages the retailer inventory, logs backorders, and calculates fill rate. Note this fill rate is not of interest in this thesis.

- **Transportation Block:** The gray "transport" block with the truck image simulates the transportation of the delivery. It adds the transportation delay, but for this thesis, it will be set to 0 since we wish to examine only stockout delay.
- **Central Warehouse Block:** The RQ warehouse block with the triangle is the central warehouse. When the central warehouse receives the order, it will check to see if the order can be fulfilled, i.e. there is stock on hand. If that answer is no, there will be a backorder. If that answer is yes, the delivery will move to the green delivery sorter to be able to reach the correct retailer or "Complete Delivery" block.
- **Replenish Block:** The green block next to the central warehouse is the supplier lead time or L_0 . Note that the block in the figure is only shown for one retailer. The other blocks are placed below the model outside the figure to preserve a cleaner look. Each one connects with a small text connector. This block just applies the lead time to an order to a supplier and is a fixed constant.
- **Delivery Type Block:** The gray "Complete Delivery" block. This was a modification to the original model, so it could handle complete deliveries or complete order fulfillment. Logically speaking, this block is still the central warehouse and can be viewed as the loading dock. The block is setup to be able to do both types of order fulfillment. The function of the block checks to see if there is enough stock to ship out. For partial deliveries, the delivery passes right through. For complete deliveries, it checks to make sure the delivery is full for the delivery to move on to the next block.

The model records many different data points that are useful along with unimportant ones. The main outputs include: Cost, Stock on Hand, Units outstanding, Fill rate, and Lead Time. Each output has both the mean and variance of the output data points for each retailer and the central warehouse (if applicable). We specifically use the lead time output as our main data comparison point. However, some of the other parameters can be of use. The central warehouse has outputs that can be used as comparison points to other tests in the experimental design. The central warehouse fill rate is the most important of these as it has a direct consequence on delay. The average stock on hand at the central warehouse can also complement the fill rate data.

Although the model outputs many data points, we are specifically interested in the delay. The lead time output in this model is the addition of stockout delay and transport delay. Since transportation delay will be 0, the lead time output will represent the delay from the central warehouse. The experimental design will incorporate the model as both a comparison for complete deliveries (complete order fulfillment) and to explore delay in partial deliveries (partial order fulfillment).

4.2 Model Verification

Out of the models presented in section 3.3, four has been implemented with Python code to provide a comparison with the simulated data. These models are the Andersson & Marklund Approach, the Axsäter Normal Approximation Model, the Kiesmüller & de Kok mixed-Erlang Distribution Model and the Berling & Farvid Lead Time Model. The complete code is in Appendix A and can easily be set up to run on another computer. The implementation has closely followed the model description in the papers as described in section 3.3. There are four python files, one for each model though they are not entirely independent. The Andersson and Marklund program uses the method *compute_fi_distribution* from the Berling and Farvid program to compute the demand distribution to the warehouse. The method can be used in two different ways depending on if the default parameter *i_exclude* is included in the method call or not.

All files are called with the method *waiting_time* and it returns either the mean or the mean and the variance of the estimated delay. The necessary arguments needed to call on the *waiting_time* method is shown in table 1 though they on some occasions vary by name. While it's difficult to verify the validity of the results provided, the program has been tested for different inputs and it has been observed if the results are reasonable and if the change of input generates the expected directional changes in delay. All programs that are

Argument	Type	Description
arr_vec	list	Arrival intensities for each retailer
Q_vec	list	Order quantities for each retailer
L0	float	Lead time to the warehouse
R0	int	Warehouse reorder point
Q0	int	Warehouse order quantity

Table 1: Input arguments

included is functioning and they have all passed the tests.

4.3 Experimental Design

Both numerical studies follow the same basic structure and will consist of a 2^3 factorial experiment with a high and low value. The main motivation for using this type of experimental design is being able to observe the relationship between certain factors. Being able to change a few factors to observe the effects it has on delay is valuable for providing data for the analysis. A high-medium-low test experiment was considered due to Grob and Bley (2018) using this combination. However, for the scope of this study, it was deemed too time consuming when a high-low combination can still provide robust results. Three factors were chosen for their importance in calculating delay, and for how they would interact with the ExtendSim program. Each factor will be explained in detail and why it was chosen.

Factor A represents the fill rate at the central warehouse (CW). Fill rate is the component that is important in this context. For delay, it is important for us to find an observable number. If the fill rate is very high, there may never be any backorders due to the inventory level being very high and fulfilling all orders it receives. Hence, if we choose a high fill rate, we will never observe any delay. Due to how the ExtendSim model is configured, the fill rate cannot be entered in. It is determined through running the simulation. However, the two main components to determine the fill rate are R_0 and Q_0 . We can use the following approach to help determine a suitable value for each based on the two different sets of λ values used for factor C. The high and low R_0 uses L_0 multiplied by μ_0 , which is added or subtracted by the the square root of μ_0 . Q_0 is simply scaled from the square root of μ_0 . These will produce two different sets of R_0 and Q_0 so they can maintain similar fill rate ranges. Q_0 will remain fixed for both + and -, while R_0 will fluctuate. These equations give a value close to the the fill rate we hope to have which is approximately 40% for - and 60% for +. The R_0 is slightly adjusted to meet this criteria based test simulations to determine the best combinations to get observable results. Table 2 includes the parameters for factor A.

Table 2: R_0 and Q_0 values for factor A.

Test	R_0	Q_0
1	17	4
2	22	4
3	17	4
4	22	4
5	8	7
6	12	7
7	8	7
8	12	7

Factor B is the batch quantities for the retailers, Q_i . Due to having 8 different retailers, we have divided them into two groups: high demand and low demand. A modified EOQ formula (Axsäter 2006) is used in determining the proper values and then multiplying than by the square route of μ_i . All values will be rounded up in the event of a decimal. The variables we set are: $A_H = 200$, $A_L = 100$, and $h = 1$.

$$Q_H = \sqrt{\frac{2A_H\mu_i}{h}} = \sqrt{\frac{2A_H}{h}} * \sqrt{\mu_i} \quad (94)$$

and

$$Q_L = \sqrt{\frac{2A_L\mu_i}{h}} = \sqrt{\frac{2A_L}{h}} * \sqrt{\mu_i} \quad (95)$$

where

$$\mu_i = 1/\lambda \quad (96)$$

Each retailer is computed for both high and low in table 3.

Table 3: High and Low Q_i values for factor B.

Retailer	Qi low	Qi high
1	5	7
2	4	5
3	4	5
4	5	6
5	5	6
6	4	5
7	5	6
8	4	5

Factor C represents both the mean and the variation of the retailers within the system. This is μ_i and σ_i . Due to having a Poisson customer demand, these are known to be the same and can be represented as λ . Each λ can be proportionally increased and decreased by 25% to create a high and low value based on the initial values. In the simulation model, the mean time between arrivals (MTBA) ends up being λ_i due to the Poisson demand. Table 4 is the initial values.

Table 4: λ initial values for factor C.

Retailer Name	λ_i
Retailer 1	10.00
Retailer 2	20.00
Retailer 3	17.00
Retailer 4	14.00
Retailer 5	13.00
Retailer 6	19.00
Retailer 7	13.75
Retailer 8	18.00

Table 5 represents each high and low value within the experiment.

Table 5: Variable values for each experiment

Factor	Low Level (−)	High Level (+)
Factor A	$R_{0L} = 17/8$ $Q_0 = 4/7$	$R_{0H} = 22/12$ $Q_0 = 4/7$
Factor B	$Q_i = Q_L$	$Q_i = Q_H$
Factor C	$\lambda_{iL} = \lambda_i - (0.25 * \lambda_i)$	$\lambda_{iH} = \lambda_i + (0.25 * \lambda_i)$

The value of each variable was selected after several test runs. This is to ensure that we use values that allow us to observe the effects of delay. For example, if R_0 is set too high and gives the central warehouse a high fill rate, then there will be little or no observable delay within the system. Taking this into account allows us to set the low and high values to a reasonable value, rather than just an extreme value.

For the setup of the experiment, we have 8 separate test cases to consider.

Table 6: Template for each experiment

Test	A (R_0, Q_0)	B (Q_i)	C (λ_i)	Stockout Delay (y)
1	−	−	−	
2	+	−	−	
3	−	+	−	
4	+	+	−	
5	−	−	+	
6	+	−	+	
7	−	+	+	
8	+	+	+	

Table 6 is our template for each of the experiments that we will run. We will compare the 5 models we previously discussed in addition to the simulation model. These will be for complete order fulfillment. The final run will be with the partial order fulfillment in the simulation study. The models we examined either assume complete order fulfillment or do not provide enough tools to convert the calculations into partial fulfillment. Therefore, we will not test those with partial order fulfillment and focus on the simulation model.

5 Results and Analysis

The results and analysis from the numerical study will be presented in this chapter. Larger and more expansive tables will be presented in full in Appendix C.

5.1 Results

The numerical study followed the structure of the previous chapter, Tables 6 and 5. All the experiments used generic/undefined time units (if applicable) with a total of 8 runs. The 8 runs were chosen from the simulation model due to the model results not becoming any more accurate. Both types of simulation model designs were ran for a total of 7000 generic time units for a total of 8 runs. For reference, Table 7 will state what each factor is.

Table 7: Factor Descriptions and Variables

Factor	Description
Factor A	CW fill rate: R_0 and Q_0
Factor B	Retailer batch size: Q_i
Factor C	Retailer MTBA: λ_i

The variables were kept the same between partial and complete order fulfillment numerical studies. Full tables and calculations are included in Appendix C.

Table 8: Results of Delay, Standard Deviation, and CW FR Averages for both simulations.

Test	Complete Fulfillment			Partial Fulfillment		
	Delay Avg	S.D. Avg	CW FR Avg	Delay Avg	S.D. Avg	CW FR Avg
1	6.9872	0.2004	0.3624	5.4448	0.1686	0.3695
2	2.9370	0.1207	0.6486	2.3569	0.1003	0.6260
3	8.2002	0.2078	0.3767	6.1859	0.2179	0.3826
4	4.0529	0.2029	0.6106	2.8546	0.1602	0.6274
5	8.9799	0.3178	0.4040	7.4089	0.2828	0.3993
6	4.3643	0.2674	0.6566	3.1934	0.1846	0.6498
7	11.3339	0.4779	0.4049	8.1100	0.3249	0.4162
8	5.9098	0.3667	0.6157	4.2472	0.2659	0.6122

Table 8 reports both the expected lead time and standard deviation averages from all 8 retailers for each test. To examine the main effects and factor relationship, the contrast method and t-tests, which are outlined in Montgomery (2017), were used to obtain the results in Table 9.

Table 9: Main Effect Comparison for Delay: Complete and Partial Order Fulfillment

Complete Order Fulfillment			Partial Order Fulfillment		
Factor	Main Effect	Margin of Error	Factor	Main Effect	Margin of Error
A	-4.559301545	significant	A	-3.624361242	significant
C	2.102675917	significant	C	1.529336231	significant
B	1.557106286	significant	B	0.748435213	significant
AC	-0.460557070	significant	AC	-0.414762209	significant
BC	0.392652153	significant	ABC	0.149006192	significant
AB	-0.226426512	significant	BC	0.129034812	significant
ABC	-0.177836320	significant	AB	0.027272069	not

The same process is used to determine the main effect for the standard deviation of the delay. The results are presented in Table 10.

Table 10: Main Effect Comparison for the Standard Deviation: Complete and Partial Order Fulfillment

Complete Order Fulfillment			Partial Order Fulfillment		
Factor	Main Effect	Margin of Error	Factor	Main Effect	Margin of Error
C	0.1744875967	significant	C	0.1027973872	significant
B	0.0872262172	significant	A	-0.0707763138	significant
A	-0.0615220812	significant	B	0.0581481949	significant
BC	0.0424380129	significant	AB	0.0124394055	not
ABC	0.0339096917	not	AC	-0.0078023772	not
AC	-0.0192581661	not	ABC	0.0071648765	not
AB	-0.0034915945	not	BC	0.003583985594	not

Note the effects in Table 9 and 10 are proportionally the same compared to the average values.

The corresponding estimate of the average $E[L]$ for four available models are presented in Table 11. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) were calculated to provide a comparison metric. Note that Grob and Bley (2018) was not included in this comparison. The differences between it and Kiesmüller et al. (2004) are too little to justify including in the comparison. The authors state that the improvement made helps in specific scenarios, which is not the case for this numerical study. Table 11 are the comparisons of the averages of all the retailers per test with the simulation result: Simulation (SIM), Andersson and Marklund (A & M), Axsäter (Ax), Kiesmüller and de Kok (K & DK), and Berling and Farvid (B & F).

Table 11: Comparison of the available models with the simulation

Test	SIM	A & M	Ax	K & DK	B & F
1	6.987161	5.735187	5.525807	8.006150	6.925325
2	2.937007	2.357466	2.223514	4.337900	2.863438
3	8.200206	6.468730	6.191531	9.198175	7.856413
4	4.052871	3.033276	2.863104	5.265988	3.635313
5	8.979906	7.480917	7.361032	11.226650	9.540425
6	4.364310	3.420895	3.301229	5.901250	4.174063
7	11.333927	8.605669	8.364809	12.930163	11.358000
8	5.909806	4.431425	4.225777	7.345838	5.541588
MAE	N/A	1.403954	1.588549	1.430865	0.254977
MSE	N/A	1.529170	1.713821	1.478539	0.313912

For a complete overview, we refer to Appendix C, which will contain the complete complied table with each retailer and test. The complete analysis will be included in the next section.

5.2 Analysis

The Analysis will focus on the results presented in the previous section and how each factor contributes as a driving force in those results. Table 9 and 10 should be referenced for this analysis. We first will examine complete order fulfillment factor relationship.

Factor A The margin of error is considered significant with a large negative value. This means that when the fill rate is low, the delay is significantly impacted. This is also the highest of all factors and should be the number one priority in a system if the aim is to minimize delay.

For the standard deviation, we see a slightly different result. While still significant, it is lower than Factor B and C. It is still negative meaning that with a low fill rate, it negatively effects the standard deviation of the delay.

This is expected because a higher fill rate increases the service level. Therefore, a high service level effectively reduces the possibility of delays.

Factor B Out of the three main factors, B is the lowest of the significant margin of errors. It is positive however, meaning that an increased Q_i value negatively effects the delay.

A similar outcome is seen for the standard deviation. It is still positive meaning that an increased Q_i value negatively effects the standard deviation of the delay.

There is a less clear reason on why factor B behaves in this way. In a complete order fulfillment policy, this is a limiting factor to how quickly an order reaches the retailer. If it is too high, there is a possibility that the central warehouse might not have enough stock on hand to cover the order, causing delays. A lower value moves it closer to a unit by unit policy. The same can be said in terms of variation of the delay in the system.

Factor C This factor comes as an expected result. With a higher λ_i value, the effects on the delay increase. This can be seen as an increase in variability and signals this in the standard deviation results being the most significant value.

Factor AB The relationship between the fill rate and Q_i remains a significant margin of error. However, it is much less than the three main factors. Being negative, we see a low value for both negatively effects the delay of the system.

There is a much different result for the standard deviation. It is not significant and therefore, it can be concluded that it does not factor into the standard deviation of the delay. This is expected as the combination causes more delay, but stays consistent in the absence of factor C.

The reasoning behind this outcome may be effected by the strong importance of factor A within the system. There is a high service impact within this relationship, indicating that this is why it is significant.

Factor AC Out of all relationships, this factor is the most significant. When both are low, they significantly effect the delay. This is expected as both factor A and C are obvious in how delay is effected due to a higher service level can mitigate the increase variation brought on by a high MTBA. Impact is lower when the service level is high when demand is more stochastic.

We see a similar outcome for standard deviation as Factor AB with it being not significant. However, this shows that factor A mitigates the variation caused by factor C due A controlling the stock on hand.

Factor BC This factor relationship faces a similar outcome with the exception that it significantly effects the standard deviation of the delay. This may be due to a higher Q_i value in combination with higher variability of λ_i is significant enough to effect the overall delay standard deviation. This is expected due to factor B being a limiting factor on speed of delivery if the value is high. Therefore, in combination with higher MTBA, we see a higher effect on the standard deviation through higher variation.

Factor ABC

The lowest of significance shows us that the individual factors are more of a priority than the combination of them. Being negative points this relationship towards being more the high values. This is expected as the individual factors have a greater effect than together. Optimization in inventory control normally focuses on a single factor, which is why the results point towards the individual factors. The main takeaway from this is that the standard deviation is not significant. More frequent deliveries does not necessarily mean longer delays. This could be from higher service levels from factor A. It does raise the question of what circumstances for the overall system could shift these values.

The main takeaway from the complete order fulfillment results are that the main factors (A, B, and C) are the largest contributors to effecting delay with A being the most. If one would want to reduce delay, it points towards the central warehouse fill rate as being the area to target.

Next we will examine partial order fulfillment. Note that the overall delay per factor is decreased.

Factor A The fill rate follows the exact same path and expected reason as the complete order fulfillment for both delay and standard deviation of the delay.

Factor B The same is almost true for factor B. The main difference is a much lower effect on the standard deviation. It is still significant and similar in value, but less important compared to other factors. This is expected based on how our simulation model defines partial fulfillment. It does this on a unit by unit basis. So, even though batch size is the same between the policy types, a larger batch size can still have units not delivered, therefore effecting the delay.

Factor C λ_i shows the exact same outcome for both delay and the standard deviation as the complete order policy.

Factor AB The largest change from complete to partial order fulfillment is factor AB. It becomes not significant, which means this relationship does not have a significant effect on the delay. The same is true for standard deviation. This is expected as factor B becomes less important due to it not having to wait for complete orders to ship out.

Factor AC This relationship importance remains the same as the complete order fulfillment for both delay and standard deviation of the delay.

Factor BC This relationship remains significant, however, less significant. This is expected as Q_i becomes less important with partial deliveries. However, it is not significant when it comes to standard deviation.

Factor ABC The combined relationship remains significant, however, in the opposite direction. In a partial delivery system, the relationship shifts toward the lower value. Note that this is still much less than the three main factors. This is expected due to how factor B has become less important. The shift to a positive value indicates this change to factor B.

The key result of this remains on the three main factors (A, B, and C).

In comparison between the two fulfillment types, we see an overall decrease in delay. The factor relationships or importance do not change too much from one to the other. The standard deviation does seem to decrease as well. Factor A is the most important factor between these two types. See Appendix C for complete results.

5.2.1 Model comparison and Ranking

The final component is to compare the available models, determine any interesting trends, and provide a ranking. The complete the results are located in Appendix C. We have compiled a summary in Table 12.

Table 12: Summary of MAE and MSE Results

Test	Metric	A & M	Ax	K & DK	B & F
Test 1	MAE	1.251974175	1.461354455	1.018988546	0.255031414
	MSE	1.306212868	1.474540495	1.115837696	0.281233757
Test 2	MAE	0.579541307	0.713493171	1.400892829	0.115204666
	MSE	0.594885852	0.735118017	1.415756711	0.142360547
Test 3	MAE	1.731475464	1.796991692	0.997969220	0.434600689
	MSE	1.772943635	1.928750569	1.216394177	0.491952836
Test 4	MAE	1.019595299	1.189767113	1.213116387	0.417558612
	MSE	1.034188603	1.208593681	1.266526915	0.449373566
Test 5	MAE	1.498988786	1.618873968	2.246744032	0.560519032
	MSE	1.585974300	1.761583300	2.293224647	0.637811897
Test 6	MAE	0.943414952	1.063081185	1.536939815	0.381391770
	MSE	1.059854143	1.167711280	1.623491295	0.427149584
Test 7	MAE	2.728258150	2.969118240	1.596235260	0.388978762
	MSE	2.762061889	3.058120193	1.761633673	0.521320263
Test 8	MAE	1.478380389	1.684028793	1.436031707	0.400272041
	MSE	1.515652064	1.736446068	1.584906047	0.498881990

First, we can examine on a model by model basis. The outlier here is Berling and Farvid (2014). Across every test, it performs the best and closest to the simulated value. It is more sensitive in Test 5 through 8, but that seems to be common across all the models. This could be explained by the increase in MTBA, which will increase variation making it harder to approximate. Axsäter (2003) tends to perform worse with the exception of Test 2 and 5. Andersson and J. Marklund (2000) is less sensitive on a test by test basis but more on a model wide. (See Table 11.) The opposite is true for Kiesmüller et al. (2004). When considering the model wide evaluation, much of the same holds. With that being said, the ranking is as follows:

1. **Berling and Farvid Lead Time Model**
2. **Kiesmüller and de Kok Mixed-Erlang Distribution Model**
3. **Andersson and Marklund METRIC Approach**
4. **Axsäter Normal Approximation Model**

Berling and Farvid are the clear best model based on the parameters ran through our code and compared to the simulation model. Kiesmüller and de Kok, and Andersson and Marklund are very close to each other. Due to the MSE being higher for the Andersson and Marklund approach, the Kiesmüller and de Kok model was chosen as the second choice. It is interesting to note that the Andersson and Marklund approach is a METRIC approach and still performs well given the other models examined. The remaining question that comes up is the computational difficulty of these models. To fully answer that question, these methods would need to be used and tested in a variety of different multi-echelon models.

The final area to examine is the trends within the test by test summary in Table 12. When λ_i is set to its high value (Test 5-8), we see a much larger jump from the MAE to the MSE. This means all models struggle more with the higher λ_i value, or a larger MTBA. The greater variability in the system cause these models to perform worse than the simulation value. This is not surprising, as stated previously, stochastic variables are difficult to approximate. Referring to the complete table in Appendix C, Some other interesting observations include Kiesmüller and de Kok’s model overshoots the simulation value while Andersson and Marklund’s model and Axsäter’s model tend to undershoot the simulation. Berling and Farvid produce a mix of over and under. No other observable trends or phenomenon were found in the tests.

Overall, all models still provide good approximations that give variety to choose from although more variety would be better. The limited available research in delay has limited comparisons. We have only found one other comparison in Grob and Bley ([2018](#)).

6 Future research

We have identified some key areas where future research can be directed. Firstly, there is still a need for more practical tests of these models. Being able to test the models with real data will be the ultimate comparison tool. It would also show which models would be best in certain situations such as different demand types or more unstable demand. Other factors could be included such as varying L_0 . Second, there is still a lack of dedicated research to stockout delay. A majority of research in stockout delay is a small component of a much larger model. The importance of being able to properly approximate the delay helps reduce uncertainty and unpredictability within a system along with being able to optimize different parameters better. Lastly, the final area of importance is the lack of partial order fulfillment representation. This is a very important fulfillment type, especially in lumpy or sporadic demand such as spare parts. This leads us into some ongoing research being done, which will be covered in the next section.

6.1 Ongoing Research

There is current and ongoing research, Marklund (2026), that introduces a new way to estimate stochastic stockout delay at a central warehouse. Due to the research being in progress, there is limited information we are allowed to share and will be done so here. There will be a general description along with how the author is determining the stockout delay within their model. Complete detailed information will be published in the future.

There are two unique variables to define and they are written as Marklund (2026) has written them.

W_i Stochastic stockout delay at the central warehouse for units ordered by retailer, i
 $\gamma_{0,i}$ proportion of the demand from retailer i that can be satisfied by the central warehouse

The system that is being examined assumes partial order fulfillment with a complex demand process that involves virtual retailers and having retailer order quantities greater than 1. The inventory policy is (R, Q) and the approximation method is referred to as SERVDIFF. The author uses the same approach approximation as a compound Poisson demand model to estimate the stockout delay. The author follows presents the stockout delay approximation in equation (97).

$$E[W_i] = \frac{\bar{B}_{0,i}}{\mu_i} = \frac{1}{\mu_i} \frac{\mu_i(1 - \gamma_{0,i})}{\sum_{i=1}^{N+1} \mu_i(1 - \gamma_{0,i})} \bar{B}_0 = \frac{(1 - \gamma_{0,i})}{\sum_{i=1}^{N+1} \mu_i(1 - \gamma_{0,i})} \bar{B}_0 \quad (97)$$

where

$$\gamma_{0,i} = \frac{\sum_{j=1}^{R_0+Q_0} \sum_{k=1}^{k_{min}} \min(j, kQ_i) P(O_{0,i} = kQ_i) P(IL_0 = j)}{\sum_{k=1}^{k_{min}} kQ_i \cdot P(O_{0,i} = kQ_i)} \quad (98)$$

The overall structure of W_i is the same, however, the calculation of $\gamma_{0,i}$ is slightly different. It includes the concept of "k orders of Q_i units are required to bring the inventory position back to $[R_i + 1, R_i + Q_i]$ if $j + (k - 1)Q_i \leq O_i \leq j - 1 + kQ_i$ Marklund (2026)." This affects the probability of the number of units ordered by a customer at a particular retailer, O_i , hence the change from d to kQ_i . This incorporates the much more complex structure of the author's model. It will be interesting to see how this is applied and tested in practice.

7 Conclusion

In this thesis, an inventory system with one warehouse and eight retailers was modeled in ExtendSim to provide a realistic benchmark against which the analytical models could be evaluated. The models are supposed to work under various circumstances such as different demand distributions, warehouse lead times and in the case of Kiesmüller et al. (2004) even with an arbitrary number of echelons. Though this study has been limited in scope and only a subset of all possible variables that could influence the stockout delay have been tested it provides an evaluation of the model's data against a measure of comparison. The results indicates that Berling and Farvid's model gives the best overall estimates and it handles both larger and smaller delays.

There are still many more things that could be considered. For example the variance estimate given by the models could be of interest. An accurate stockout delay is important in regard to computing an accurate retailer lead time distribution which in turn gives more accurate inventory level probabilities. Some models may score better on an estimate of the mean but the most important aspect is that it yields an accurate lead time demand distribution. In Andersson and J. Marklund (2000) they provide a method for deriving a probability distribution of the delay, while in Berling and Farvid (2014) they fit the mean and variance estimate of the delay to a log-normal distribution. Further, Kiesmüller et al. (2004) fits the first and second moment of the delay to a mixed-Erlang distribution to model the effective lead time. The next step could be to evaluate both the estimates and the distributions together and find models that captures the important parts of the lead time demand distribution. To widen the scope and applicability of the results, an important extension would be to evaluate a delay model accommodating partial deliveries. A first step could be the model in section 6.1 which could be compared to the results in this thesis.

References

- Andersson, J., S. Axsäter, and J. Marklund (1998). “Decentralized multiechelon inventory control”. In: *Production and Operations Management* 7.4, pp. 370–386.
- Andersson, J. and J. Marklund (2000). “Decentralized inventory control in a two-level distribution system”. In: *European Journal of Operational Research* 127.3, pp. 483–506.
- Axsäter, S. (1990). “Modelling emergency lateral transshipments in inventory systems”. In: *Management Science* 36.11, pp. 1329–1338.
- (2003). “Approximate optimization of a two-level distribution inventory system”. In: *International Journal of Production Economics* 81, pp. 545–553.
- (2006). *Inventory Control*. Boston, MA: Springer US.
- Berling, P. and M. Farvid (2014). “Lead-time investigation and estimation in divergent supply chains”. In: *International Journal of Production Economics* 157, pp. 177–189.
- Berling, P., L. Johansson, and J. Marklund (2023). “Controlling inventories in omni/multi-channel distribution systems with variable customer order-sizes”. In: *Omega* 114, p. 102745.
- Berling, P. and J. Marklund (2013). “A model for heuristic coordination of real life distribution inventory systems with lumpy demand”. In: *European Journal of Operational Research* 230.3, pp. 515–526.
- (2014). “Multi-echelon inventory control: an adjusted normal demand model for implementation in practice”. In: *International Journal of Production Research* 52.11, pp. 3331–3347.
- Chatfield, D. C., J. C. Hayya, and D. P. Cook (2013). “Stockout propagation and amplification in supply chain inventory systems”. In: *International Journal of Production Research* 51.5, pp. 1491–1507.
- Cho, D. I. and M. Parlar (1991). “A survey of maintenance models for multi-unit systems”. In: *European Journal of Operational Research* 51.1, pp. 1–23.
- De Kok, T. et al. (2018). “A typology and literature review on stochastic multi-echelon inventory models”. In: *European Journal of Operational Research* 269.3, pp. 955–983.
- Deuermeyer, B. and L. B. Schwarz (1981). “A model for the analysis of system service level in warehouse/retailer distribution systems: the identical retailer case”. In: *Multi-level production, inventory control systems: theory and practice*. Ed. by L. B. Schwarz. Amsterdam: North-Holland Publishing Company, pp. 163–193.
- Fleischmann, M., J. A. Van Nunen, and B. Gräve (2003). “Integrating closed-loop supply chains and spare-parts management at IBM”. In: *Interfaces* 33.6, pp. 44–56.
- Grob, C. and A. Bley (2018). “Comparison of wait time approximations in distribution networks using (R, Q)-order policies”. In: *arXiv preprint arXiv:1801.09617*.
- Hadley, G. and T. M. Whitin (1963). *Analysis of Inventory Systems*. Prentice-Hall, Inc.
- Hawkins, D. T. and R. Wagers (1982). “Online Bibliographic Search Strategy Development”. In: *Online* 6.3, pp. 12–19.
- Hillier, Frederick S. and Gerald J. Lieberman (2015). *Introduction to Operations Research*. 10th. New York, NY: McGraw-Hill.
- Johansson, L. et al. (2020). “Controlling distribution inventory systems with shipment consolidation and compound Poisson demand”. In: *European Journal of Operational Research* 280.1, pp. 90–101.
- Kiesmüller, G. P. et al. (2004). “Evaluation of divergent N-echelon (s, nQ)-policies under compound renewal demand”. In: *OR Spectrum* 26.4, pp. 547–577.

- Kruse, W. K. (1980). “Waiting time in an S-1, S inventory system with arbitrarily distributed lead times”. In: *Operations Research* 28.2, pp. 348–352.
- Lee, H. L. and C. Billington (1993). “Material management in decentralized supply chains”. In: *Operations Research* 41.5, pp. 835–847.
- Marklund, J (2026). “Work in process paper”. Manuscript in preparation. URL: <https://portal.research.lu.se/en/persons/johan-marklund/>.
- Montgomery, Douglas C. (2017). *Design and Analysis of Experiments*. 9th. Hoboken, NJ: John Wiley & Sons. ISBN: 978-1119113478.
- Sherbrooke, C. C. (1968). “METRIC: A multi-echelon technique for recoverable item control”. In: *Operations Research* 16.1, pp. 122–141.
- Stenius, O. et al. (2016). “Exact analysis of divergent inventory systems with time-based shipment consolidation and compound poisson demand”. In: *Operations Research* 64.4, pp. 906–921.
- Svoronos, A. and P. Zipkin (1988). “Estimating the performance of multi-level inventory systems”. In: *Operations Research* 36.1, pp. 57–72.
- Tempelmeier, H. (1983). “Inventory control using a service constraint on the expected customer order waiting time”. In: *European Journal of Operational Research* 19.3, pp. 313–323.

A Model Code for the Tested Models

Axsater.py

```
import sys
import scipy.stats as sp
import numpy as np

def H(x):
    return (1/2)*((x**2+1)*(1-sp.norm.cdf(x)) - x*sp.norm.pdf(x))
def G(x):
    return sp.norm.pdf(x)-x*(1-sp.norm.cdf(x))

def waiting_time(arr_vec,Q_vec,L0, Q0,q, R0):

    def p_ik_fun(k,mu_i_L0,sigma_i_L0,Q_i,L0):
        term1 = (sigma_i_L0/Q_i)
        term2 = G(((k-1)*Q_i-mu_i_L0)/sigma_i_L0)
        term3 = G(((k+1)*Q_i-mu_i_L0)/sigma_i_L0)
        term4 = 2*G((k*Q_i-mu_i_L0)/sigma_i_L0)
        return term1*(term2+term3-term4)

    mu_w_vec = []
    var_w_vec = []
    for i in range(len(arr_vec)):
        mu_i_L0 = arr_vec[i]*L0
        var_i_L0 = arr_vec[i]*L0

        lowest_order_prob = 1e-5
        k = 1
        p_ik = {}

        while sum(p_ik.values()) < 1- lowest_order_prob:
            p_ik[k] = p_ik_fun(k,mu_i_L0,np.sqrt(var_i_L0),Q_vec[i],L0)

            if len(p_ik) == 1:
                k = k +1
            else:
                k_min = min(p_ik.keys())
                k_max = max(p_ik.keys())
                if p_ik[k_max] > p_ik[k_min]:
                    k = k_max+1
                else: k = k_min-1

        mu_w_vec.append(mu_i_L0)
        var_w_vec.append(np.sum([(k*Q_vec[i]-mu_i_L0)**2*p for k,p in p_ik.items()]))

    mu_w_L0 = np.sum(mu_w_vec)
    var_w_L0 = np.sum(var_w_vec)

    #calculate  $E[IO^-]$ 
    if q != Q0:
        term1 = var_w_L0/(Q0-q)
        term2 = H((R0+q-mu_w_L0)/np.sqrt(var_w_L0))
        term3 = H((R0+Q0-mu_w_L0)/np.sqrt(var_w_L0))
```

```

    E_IO_m = term1*(term2-term3)
else:
    E_IO_m = np.sqrt(var_w_L0)*G((R0+Q0-mu_w_L0)/np.sqrt(var_w_L0))

#mu_d = mean of delay
mu_d = E_IO_m/np.sum(arr_vec)

if Q0 > q:
    term1 = np.sqrt(var_w_L0/(Q0-q))
    term2 = G((R0+Q0-mu_w_L0)/np.sqrt(var_w_L0))
    term3 = G((R0+q-mu_w_L0)/np.sqrt(var_w_L0))
    p_d_is0 = term1*(term2-term3) +1
elif Q0 == q:
    p_d_is0 = sp.norm.cdf((R0+Q0-mu_w_L0)/np.sqrt(var_w_L0))
else:
    sys.exit("q can 't be larger than Q0")

alpha = sp.norm.ppf(p_d_is0)
sigma_delta = mu_d/G(alpha)
mu_delta = -1*alpha*sigma_delta
var_d = sigma_delta**2*(1-p_d_is0)+mu_delta*mu_d-mu_d**2

return mu_d, var_d

```

Berling_Farvid.py

```

import numpy as np
import scipy.stats as sp

def convolve(f, g):
    length = len(f) + len(g) - 1
    return [conv_t(f, g, t) for t in range(length)]

def conv_t(f, g, t):
    return sum([
        f[i]*g[t-i] for i in range(len(f))
        #checking the bounds
        if t-i < len(g) and (t-i) >= 0
    ])

def compute_fi_distribution(arr_vec, Q_vec, L0, i_exclude=None):
    """
    Computes the PMF of  $f_{i_e0i}$  with convolution
    """
    nbr_of_retailers = len(arr_vec)
    # Compute g_j for each retailer in units
    raw_list = [g_j(arr_vec[j], Q_vec[j], L0, modification=True) if j == i_exclude
                else g_j(arr_vec[j], Q_vec[j], L0)
                for j in range(nbr_of_retailers)]

    # Start with first retailer
    combined = raw_list[0]

    for j in range(1, nbr_of_retailers):
        combined = np.convolve(combined, raw_list[j])
        #combined = convolve(combined, raw_list[j])

```

```

    return combined

def g_j(arr_j, Q_j, L0, modification=False):
    '''
    s_j : lambda function that returns P(retailer j orders <= n batches)
    d_j : lambda function that return P(Demand at j < x during time t)
    '''
    d_j = lambda x,t : sp.poisson.cdf(x-1,t*arr_j)
    lowest_order_prob = 1e-5
    g_j_vec = []
    total_prob = 0
    n = 0 #nbr of orders

    if modification:
        s_j = lambda n : d_j((1+n)*Q_j, L0)
    else:
        s_j = lambda n : (1/Q_j)*np.sum([d_j(n*Q_j+z, L0) for z in range(1, Q_j+1)])

    while total_prob <= 1 - lowest_order_prob:
        if n == 0: #same as u == 0
            g_j_vec.append(s_j(n))
            total_prob += s_j(n)
        else:
            g_j_vec.append(s_j(n)-s_j(n-1))
            total_prob += s_j(n) - s_j(n-1)
        g_j_vec.extend([0]*(Q_j-1))
        n += 1

    return g_j_vec

def waiting_time(arr_vec, Q_vec, R0, Q0, L0):
    '''
    :param arr_vec: vector with arrival intensity, same length as Q_vec
    :param Q_vec: vector with order quantity for the retailers
    :param Q0: order quantity for warehouse
    :param R0: warehouse reorder point
    :param L0: leadtime to warehouse
    :return E_delta : vector with the estimated delay for every retailer i
    :return E_delta2 : vector with the second moment of the delay for every retailer i
    '''
    if R0 < 0:
        raise ValueError("R0 must be non-negative")

    E_delta = []
    E_delta2 = []
    for i in range(len(arr_vec)):
        E_delta_i = 0.0
        E_delta2_i = 0.0
        combined = compute_fi_distribution(arr_vec, Q_vec, L0, i_exclude=i)
        if Q_vec[i] <= R0:
            for x in range(R0 + 1, R0 + Q0 + 1):
                for xi in range(x-Q_vec[i]+1, len(combined)):
                    if combined[xi] > 0:
                        delay = 1 - (x-Q_vec[i])/xi

```

```

        E_delta_i += combined[xi]*delay
        E_delta2_i += combined[xi]*delay**2
    else:
        E_delta_i = Q_vec[i] - R0 - 2
        E_delta2_i = Q_vec[i] - R0 - 2
        for x in range(Q_vec[i], R0 + Q0 + 1):
            for xi in range(x-Q_vec[i]+1, len(combined)):
                if combined[xi] > 0:
                    delay = 1 - (x-Q_vec[i])/xi
                    E_delta_i += combined[xi]*delay
                    E_delta2_i += combined[xi]*delay**2

        E_delta_i = (L0 / Q0) * E_delta_i
        E_delta2_i = (L0 ** 2 / Q0) * E_delta2_i
        E_delta.append(E_delta_i)
        E_delta2.append(E_delta2_i)
        var_delta = [E_delta2 - E_delta**2 for E_delta2, E_delta in zip(E_delta2, E_delta)]
    return E_delta, var_delta

```

Andersson_Marklund.py

```

import Berling_Farvid as bf
import numpy as np
import scipy.stats as sp
import sys

def waiting_time(arr_vec, Q_vec, L0, R0, Q0, Q):
    """
    Q_vec: vector with Q_i for each retailer, i.e expressed in units
    Q : size of sub-batch in nbr of units
    arr_vec : vector with the arrival intensities
    Q0: expressed in units of Q
    R0 : expressed in units of Q
    """
    #check if supplied Q_vec is integer of Q
    if(any([Q_i % Q != 0 for Q_i in Q_vec])):
        sys.exit("All Q_i must be integer multiples of Q")

    Q_vec = np.asarray(Q_vec)
    arr_vec = np.asarray(arr_vec)

    B0_pmf = backorder_avg(arr_vec, Q_vec, L0, R0, Q0, Q)
    lmbda_0 = np.sum(arr_vec/Q_vec) #demand intensity at the WH in number of retailer orders

    Q_bar = (np.sum(arr_vec) / lmbda_0)/Q
    j_values = np.arange(len(B0_pmf))
    E_B0 = np.sum(j_values*B0_pmf)
    E_ceil = np.sum(np.ceil(j_values/Q_bar)*Q_bar*B0_pmf)
    E_B0_r = E_ceil - E_B0

    mu_0 = np.sum(arr_vec) #mu_0 = sum of mu_i
    delta1_bar = E_B0*Q/mu_0
    delta2_bar = E_B0_r*Q/mu_0

    n1 = np.sum(arr_vec*(Q_vec - Q))
    N = len(arr_vec)

```

```

#li is set to 0 in our simulations
Li_bar_vec = [delta1_bar + delta1_bar*((arr_vec[i]*(Q_vec[i]-Q)-(n1/N))/n1) + delta2_bar*(arr_vec[i]
return Li_bar_vec

def backorder_avg(arr_vec, Q_vec,L0,R0,Q0,Q):
    """
    Q_vec: vector with Q for each retailer expressed in units
    Q0: expressed in units of Q
    R0: expressed in units of Q
    """
    demand_pmf = compute_D0_distribution(arr_vec, Q_vec, Q,L0)
    max_j = len(demand_pmf)
    BO_pmf = np.zeros(max_j)

    for j in range(max_j):
        for y0 in range(R0+1, R0+Q0+1):
            idx = y0 + j
            if idx < len(demand_pmf):
                BO_pmf[j] += demand_pmf[idx]
            BO_pmf[j] /= Q0
    return BO_pmf

def compute_D0_distribution(arr_vec, Q_vec,Q, L0):
    """
    Computes the PMF of D0 in units of sub-batches (Q) with convolution
    arr_vec : vector with arrival intensities
    Q_vec : vector with Q_i for each retailer in units
    Q : size of sub-batch in units
    L0 : warehouse lead-time
    """
    nbr_of_retailers = len(arr_vec)
    # Compute g_j for each retailer in units
    raw_list = [g_j(arr_vec[j], Q_vec[j],Q,L0)
                for j in range(nbr_of_retailers)]
    # Start with first retailer
    combined = raw_list[0]

    for j in range(1, nbr_of_retailers):
        combined = np.convolve(combined, raw_list[j])

    return combined

def g_j(arr_j, Q_j,Q,L0):
    """
    s_j : lambda function that returns P(retailer j orders <= n batches) = P(Di(L0) < n*(qi+1)
    d_j : lambda function that return P(Demand at j < x during time t) in units
    """
    q_i = Q_j // Q
    d_j = lambda x, t: sp.poisson.cdf(x - 1, t * arr_j)
    s_j = lambda n : (1/Q_j)*np.sum([d_j(n*Q_j + z,L0) for z in range(1,Q_j+1)])

    lowest_order_prob = 1e-5
    g_j_vec = []
    total_prob = 0
    n = 0 # nbr of orders

```

```

while total_prob < 1-lowest_order_prob:
    if n == 0: # same as u == 0
        p = s_j(n)
    else:
        p = s_j(n) - s_j(n-1)
    g_j_vec.append(p)
    total_prob += p
    g_j_vec.extend([0] * (q_i - 1)) #the index should match the demanded nbr of sub_batches
    n += 1
return np.array(g_j_vec)

```

Kiesmuller_DeKok.py

```

import scipy.stats as sp
import numpy as np
import pandas as pd
import math as m
from scipy.integrate import quad
from scipy.special import factorial

def E_A_j_mean(arr_vec, Q_vec):
    """
    Here we find the first moment of A_j by implementing eq. 3. Here j refers to the central warehouse
    that j <- M{w/o}E_1 and k refers to the retailers. To find the first moment of T_k eq.16 is implemented
    inferred that E[R_k] = Q_k and that E[A_k] = 1/arrival rate
    1/E[T_k] = 1/(Q_k*(1/arrival_rate)) = arrival_rate/Q_k
    It returns a scalar
    """
    arr_vec = np.asarray(arr_vec, dtype=float)
    Q_vec = np.asarray(Q_vec, dtype=float)
    return 1 / np.sum(arr_vec / Q_vec)

def _inner_integral(x, arr, Q):
    n = np.arange(Q)
    mu = arr * x
    pmf = np.exp(-mu) * mu**n / factorial(n) # P(K=n) for each n
    cdf = np.cumsum(pmf)
    return np.sum(cdf) / arr

def _inner_integral2(x, arr, Q):
    mu = arr * x
    result = 0.0
    exp_term = np.exp(-mu) / arr
    for n in range(int(Q)):
        m = np.arange(n + 1)
        inner = np.sum(mu**m / factorial(m))
        result += exp_term * inner
    return result

def _integrand(x, arr_vec, Q_vec):
    prod = 1.0
    for arr, Q in zip(arr_vec, Q_vec):
        prod *= _inner_integral(x, arr, Q)
    return prod

```

```

def E_A_j_second(arr_vec, Q_vec):
    """
    Second moment of A_j via eq. (4):
    """
    arr_vec = np.asarray(arr_vec, dtype=float)
    Q_vec = np.asarray(Q_vec, dtype=float)

    # Reuse function
    E_A_j = E_A_j_mean(arr_vec, Q_vec)

    #  $\rho_k / Q_k$ 
    prod_inv_T = np.prod(arr_vec / Q_vec)

    #  $\int_0^\infty \rho_k I_k(x) dx$ 
    outer, _ = quad(
        _integrand, 0, np.inf,
        args=(arr_vec, Q_vec),
        limit=200
    )

    return 2 * E_A_j * prod_inv_T * outer

def E_D_j_mean(arr_vec, Q_vec):
    """
    First moment of D_j (warehouse aggregate demand size) from eq. (9)
     $E[D_k] = Q_k$ ,  $E[T_k] = Q_k / \rho_k$ 
    """
    arr_vec = np.asarray(arr_vec, dtype=float)
    Q_vec = np.asarray(Q_vec, dtype=float)

    E_A_j = E_A_j_mean(arr_vec, Q_vec)
    return E_A_j * np.sum(arr_vec)

def L_hat_tilde_moments(L_0):
    """
    Calculates step 1 as in eq. 31 and 32
    """
    E_L_hat = L_0 / 2
    E_L_hat2 = L_0**2 / 3
    E_L_tilde = L_0 / 3
    E_L_tilde2 = L_0**2 / 6

    return (E_L_hat, E_L_hat2), (E_L_tilde, E_L_tilde2)

def fit_mixed_erlang(mean, second_moment):
    """
    Fit a mixed-Erlang distribution to the first two moments.
    Section 2.2 of Kiesmuller DeKok
    """
    var = second_moment - mean ** 2
    c2 = var / mean ** 2

    if c2 < 1:

```

```

    # Case 1: mixture of two Erlangs with same rate
    k1 = int(1 / c2)
    k2 = k1 + 1

    p1 = (1 / (1 + c2)) * (k2 * c2 - np.sqrt(k2 * (1 + c2) - k2 ** 2 * c2))
    p2 = 1 - p1

    mu1 = (k2 - p1) / mean
    mu2 = mu1

else:
    # Case 2: mixture of two exponentials
    k1 = 1
    k2 = 1

    mu1 = (2 / mean) * (1 + np.sqrt((c2 - 1/2) / (c2 + 1)))
    mu2 = 4 / mean - mu1

    p1 = mu1 * (mu2 * mean - 1) / (mu2 - mu1)
    p2 = 1 - p1

return k1, k2, mu1, mu2, p1, p2

def mixed_erlang_third_moment(k1, k2, mu1, mu2, p1, p2):
    """
    Third moment of a mixed-Erlang distribution.
    Used to get  $E[A_j^3]$ 
    """
    E3 = p1 * k1 * (k1 + 1) * (k1 + 2) / mu1 ** 3 \
        + p2 * k2 * (k2 + 1) * (k2 + 2) / mu2 ** 3
    return E3

def N_moments(E_L, E_L2, E_Aj, E_Aj2, E_Aj3):
    """
    Works for both  $L_{\hat{}}$  and  $L_{\tilde{}}$ .
    E_L : first moment of L
    E_L2 : second moment of L
    E_Aj : first moment of Aj
    E_Aj2 : second moment of Aj
    E_Aj3 : third moment of Aj

    E_N : first moment of  $N(L_j)$ 
    E_N2 : second moment of  $N(L_j)$ 
    """
    #First moment eq.33
    E_N = E_L/E_Aj + E_Aj2/(2*E_Aj**2) - 1
    #Second moment eq. 34
    E_N2 = E_L2/(E_Aj**2) + E_L*(2*E_Aj2/(E_Aj**3) - 3/E_Aj)\
        + (3*E_Aj2**2)/(2*E_Aj**4) - (2*E_Aj3)/(3*E_Aj**3) - (3*E_Aj2)/(2*E_Aj**2) + 1

    return E_N, E_N2

def D_L_moments(E_N, E_N2, E_Dj, E_Dj2):
    """

```

```

Step 3: Moments of  $D_j(L)$  / total warehouse demand during
Works for both  $D_j(L_{\text{hat}})$  and  $D_j(L_{\text{tilde}})$  . )

E_N : first moment of N
E_N2 : second moment of N
E_Dj : first moment of Dj
E_Dj2 : second moment of Dj

E_DL :  $E[D_j(L)]$ 
E_DL2 :  $E[D_j(L)^2]$ 
"""
var_Dj = E_Dj2 - E_Dj**2

# First moment eq. (35)
E_DL = E_N * E_Dj

# Second moment eq. (36)
E_DL2 = E_N * var_Dj + E_N2 * E_Dj**2

return E_DL, E_DL2

def E_D_j_second(arr_vec, Q_vec):
    arr_vec = np.asarray(arr_vec, dtype=float)
    Q_vec = np.asarray(Q_vec, dtype=float)

    E_A_j = E_A_j_mean(arr_vec, Q_vec)
    return E_A_j * np.sum(arr_vec * Q_vec)

def fit_mixed_erlang_DL_Ok(E_DL, E_DL2, Q_k):
    """
    Step 4: Fit mixed-Erlang to  $D_j(L) + O_k$  via Section 2.2.
    Works for both  $L_{\text{hat}}$  and  $L_{\text{tilde}}$ .
    """

    mean = E_DL + Q_k
    second = E_DL2 + 2*E_DL*Q_k + Q_k**2

    return fit_mixed_erlang(mean, second)

def mixed_erlang_loss(z, k1, k2, mu1, mu2, p1, p2):
    """
    Step 5: Loss function  $E[(X - z)^+]$  for  $X$  distributed as mixed-Erlang
     $z$  is either  $s_j$  or  $(s_j + Q_k)$ 
    """

    def component_loss(k, mu, p):
        lam = mu * z
        n1 = np.arange(k + 1)
        n2 = np.arange(k)
        fac1 = factorial(n1)
        fac2 = factorial(n2)
        cdf_k = np.sum(lam ** n1 * np.exp(-lam) / fac1)
        cdf_km1 = np.sum(lam ** n2 * np.exp(-lam) / fac2)
        return p * (k / mu * cdf_k - z * cdf_km1)

```

```

return component_loss(k1, mu1, p1) + component_loss(k2, mu2, p2)

def waiting_time(L_0, arr_vec, Q_vec, s_j, Q_j):
    """
    Compute first two moments of stockout waiting time  $W_k$  for each retailer  $k$ .
    See Proposition 2 and 3 (eqs. 27, 29) (step 6)
    Puts it all together step 1-6.

    Parameters
    L_0      : deterministic supplier-warehouse lead time
    arr_vec  : array Poisson arrival rates at each retailer
    Q_vec    : array for batch sizes at each retailer
    R_j      : warehouse reorder level
    Q_j      : warehouse batch size

    Returns
    E_W      : array  $E[W_k]$  for each retailer
    E_W2     : array  $E[W_k^2]$  for each retailer
    """
    arr_vec = np.asarray(arr_vec, dtype=float)
    Q_vec = np.asarray(Q_vec, dtype=float)

    E_Aj = E_A_j_mean(arr_vec, Q_vec)
    E_Aj2 = E_A_j_second(arr_vec, Q_vec)
    E_Aj3 = mixed_erlang_third_moment(*fit_mixed_erlang(E_Aj, E_Aj2))
    E_Dj = E_D_j_mean(arr_vec, Q_vec)
    E_Dj2 = E_D_j_second(arr_vec, Q_vec)

    #Step 1:
    (E_L_hat, E_L_hat2), (E_L_tilde, E_L_tilde2) = L_hat_tilde_moments(L_0)

    #Step 2:
    E_N_hat, E_N2_hat = N_moments(E_L_hat, E_L_hat2, E_Aj, E_Aj2, E_Aj3)
    E_N_tilde, E_N2_tilde = N_moments(E_L_tilde, E_L_tilde2, E_Aj, E_Aj2, E_Aj3)

    #Step 3:
    E_DL_hat, E_DL2_hat = D_L_moments(E_N_hat, E_N2_hat, E_Dj, E_Dj2)
    E_DL_tilde, E_DL2_tilde = D_L_moments(E_N_tilde, E_N2_tilde, E_Dj, E_Dj2)

    E_W = np.zeros(len(Q_vec))
    E_W2 = np.zeros(len(Q_vec))

    for k, Q_k in enumerate(Q_vec):
        # Step 4:
        params_hat = fit_mixed_erlang_DL_0k(E_DL_hat, E_DL2_hat, Q_k)
        params_tilde = fit_mixed_erlang_DL_0k(E_DL_tilde, E_DL2_tilde, Q_k)

        # Step 5:
        loss_hat_sj = mixed_erlang_loss(s_j, *params_hat)
        loss_hat_sj_Qj = mixed_erlang_loss(s_j + Q_j, *params_hat)

        loss_tilde_sj = mixed_erlang_loss(s_j, *params_tilde)
        loss_tilde_sj_Qj = mixed_erlang_loss(s_j + Q_j, *params_tilde)

```

```

# Step 6:
E_W[k] = (L_0 / Q_j) * (loss_hat_sj - loss_hat_sj_Qj)
E_W2[k] = (L_0 ** 2 / Q_j) * (loss_tilde_sj - loss_tilde_sj_Qj)

var_w = [EW2 - EW**2 for EW2,EW in zip(E_W2, E_W)]

return E_W, var_w

```

B ExtendSim Model Description

The functional details of the ExtendSim Model will be included here. We assume basic to intermediate understanding of ExtendSim in these descriptions. Note that the software often allows multiple different ways to achieve the same function or result. The model uses hierarchical blocks to handle and separate the complex system functions. We will start first with the central warehouse and continue to customer demand. Figure 3 is provided again for reference.

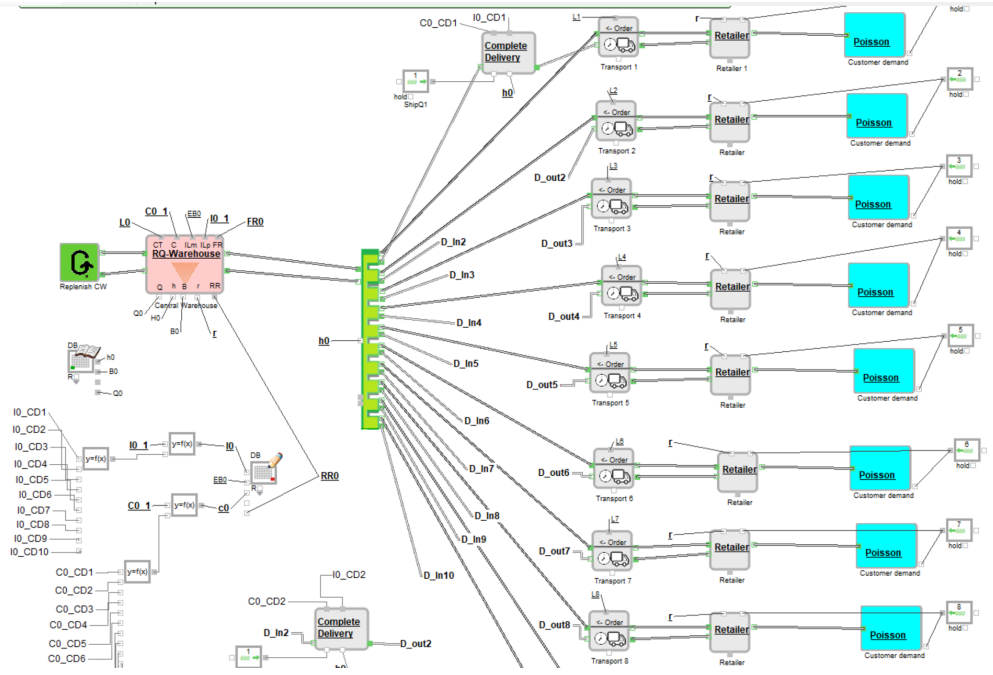


Figure 2: ExtendSim simulation model

Central Warehouse Replenishment Block

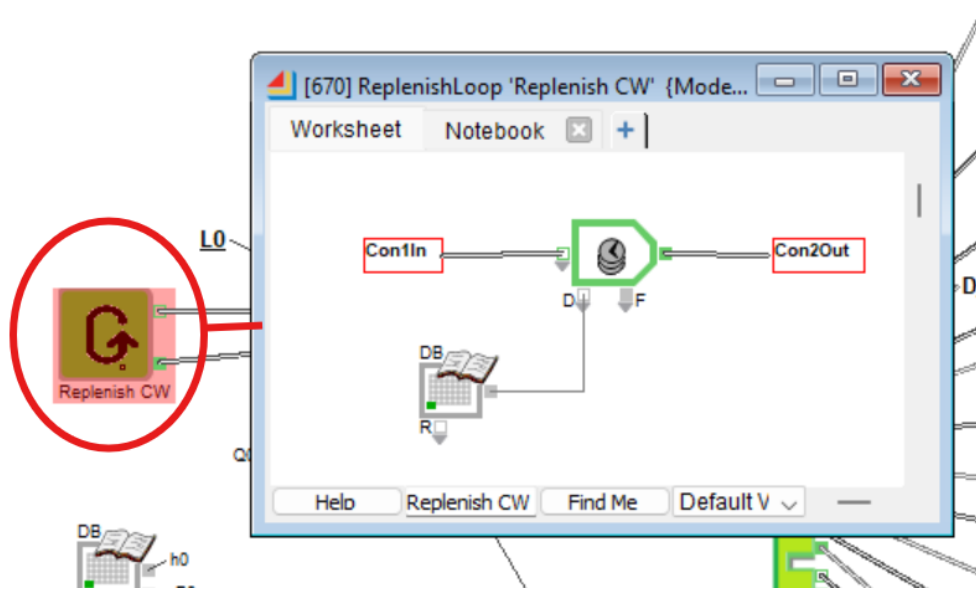


Figure 4: Central Warehouse Replenishment in the model

The replenishment order for the central warehouse is received in Con1In passes through a delay block and then into Con2Out. The delay block here applies the L_0 delay from the data base into the activity block, which functions as a delay block.

Central Warehouse Block

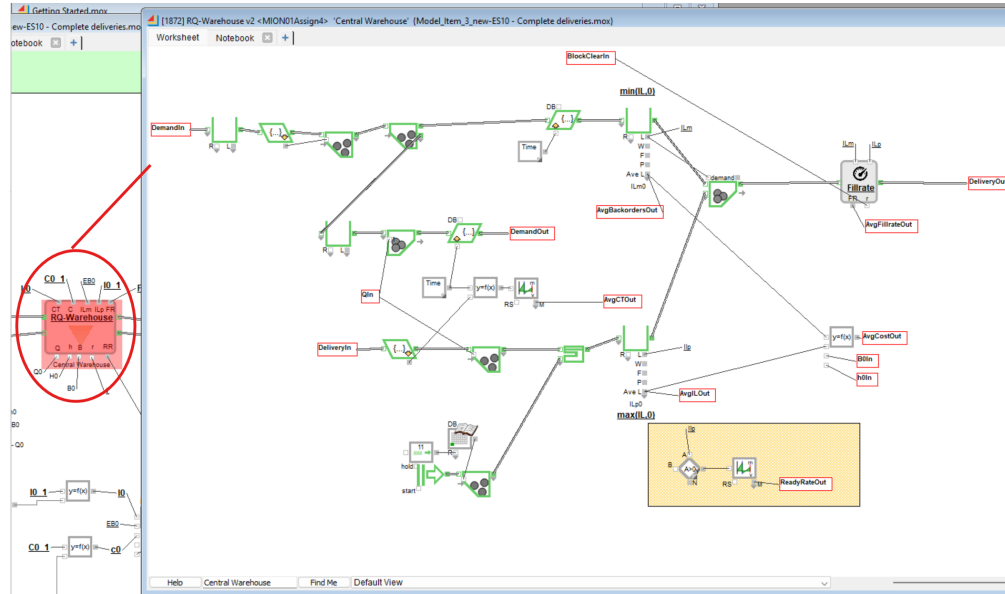


Figure 5: Central Warehouse block in the model

There are two flows of items. The DemandIn flow handles the demand (requested order) submitted from a retailer while the DeliveryIn handles both the replenishment of the central warehouse and the on hand inventory within the warehouse. Each flow is matched 1 to 1 into a batch block to ensure there is sufficient inventory to meet the demand. In the case of there not being enough inventory, a backorder occurs. The fill

rate is calculated from this and is covered separately in the next section.

The demand flow in the central warehouse follows from a queue block immediately into a Get block. The Get block puts the order quantity which is needed for it to be unbatched into individual items. The following batch block copies the demand so it used to for the replenishment cycle. The replenishment portion takes the Q_0 into a batch block so it can set the required amount needed for replenishment. This also includes the calculation for the average cycle time out. The final portion is the set block which sets the current time an item passes through and then the queue block which is used to handle backorders (average, IL_m , and used for the average cost).

The delivery in flow first gets (from a Get block) the set cycle time from the previously set cycle time from the replenishment portion. A delay has been applied to this in the central warehouse replenishment block. Note that the averageCTout will only be constant (L_0) unless a stochastic delay is applied. The unbatch block is used to convert the order into Q_0 units and then deposit it into the queue inventory block or on hand inventory. This block is uses the average IL for cost calculations within the model. Note that the create and batch block connected to the select item in block represents $R_0 + Q_0$.

The final component comes from the the orange/yellow box. This is the average ready rate, which is a simple If/then calculation. If the inventory level is greater than 0, output 1. If not, output 0. The ready rate is the average over the entire simulation and number of runs. It should between 0 and 1.

Fill rate Block

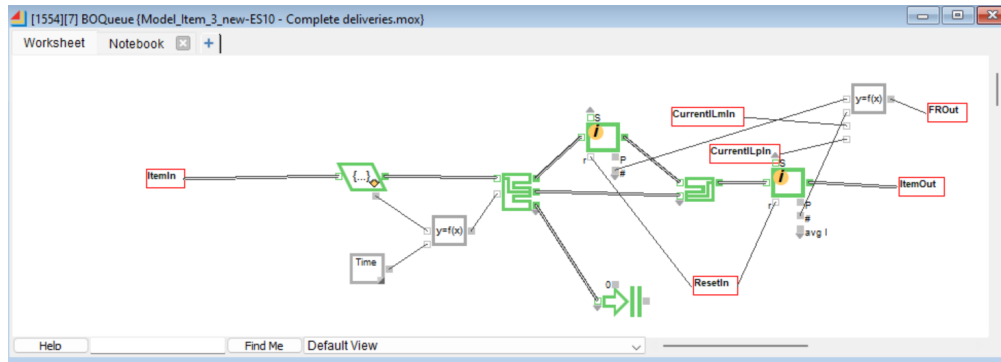


Figure 6: Fill rate block in the model

The fill rate block in Figure 6 functions the same across the central warehouse and its retailers. The Get block gets the arrival time set previously and immediately compares it to the current time in the gray equation block. If both times do not equal each other, it is a backordered item. This is important for when it arrives at the select item out block. If the item is backordered, it goes to the top output to go into a information block. If it was not backordered, it goes through normally. After merging, another information block is used to count the total number of items. These counts, along with the inventory levels from the input (central warehouse or retailer), are used to calculate the fill rate. All items exit the block after.

Demand Delivery System Block

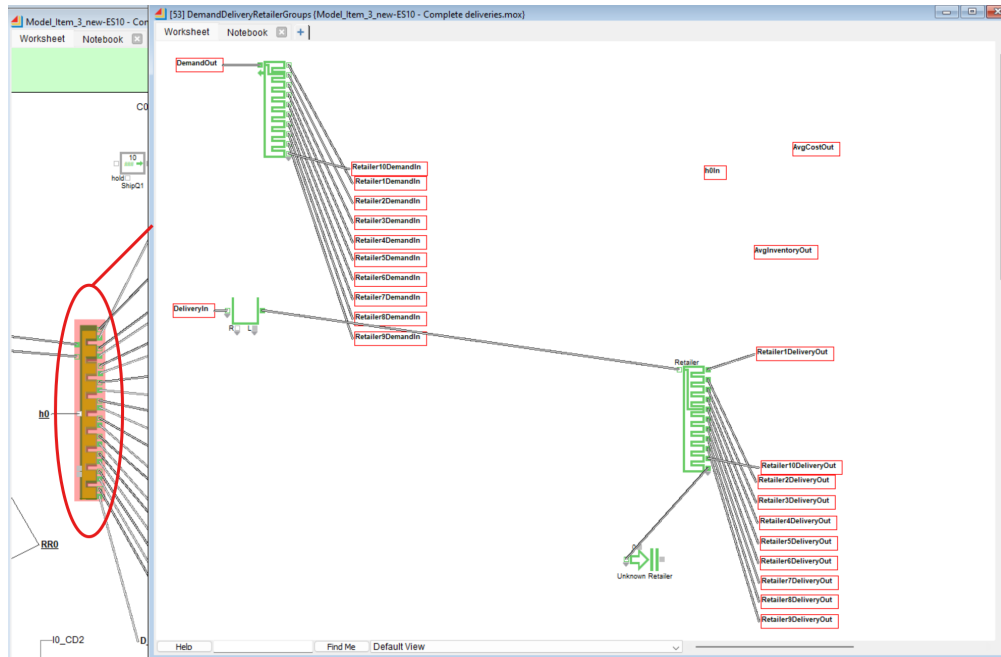


Figure 7: Demand Delivery System block in the model

Figure 7 handles both the demand and delivery flow. The demands from the retailers are simply merged into a select item in block and sent to the warehouse.

The delivery section filters by property in the select item out block to sort the deliveries to the correct retailer. Note that there is a fail safe exit block in the case of an unknown retailer.

Delivery Type Block

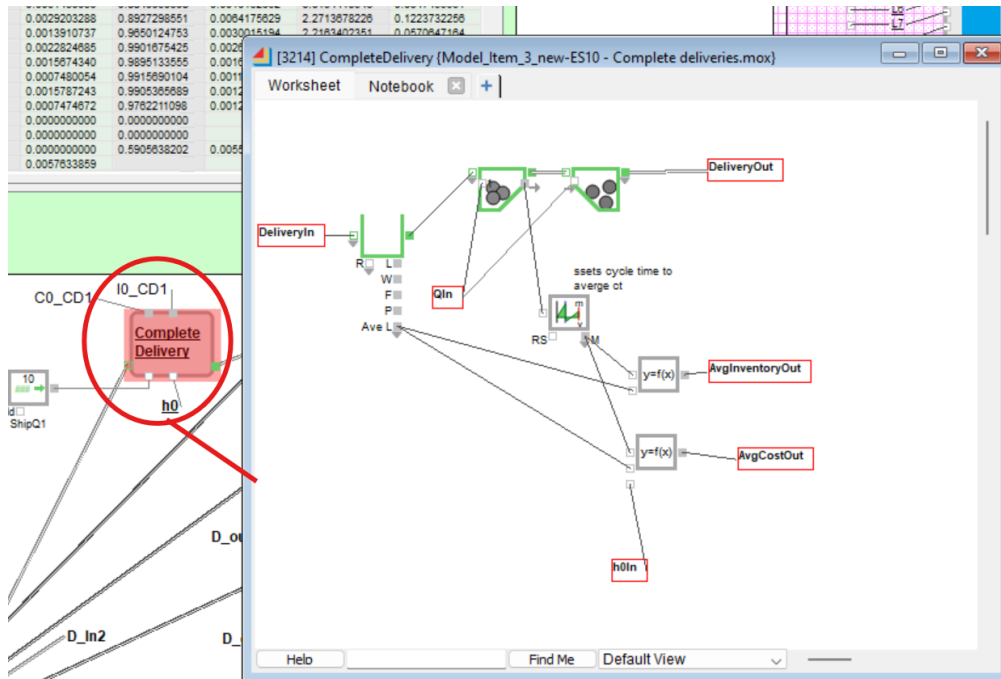
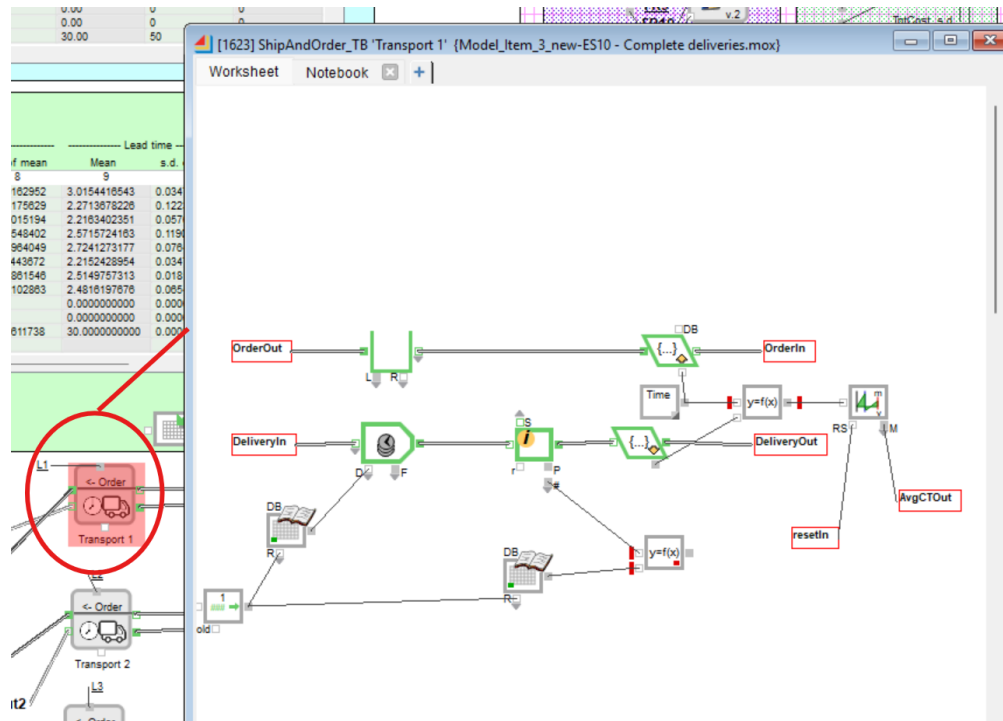


Figure 8: Delivery Type block in the model

Figure 8 operates as a loading dock. When deliveries come in, they are directed to the queue and then into the batch block. The batch block takes the batch quantity into account and waits until enough items arrive before moving on. This is manually set by a constant outside the block. If it is set to 1, the block functions as partial deliveries. Greater values change it to complete order deliveries. The unbatch block after just unbatches it as the model was built to use single units, not batched units. This block also adds calculations to the average cost and inventory out as there is the possibility for inventory to be located there.



The transportation block in Figure 9 has two components: demand and delivery. The demand that passes through on the top just sets (with a set block) the current time that it passes through.

Retailer Block

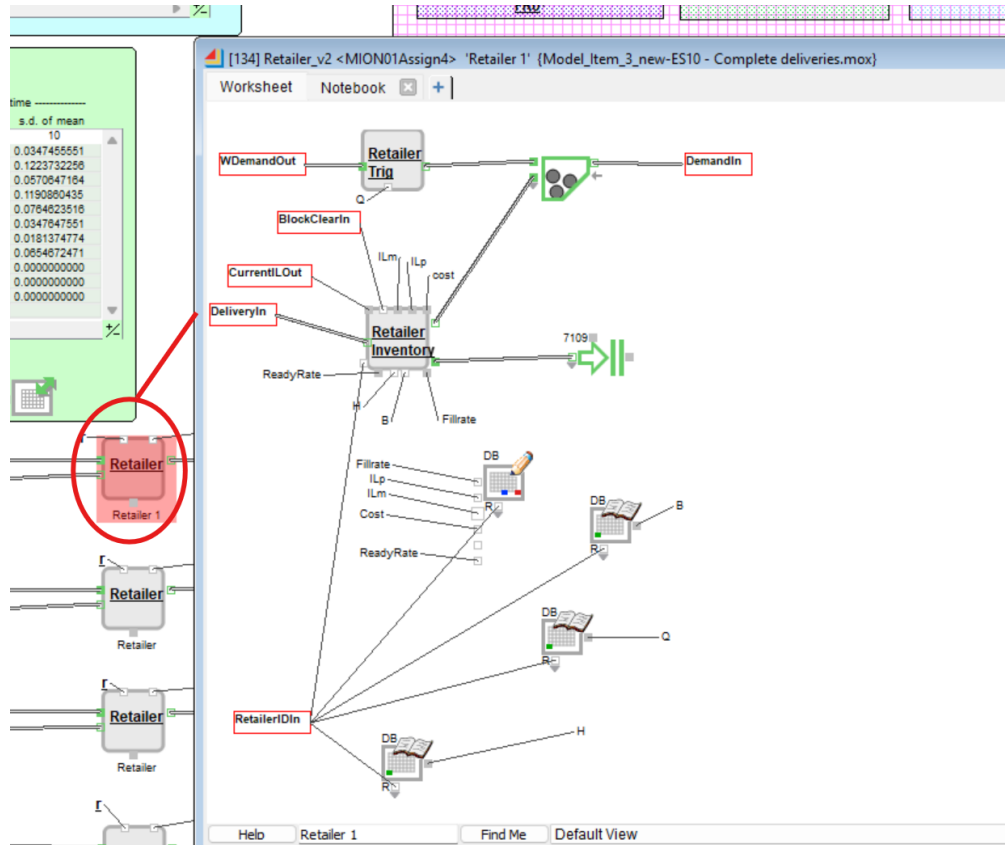


Figure 10: Retailer block in the model

The retailer block in Figure 10 mainly comprises of its sub blocks: retailer trigger and inventory. The exit block removes items from the simulation model while unbatch block adds the demand quantity to the inventory block. There are three database read blocks that pull database values to be used in for the varies inputs. There is a database write block that writes to the database such as fill rate or cost.

Retailer Inventory Block

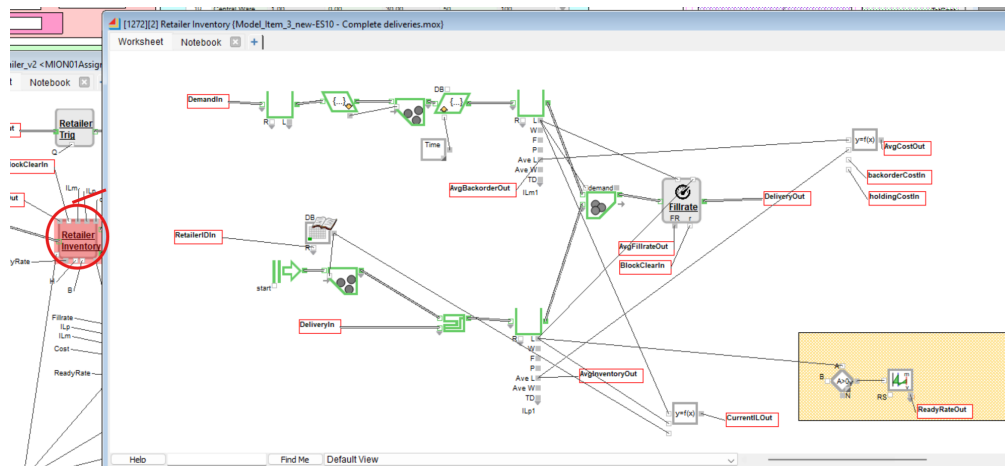


Figure 11: Retailer Inventory block in the model

average cost and inventory on hand in the model was added here. The complete delivery block can be completely included inside the central warehouse block, for simplicity and to save time, it is included here along with the calculations. It provides a summation of all values and writes it to the database. This is done through a series of equation blocks.

Statistics Section

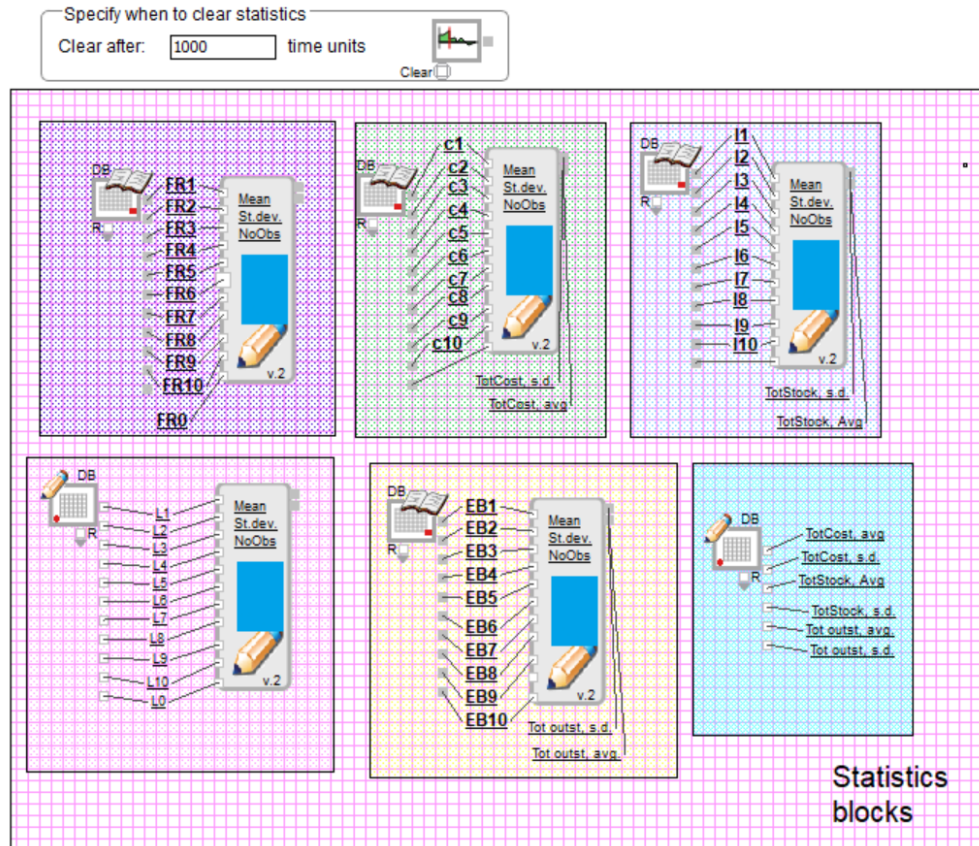


Figure 15: Statistics Section in the model

The statistics section in Figure 15 calculates all the retailer average values. These are the fill rate, cost, stock on hand, lead time, and outstanding units. Each of these uses a series of mean and variance blocks to obtain the averages gathered from the model and write it to the database. This part also includes a clear statistics section to factor in a warm up period.

C Complete Result Tables and Result Calculations

Table 13: Simulation Complete Order Fulfillment Results. 7000 time units and 8 runs.

	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6		Test 7		Test 8	
	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.
R1	6.741	0.136	3.166	0.108	8.267	0.215	4.263	0.168	9.049	0.253	3.979	0.280	12.175	0.491	6.256	0.406
R2	7.027	0.268	3.062	0.126	8.053	0.228	3.782	0.213	7.881	0.243	3.709	0.214	10.486	0.405	6.038	0.480
R3	6.907	0.137	2.651	0.124	8.284	0.165	4.155	0.279	8.205	0.189	4.047	0.260	10.965	0.782	5.870	0.354
R4	7.195	0.246	3.073	0.097	8.304	0.268	4.269	0.171	10.005	0.345	5.253	0.169	12.224	0.549	6.322	0.461
R5	6.892	0.158	2.904	0.081	8.612	0.168	4.199	0.258	9.486	0.465	4.204	0.257	11.627	0.429	6.348	0.383
R6	6.919	0.240	2.719	0.117	8.190	0.073	4.047	0.129	8.764	0.292	4.240	0.337	10.941	0.483	5.374	0.306
R7	7.386	0.270	3.088	0.161	8.007	0.343	4.061	0.171	9.759	0.390	4.935	0.324	12.024	0.332	5.964	0.299
R8	6.829	0.148	2.833	0.151	7.885	0.202	3.646	0.234	8.690	0.364	4.547	0.298	10.229	0.352	5.107	0.243
Average	6.987	0.200	2.937	0.121	8.200	0.208	4.053	0.203	8.980	0.318	4.364	0.267	11.334	0.478	5.910	0.367
CW FR	36.24%		64.86%		37.67%		61.06%		40.40%		65.66%		40.49%		61.57%	

Table 14: Simulation Partial Order Fulfillment. 7000 time units and 8 runs.

	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6		Test 7		Test 8	
	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.	Delay	Var.
R1	5.038	0.137	2.307	0.081	5.781	0.307	2.752	0.119	7.311	0.213	2.984	0.121	8.450	0.247	4.347	0.209
R2	5.511	0.101	2.538	0.121	6.175	0.223	2.928	0.140	6.991	0.343	3.317	0.172	8.025	0.361	4.462	0.408
R3	5.304	0.171	2.295	0.128	6.312	0.233	2.922	0.145	7.860	0.317	3.060	0.246	7.712	0.363	4.100	0.218
R4	5.437	0.157	2.352	0.101	6.335	0.122	3.061	0.197	8.314	0.291	3.572	0.112	8.592	0.386	4.181	0.213
R5	5.394	0.207	2.351	0.108	5.809	0.209	2.672	0.110	7.072	0.239	3.232	0.184	7.235	0.351	4.123	0.302
R6	5.648	0.230	2.571	0.094	6.663	0.295	2.890	0.172	7.382	0.359	2.903	0.219	7.748	0.303	4.229	0.253
R7	5.623	0.227	2.086	0.087	6.285	0.179	2.586	0.214	7.020	0.148	3.091	0.215	8.282	0.185	4.625	0.233
R8	5.604	0.119	2.355	0.083	6.127	0.177	3.026	0.185	7.321	0.351	3.389	0.208	8.835	0.404	3.911	0.290
Average	5.445	0.169	2.357	0.100	6.186	0.218	2.855	0.160	7.409	0.283	3.193	0.185	8.110	0.325	4.247	0.266
CW FR	36.95%		62.60%		38.26%		62.74%		39.93%		64.98%		41.62%		61.22%	

Table 15: Full Comparison of the Simulation model and Approximation Methods

Rep.	SIM	A & M	Ax	K & DK	B & F
Test 1					
R1	6.741488079	6.324207589	5.525807	8.487900	7.187300
R2	7.027164441	5.392694163	5.525807	7.524400	6.674500
R3	6.907057954	5.491324997	5.525807	7.524400	6.621900
R4	7.195023690	5.898372880	5.525807	8.487900	7.220400
R5	6.891508392	5.980264170	5.525807	8.487900	7.193100
R6	6.919154968	5.422110377	5.525807	7.524400	6.655200
R7	7.386488801	5.917729003	5.525807	8.487900	7.212700
R8	6.829405311	5.454795058	5.525807	7.524400	6.637500
MAE	N/A	1.251974175	1.461354455	1.018988546	0.255031414
MSE	N/A	1.306212868	1.474540495	1.115837696	0.281233757
Test 2					
R1	3.166433500	2.616472764	2.223514	4.617400	3.013500
R2	3.061600914	2.206863110	2.223514	4.058400	2.746000
R3	2.651280690	2.250233544	2.223514	4.058400	2.721200
R4	3.072633732	2.429222637	2.223514	4.617400	2.990800
R5	2.904219191	2.465232277	2.223514	4.617400	2.983500
R6	2.718960140	2.219798152	2.223514	4.058400	2.736300
R7	3.088319338	2.437734006	2.223514	4.617400	2.988300
R8	2.832609867	2.234170420	2.223514	4.058400	2.727900
MAE	N/A	0.579541307	0.713493171	1.400892829	0.115204666
MSE	N/A	0.594885852	0.735118017	1.415756711	0.142360547

Continued on next page

Table 15 – Continued from previous page

Rep.	SIM	A & M	Ax	K & DK	B & F
Test 3					
R1	8.267057168	7.369557179	6.191531	10.742300	8.559100
R2	8.052872747	6.074895272	6.191531	8.546400	7.583100
R3	8.283824033	6.189130146	6.191531	8.546400	7.463500
R4	8.303623112	6.583512450	6.191531	9.552500	8.096100
R5	8.611960807	6.672431537	6.191531	9.552500	8.025400
R6	8.190499477	6.108965322	6.191531	8.546400	7.543100
R7	8.006813195	6.604529689	6.191531	9.552500	8.078000
R8	7.884499570	6.146820934	6.191531	8.546400	7.503000
MAE	N/A	1.731475464	1.796991692	0.997969220	0.434600689
MSE	N/A	1.772943635	1.928750569	1.216394177	0.491952836
Test 4					
R1	4.262650550	3.475772587	2.863104	6.237400	3.938600
R2	3.782311426	2.839819398	2.863104	4.862400	3.528700
R3	4.155410661	2.895932914	2.863104	4.862400	3.450900
R4	4.269251461	3.089658151	2.863104	5.480300	3.746100
R5	4.198729763	3.133336254	2.863104	5.480300	3.705300
R6	4.047256480	2.856555008	2.863104	4.862400	3.501900
R7	4.061064691	3.099982066	2.863104	5.480300	3.735200
R8	3.646293871	2.875150130	2.863104	4.862400	3.475800
MAE	N/A	1.019595299	1.189767113	1.213116387	0.417558612
MSE	N/A	1.034188603	1.208593681	1.266526915	0.449373566
Test 5					
R1	9.049424292	8.267906390	7.361032	12.109300	9.806000
R2	7.880839502	7.023312610	7.361032	10.344000	9.048600
R3	8.204901389	7.155093128	7.361032	10.344000	8.910600
R4	10.00546454	7.698949233	7.361032	12.109300	10.240100
R5	9.485644065	7.808364071	7.361032	12.109300	10.138900
R6	8.763935656	7.062615571	7.361032	10.344000	9.004900
R7	9.759362402	7.724810922	7.361032	12.109300	10.215500
R8	8.689675897	7.106285529	7.361032	10.344000	8.958800
MAE	N/A	1.498988786	1.618873968	2.246744032	0.560519032
MSE	N/A	1.585974300	1.761583300	2.293224647	0.637811897
Test 6					
R1	3.979165828	3.810581038	3.301229	6.433400	4.244100
R2	3.709335309	3.194307621	3.301229	5.369100	3.967200
R3	4.047361314	3.259560100	3.301229	5.369100	3.876900
R4	5.252773496	3.528856047	3.301229	6.433400	4.514800
R5	4.204022524	3.583033930	3.301229	6.433400	4.445800
R6	4.239733522	3.213768886	3.301229	5.369100	3.938000
R7	4.934749377	3.541661728	3.301229	6.433400	4.497800
R8	4.547340112	3.235392515	3.301229	5.369100	3.907900
MAE	N/A	0.943414952	1.063081185	1.536939815	0.381391770
MSE	N/A	1.059854143	1.167711280	1.623491295	0.427149583
Test 7					
R1	12.17521948	9.889771800	8.364809	15.530100	13.375300
R2	10.48560119	8.044268594	8.364809	11.810800	10.551000
R3	10.96537406	8.207107112	8.364809	11.810800	10.370600
R4	12.22356380	8.769287711	8.364809	13.556000	11.928000
R5	11.62666645	8.896039304	8.364809	13.556000	11.806800
R6	10.94129938	8.092834468	8.364809	11.810800	10.495800
R7	12.02438677	8.799247178	8.364809	13.556000	11.899700
R8	10.22930679	8.146796550	8.364809	11.810800	10.435900
MAE	N/A	2.728258150	2.969118240	1.596235260	0.388978762
MSE	N/A	2.762061889	3.058120193	1.761333673	0.521320263

Continued on next page

Table 15 – Continued from previous page

Rep.	SIM	A & M	Ax	K & DK	B & F
Test 8					
R1	6.255831170	5.107358544	4.225777	9.160400	6.360300
R2	6.037919242	4.135912283	4.225777	6.595900	5.217600
R3	5.869657083	4.221628129	4.225777	6.595900	5.082800
R4	6.322082208	4.517551886	4.225777	7.740900	5.827800
R5	6.347901368	4.584272096	4.225777	7.740900	5.732500
R6	5.373504794	4.161476658	4.225777	6.595900	5.175900
R7	5.964296643	4.533322117	4.225777	7.740900	5.804800
R8	5.107253836	4.189881520	4.225777	6.595900	5.131000
MAE	N/A	1.478380389	1.684028793	1.436031707	0.400272041
MSE	N/A	1.515652064	1.736446068	1.584906047	0.498881991

Pooled Variance and Standard Error Calculations

To determine whether a particular main effect was significant or not, we used the pooled variance for calculating the Standard error. This method was used from (Montgomery 2017, p. 237).

$$\sigma_p^2 = \frac{\sum_{i=1}^N \sigma_i^2}{N} \quad (99)$$

$$V(Effect) = \sqrt{\frac{4 \cdot \sigma_p^2}{N \times n}} \quad (100)$$

Note that the pooled variance is simplified due to having the same sample sizes. The $V(Effect)$ is also known as the standard error.