JOINT OPTIMISATION OF FACILITY LOCATION AND TWO-ECHELON INVENTORY CONTROL WITH RESPONSE TIME REQUIREMENT AND LATERAL TRANSSHIPMENT

SAMUEL CHIABOM, ZELIBE

DECEMBER, 2019

JOINT OPTIMISATION OF FACILITY LOCATION AND TWO-ECHELON INVENTORY CONTROL WITH RESPONSE TIME REQUIREMENT AND LATERAL TRANSSHIPMENT

BY

SAMUEL CHIABOM, ZELIBE
Matric. No. 152811;
B.Sc. Mathematics (Abraka), M.Sc. Mathematics (Ibadan)

A Thesis in the Department of Mathematics
Submitted to the Faculty of Science
in partial fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

of the

UNIVERSITY OF IBADAN

DECEMBER, 2019

Abstract

Lateral Transshipment (LT) (stock movement between facilities on the same echelon), has been used as an option for reducing the occurrences of stockout and excess stock in many multi-echelon environments. Several LT models have been formulated for many supply chain systems. However, the incorporation of LT into a system which jointly optimises facility location and two-echelon inventory decisions with Response Time Requirement (RTR) has not been considered. Therefore, this study was designed to incorporate LT into a two-echelon system which jointly minimises expected cost emanating from facility location and inventory decisions subject to RTR.

The customer arrival at facilities was modelled as a single server queue with Poisson arrivals and exponential service rate. The balance equation of this queue along with the distribution of the number of orders in replenishment (N_{vw}) was used to derive service center steady state expected level for on-hand inventory (I_{vw}), backorder (B_{vw}), and LT (T_{vw}). The derived steady state expected levels were used to formulate the two-echelon LT model. This model was decomposed using Lagrange relaxation. Relaxation of the assignment variable's integrality was used to further reduce the model. The reduced model was checked for convexity using second order conditions. Karush-Kuhn-Tucker (KKT) conditions were used to investigate global optimality, which was also examined for the case of stochastic occurrences. Multiple computational experiments were performed on three data sets using general algebraic modelling system for the values: $d_{uvw_{(max)}} = 100, 150$; $\rho = 0.5, 0.9$ and $\tau = 0.2, 0.3, 0.5$, where, $d_{uvw_{(max)}}$, ρ and τ are customer distance, utilisation rate and RTR, respectively.

The expected number of customers in queue at a service center was: $E[N_{vw}] = \frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_0}\frac{\rho^{S_0+1}}{1-\rho} + \sum_{u\in U}\lambda_u Y_{uvw}\alpha_w$. The derived steady state expected levels were: $I_{vw} = \sum_{s=0}^{S_{vw}-1}(S_{vw} - s)P\{N_{vw} = s\}$, $B_{vw} = \frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_0}\frac{\rho^{S_0+1}}{1-\rho} + \sum_{u\in U}\lambda_u Y_{uvw}\alpha_w + \frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}\left(\sum_{s=0}^{|w|S_{vw}-1} F_w(s) - |w|S_{vw}\right)$ and
$T_{vw} = \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}\left(\sum_{s=0}^{|w|S_{vw}-1} F_w(s) - |w|S_{vw}\right)$

The two-echelon LT model formulated was:

$$\min \sum_{w\in W}\sum_{v\in V}\left(f_{vw}X_{vw} + h_{vw}I_{vw} + p_{vw}B_{vw} + q_{vw}T_{vw} + \sum_{u\in U}\lambda_u Y_{uvw}d_{uvw}\right) + h_0 S_0$$

Subject to

$$\sum\nolimits_{v\in V} Y_{uvw} = 1$$
$$Y_{uvw} \le a_{uvw}X_{vw}$$
$$S_{vw} \le C_{vw}$$
$$S_0 \le C_0$$
$$\left[\tfrac{\rho^{S_0+1}}{\lambda_0(1-\rho)} + \alpha_w - \tau\right] \le \tfrac{\sum_{s=0}^{|w|S_{vw}-1}[1-F_w(s)]}{\lambda_w}$$
$$X_{vw}, Y_{uvw} \in \{0,1\}.$$

The Lagrange dual problem was:

$$\max_{\theta,\pi\ge 0}\min_{X,Y,S}\sum_{w\in W}\sum_{v\in V}\Bigg\{f_{vw}X_{vw} + (h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s) - q_{vw}S_{vw}$$
$$+ (p_{vw} - q_{vw} + \theta_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}$$
$$+ (p_{vw} - q_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}(\sum_{s=0}^{|w|S_{vw}-1}F_w(s) - |w|S_{vw}) + \sum_{u\in U}\lambda_u Y_{uvw}\frac{(p_{vw}+\theta_{vw})\rho^{S_0+1}}{\lambda_0(1-\rho)}$$
$$+ \sum_{u\in U}(((p_{vw}+\theta_{vw})\alpha_w + d_{uvw} - \theta_{vw}\tau)\lambda_u - \pi_u)\,Y_{uvw}\Bigg\} + \sum_{u\in U}\pi_u$$

Subject to

$$Y_{uvw} \le a_{uvw}X_{vw}$$
$$S_{vw} \le C_{vw}$$
$$S_0 \le C_0$$
$$X_{vw}, Y_{uvw} \in \{0,1\}$$

The reduced model obtained was:

$$\min_{0\le Y_{uvw}} (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw}S_{vw}$$
$$+ (p_{vw} - q_{vw} + \theta_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}$$
$$+ (p_{vw} - q_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}(\sum_{s=0}^{|w|S_{vw}-1} F_w(s) - |w|S_{vw}) + \sum_{u\in U}\lambda_u Y_{uvw}\frac{(p_{vw}+\theta_{vw})\rho^{S_0+1}}{\lambda_0(1-\rho)}$$
$$+ \sum_{u\in U}(((p_{vw}+\theta_{vw})\alpha_w + d_{uvw} - \theta_{vw}\tau)\lambda_u - \pi_u)\, Y_{uvw}$$

where $\lambda_u, \lambda_w, \lambda_0, Y_{uvw}, L_w, (S_{vw}, S_0), (C_{vw}, C_0), X_{vw}, a_{uvw}, \tau, f_{vw}, h_{vw}, p_{vw}, q_{vw}$ and d_{uvw} are, customer demand, pool demand, plant demand, assignment variable, lead time, base-stock levels, capacity, location variable, distance variable, facility, holding, backorder, LT and transportation costs, while, θ_{vw}, π_u are Lagrange multipliers and F_{vw}, F_w are facility and pool distribution functions, respectively. The reduced model was convex and satisfied KKT conditions, establishing the existence of global minimum for the two-echelon LT model. The stochastic case was also shown to be convex. The computational experiment showed that expected cost remained stable with increasing RTR, and that the model resulted to lower cost when compared with the model without LT.

The two-echelon joint location-inventory model with response time requirement and lateral transshipment obtained lower expected cost than the model without lateral transshipment. Stability of expected cost with varying response time requirement was also established.

Keywords: Supply chain, Convexity, Karush-Khun-Tucker conditions, Global optimality, Basestock level.

Word count: 473

Dedication

To the wife of my youth, Mrs Osivwi Avwerosuo Zelibe. Your prayers, love and support have been invaluable all the way.

Acknowledgements

Except the Lord builds a house, they labour in vain that build it..." Indeed, without the involvement of the Almighty God this work would not have been a success. *Great is thy faithfulness O Lord! ...*

I appreciate the effort of my supervisor Dr U. N. Bassey, whose guidance, encouragements, constructive criticisms and comments were extremely helpful towards the completion of this thesis.

I am grateful to late Dr C. R. Nwozo for the role he played in setting me up on the path of research in Operations Research. He was my supervisor for this Ph.D. until his untimely death. May his soul rest in peace. May the Almighty God help his family.

I would like to thank all staff of the Department of Mathematics, University of Ibadan, headed by Dr Deborah O. A. Ajayi. Many thanks to Prof. E. O. Ayoola, who has been a father to me since my M.Sc. days. Special thanks to Prof. O. O Ugbebor, who has been a cherished mentor to me. I am grateful to Dr P.O Arawomo for all that he is. I am grateful to Dr EniOluwafe for the extra length he goes as PG coordinator. I would love to mention the names of all staff of the Department but for space. I am eternally grateful to you all.

Many thanks to Dr D. Allenotor, Head of Department of Mathematics and Computer Science, Federal University of Petroleum Resources, Effurun, for the support given to me towards the completion of this programme. I am also grateful to all staff of the Department.

I appreciate my friend and brother Celestine Achudume, who has been with me through every step of my postgraduate studies.

Special thanks to my parents Prof. Samuel A. A. Zelibe and Mrs. C. O. Zelibe, who are my first academic and life mentors. I also want to appreciate my non-biological parents Mr. and Mrs. Ogidiama who happen to be the biological parents of my wife, thanks for helping us out with the kids when we had to be away from home.

I have saved the best for last. My wife (Osivwi Zelibe), our daughter (Chimnoya

Zelibe) and son (Praise Zelibe) are the most important people in my life, and I cannot thank them enough for their unconditional love and support. Thank you.

Certification

I certify that this work was carried out by **Mr. Samuel Chiabom ZELIBE** in the Department of Mathematics, University of Ibadan, Ibadan Nigeria under my supervision.

. . . .

Date

Supervisor

Dr U. N. Bassey *Ph.D. (Ibadan)*

Department of Mathematics, University of Ibadan, Nigeria.

Contents

List of Tables

List of Figures

List of Notations

The following notations were used for this study:

SVC stands for Service Center

SVCs stand for Service Centers

LT stands for Lateral Transshioment

U represents the set of Customers

V represents the set of candidate SVC Locations

W represents the set of Pools

f_{vw} is the fixed opening cost for SVC at location v in pool w

h_{vw} is the per unit holding cost of inventory at SVC v in pool w per unit time

p_{vw} is the per unit cost of backorder per unit inventory per unit time

q_{vw} is the LT cost per unit inventory

$\bar{d}_{uvw}$ is the distance of customer u from SVC v in pool w

d_{uvw} is the Transportation cost from SVC v in pool w to customer u

λ_u is customer u's demand rate

λ_{vw} is the SVC demand rate of SVC v in pool w demand rate $= \sum_{u \in U} \lambda_u Y_{uvw}$

λ_w is pool w's demand rate $= \sum_{v \in V} \lambda_{vw} = \sum_{v \in V} \sum_{u \in U} \lambda_{uvw} Y_{uvw}$

λ_0 is the plant demand rate $= \sum_{w \in W} \lambda_w$

μ is the plant's order processing rate

ρ is the plant's utilisation rate $(= \frac{\lambda_0}{\mu})$

τ is the response time requirement

α_w is the exact plant to pool w lead time

d_{max} is the maximum distance allowable from a customer to its assigned SVC

$a_{uvw} = 1$ if the distance from customer u to candidate location v in pool w is not greater than d_{max}, 0 otherwise

C_{vw} is the capacity of SVC v in pool w in number of units of the product, this is uniform for all SVCs in pool w

$C_w = \hat{w} C_{vw}$ is the total space available for storage at pool w, where $\hat{w}$ is the number of SVCs in pool w

C_0 is the total space available for storage space at the plant $X_{vw} = 1$ if a SVC is located at v in pool w, 0 otherwise

$Y_{uvw} = 1$ if customer u's demand is assigned to SVC v in pool w

S_{vw} is the stock level required at SVC v in pool w, this is uniform for all SVCs in pool w

$S_w = \hat{w} S_{vw}$ is the total stock level required at pool w

S_0 is the stock level required at the plant

I_{vw} is the expected SVC inventory level at SVC v in pool w in steady state

$I_w = \sum_{v \in V} I_{vw}$ is the expected pool w inventory level in steady state

B_{vw} is the expected SVC backorder level at SVC v in pool w in steady state

B_w is the expected pool w backorder level in steady state

T_{vw} is the expected LT level at SVC v in pool w in steady state

Wt_{vw} is the expected SVC response time at SVC v in pool w

I_0 is the expected plant inventory level in steady state

B_0 is the expected plant backorder in steady state

$N_{vw}(t)$ is the number of replenishment orders placed by SVC v in pool w which are yet to arrive at time t

N_{vw} is the steady state expected number of replenishment orders placed by SVC v in pool w which are yet to arrive

$N_w(t)$ is the number of replenishment orders placed by pool w which are yet to arrive at time t.

N_w is the steady state expected number of replenishment orders placed by pool w which are yet to arrive.

$N_0(t)$ is the number of replenishment orders placed by the plant that are yet to arrive by time t.

N_0 is the steady state expected number of replenishment orders placed by the plant that are yet to arrive by time.

Chapter 1
INTRODUCTION

1.1 Introduction

The day to day running of man's life is largely dependent on the usage of various machines or eqipments in various areas such as transportation, electricity, medicine, communication, finance, etc. A common trait of these machines is that their components are subject to failure. Hence, there is always a requirement from customers to replace failed parts within an acceptable time frame. To satisfy this requirement, manufacturers need to have efficient service parts supply systems that guarantee customers' desired response times. Items kept aside for purposes of meeting this customer requirement are called service parts. The relevance of service parts in various sectors and in individuals personal lives cannot be overemphasized. Inavailability or inadequacy of these parts will almost surely lead to a halt or slow down activities in businesses and in man's everyday life.

Generally, there are three classifications of decisions made in service parts supply systems; namely, tactical, strategic and operational decisions. Strategic decisions have to do with determining what customer requirements are and how to distribute resources to meet these requirements. Decisions that have to do with facility location planning are strategic decisions. Tactical decisions are made to ascertain inventory level necessary to meet given operational objectives at certain future time. Operational decisions drive the everyday running of the system. Distribution planning decisions are examples of operational decisions.

Decisions on inventory and location are critical for efficient running of service parts supply system. If available parts in the system are scarce, customers' waiting times (response times) will be prolonged. If surplus parts are available, the operating cost of the system will be very high. If available facilities are too few, customer satisfaction and service levels (response time requirements) may not be guaranteed, whereas, an increase in the total number of facilities will also result in increase in the total operating cost. This highlights the need for efficient design of replenishment systems and the optimal allocation of resources within these systems if the decision maker aims to attain cost-effective management of service parts. Consequently, decision makers are always on the look out for more efficient means of optimising supply chain decisions in their firms.

Conventionally, decisions on distribution and storage are not jointly considered, this is partly as a result of the complexity that results from their joint consideration. Moreover, decisions on distribution (operational) and stocking (tactical) are considered independent of decisions on facility location and network design which are strategic decisions. Integrated problems have been attracting the attention of numerous researchers who have highlighted the benefits of joint consideration of location and inventory decisions. Researchers have established that considering facility location decisions independent of inventory decisions can result in supply chain systems that are below optimal Daskin *et al.* (2002), Candas and Kutanoglu (2007). Research on joint location inventory systems further evolved to consider joint two-echelon systems with service (time) constraints Mak and Shen (2009), Riaz (2013). Nonetheless, jointly considering facility location, inventory and distribution decisions with customer service level considerations (response time requirements) remains a very challenging mathematical modelling task.

Customers in need of service parts usually have desired response times because they intend to get their machines fixed and operational within their desired response times. This makes firms try to locate facilities close enough to customers and this is the motivation for two-echelon systems with a plant at the top echelon and Service Centers (SVCs) at the lower echelon. For systems with slow demand arrival process, it makes sense to assume that direct shipment to the demand node from the plant is a better alternative to keeping inventory at SVCs. Yet in many situations, customers are service time sensitive. Hence, most firms try to maintain positive inventory levels at SVCs which are close enough to customers whose locations are not close to the plant. From Caglar *et al.* (2004), such system is appropriate for service parts structures that have SVCs stocked with service parts inventory. Also situations could arise in which a Service Center (SVC) experiences a stock out situation and the customer's sensitivity to response time could make the customer seek for other alternatives. Thus, the decision maker has to factor in a way to manage stock out situations.

The sensitivity of customers to response times and the desire of the decision maker to satisfy this customer requirement while working within a given budget, makes the incorporation of lateral transshipment an interesting and attractive area of possible research for many two-echelon joint systems. Lateral Transshipments (LTs) are stock transference that occur among same-echelon locations in inventory systems and their effect on two-echelon joint systems with response time constraints has not been studied. If a SVC experiences stockout, demand at that SVC can be fulfilled by means of stock transference from another SVC. LTs are also useful to decision makers with the objective of reducing the penalties of stockouts at facilities Axsater (1990). The complexity that arises from LTs lies in determining the right time to initiate a stock transfer and the destination which will be optimal for the system. It is likely that an LT reduces the immediate stock out risk at its destination but it unavoidably increases future risk of stockout at the origin. Therefore an appropriate LT policy must evaluate these risks and determine when the LT cost is dominated by its expected benefit.

The appropriate LT policy is usually dependent on the characteristics of the inventory system in use. In this study, the SVCs and plant inventory are controlled with a base-stock policy. Base stock or (S -1, S) policies are apparently fitting for slow moving items that have high holding cost for inventory. Moinzadeh and Lee (1986) analytically check the optimality of base stock policies given specific problem parameters. Their findings imply that the optimality of base stock policy holds in a setting that admits low rate of demand and low setup costs in comparison with holding costs. This policy is widely applied in service parts inventory structures in which malfunctioning parts are replaced with new ones from on hand stock. A low malfunctioning system implies a base-stock policy is just appropriate. This is found in many areas, namely, transport, oil and gas, transport, and IT firms. The problem considered in this study may be treated as creating a structure for spare parts inventory.

In this study, we further stretch the research on joint location-inventory systems by incorporating lateral transhipment into a centralised two-echelon spare parts system which jointly considers decisions on facility location and inventory with customer service constraint (response time requirement). This system involves a central plant at the top echelon, which has constraints on capacity and production. This system also has a finite number of SVCs at the lower echelon, which satisfy demand from geographically dispersed customers. A SVC satisfies its assigned demand via on-hand inventory, LT or backorder. The system also has a response time requirement across all facilities.

The problem here is to incorporate LT into a two-echelon joint inventory-location system and simultaneously determine optimal number of SVCs, optimal assignment of customer to SVCs, the on-hand inventory level, backorder level and lateral transshipment level at SVCs, backorder and on-hand inventory level at the plant, under a response time requirement for demand. We consider the problem by formulating and solving a model which minimises total cost for the system.

1.2 Motivation for the study

Cost minimisation and service improvement are two contrasting management objectives. Cost minimisation seeks to reduce expenses, while, improving service could lead to increase in total system cost with the firm experiencing either an increase or a decrease in profit margins. If the decision maker's only objective is to minimise costs, it is very likely that in the long run, customer service will suffer. Also, if the decision maker's only objective is to improve service, there is a chance that the firm might not make tangible profit due to the cost required to improve service. Thus, the need for more efficient means of balancing the contrasting objectives of cost minimisation and service improvement exists and will always exist.

Integrating decisions on facility location, inventory and distribution has been a very interesting and challenging aspect in the study of supply chain for spare parts. The consideration of facility location decisions independent of inventory decisions can result in supply chain designs that are below optimal Daskin *et al.* (2002). Some authors have investigated integrated supply chain systems with service considerations (response time requirements) Caglar *et al.* (2004) and Mak and Shen (2009). Also, many authors have considered the effects of LT in many systems and LTs have been found to improve service levels in inventory problems Lee (1987). However, LT has not been considered for the system treated in this thesis.

We are thus motivated to consider a new approach for formulating models which helps the decision maker to balance the contrasting objectives of cost minimisation and service improvement. We do this by incorporating LT into an integrated two-echelon system with response time requirement. So far, researchers have not yet formulated a model that incorporates lateral transshipment into a two-echelon system that simultaneously considers facility, inventory and distribution decisions subject to response time requirement.

1.3 Statement of the problem

Lateral Transshipment (LT) (stock movement between facilities on the same echelon), has been used as an option for reducing the occurrences of stockout and excess stock in many multi-echelon environments. Caglar *et al.* (2004) studied a model for a two-echelon spare parts system with RTR controlled with continuous (S-1,S) policies at both echelons. The system considered by Caglar *et al.* (2004) is made up of a plant, multiple service centers and geographically spaced customers. Mak and Shen (2009) integrated facility location into the problem considered by Caglar *et al.* (2004). Several LT models have been formulated for many supply chain systems. However, the incorporation of LT into a system with similar structures as those of Caglar *et al.* (2004) and Mak and Shen (2009) has not been considered. In other words, the incorporation of LT into a system which jointly optimises facility location and two-echelon inventory decisions with response time requirement has not been considered. Therefore, this study was designed to incorporate LT into a two-echelon system which jointly minimises expected cost emanating from facility location and inventory decisions subject to response time requirement.

1.4 Research aim and objectives

The aim of this study is to formulate and solve a model which incorporates LT into a two echelon system with response time requirement which jointly determines optimal facility locations, inventory stocking levels at SVCs, lateral transshipment levels at the SVCs, backorder levels at SVC, inventory stocking levels at the plant, and backorder levels at the plant.

The following are the objectives of this study:

1. To evaluate the distribution of number of units in transit to replenish inventory at SVCs and at the plant.
2. To determine steady state expected inventory levels at both echelons.
3. To formulate mathematical models for joint two-echelon systems with response time requirement and lateral transshipment.
4. To investigate properties of the models.

5. To examine optimality conditions for cases of probabilistic failure of SVCs and stochastic demand.

6. To perform computational experiments on the model so as to highlight model properties using General Algebraic Modeling System (GAMS).

1.5 Research methodology

The arrival of customers at SVCs followed the arrival process of the M/M/1 queue system. The balance equation of this queue along with the property of the base stock inventory control policy was used to derive service center steady state expected level for on-hand inventory I_{vw}, backorder B_{vw}, and LT T_{vw}. The derived steady state levels were used to formulate the two-echelon LT model. This model was then decomposed via Lagrange relaxation. This model was further reduced by relaxing the assignment variable's integrality constraint and fixing the base stock levels. The reduced model was investigated for convexity. Karush-Kuhn-Tucker (KKT) conditions were used to investigate global optimality, which was also examined for the cases of probabilistic facility failure and stochastic demand. Computational experiments were carried out using GAMS.

1.6 Structure of the study

This thesis' structure is as follows. In Chapter One, we present the introduction. In Chapter Two, we present the literature review. Chapter Three consists of methodology. In Chapter Four, we present results and discussion, and in Chapter Five, we presents our summary and conclusions.

Chapter 2
LITERATURE REVIEW

2.0 Introduction

The research carried out in this study consists of two major areas in service parts supply chain. These areas are:

1. inventory-location

2. lateral transshipment

This chapter began with review of literature on inventory-location which are relevant to this study . Thereafter, a review was also presented of relevant literature on lateral transshipment.

2.1 Inventory-location

Significant portions of operations research literature have been fully dedicated to the study of facility location models. Location models are usually designed to answer questions like what number of facilities should be opened, what should be the capacity of opened facilities, where should the facilities be located and number of customers per facility? Daskin (1995). Some examples of location models are; fixed charge location problems, covering problems, median problems and center problems. Drezner (1995), Daskin (1995) along with Drezner and Hamacher (2002) treat location problems extensively. Many conventional facility location model formulations are deterministic; that is all of the model parameters are assumed to be known and constant. While the assumption for stochastic location models is that there are model parameters that are uncertain and the aim is to ascertain the best decision given uncertainty. Snyder (2006) presents a review on stochastic location models.

Until recently, location decisions have been considered independent of inventory decisions. The literature on inventory is very wide, so we restrict this review to models that have similar structure to our problem.

The paper by Shebrooke (1968) has made one of the greatest impacts in multi-echelon inventory research. He considered a mathematical model applicable to repairable items called "Multi-Echelon Technique for Recoverable Item Control (METRIC)". METRIC model uses an approximation to the distribution of items in replenishment to circumvent the computationally burdensome exact representation, this has given it wide applicability. All facilities are controlled using (S-1,S) or basestock policies. METRIC model is used to get an approximate value for the total expected backorder in the system for minimisation. Generalising Sherbrooke's model, Muckstadt (1973) developed the MOD-METRIC model for the consideration of items using hierarchical parts structures. The model allows two levels of parts to be considered, an assembly (e.g. an engine) and it's components.

Graves (1985) considered a model which determines inventory stock level in multi-echelon systems. He presented an exact procedure for determining expected inventory level. He showed numerically that his approach was more accurate than the METRIC approach under same problem structure. The downside of his approach is that it is computationally onerous. Sherbrooke (2004) gives full treatment of inventory in multi-echelon arena.

Caglar *et al.* (2004) studied a model for a two-echelon spare parts system controlled with continuous (S-1,S) policies at both echelons. They imposed a constraint on response time and created efficient algorithms to determine optimal stock level at both echelons.

Shen *et al.* (2003) looked at an integrated location-inventory system that employed a continuous review (r,Q) policy for inventory management. They utilise an economic order quantity based approximation of the stochastic demand. The model displays the effect of economies of scale, Eppen (1979), since demand pooling from a number of retailers to a single facility causes a reduction in safety stock (Eppen's model clearly shows the savings achieved when risk pooling is allowed). As a result of the term dependent on inventory cost, which is approximated by the EOQ, the objective function happens to be nonlinear. They assume identical mean-variance ratio for each retailer demand. Hence the problem was solved by column generation algorithm following the combination of both square root terms in the objective function. Computational results of their study suggest that when decisions on location and decisions on inventory are not jointly optimised, the facilities opened will exceed the optimal number. Daskin *et al.* (2002) consider nonlinear integer programming location problem that integrates inventory costs and cost of safety stock. They utilised Lagrange relaxation and showed that their relaxation technique improved the computational time of Shen *et al.* (2003). Shu *et al.* (2005) slackened the identical mean-variance ratio assumption. They utilised submodular function minimisation to solve the problem. Shen and Qi (2006) introduced costs on operational routing and obtained a model which was a nonlinear integer problem. Shen and Daskin (2005) examined the relationship between minimising cost and maximising service. Their service level definition is dependent on distance from distribution centers to retailers. They showed that the structure of the model having service constraints was similar to the structure of the model of Shen *et al.* (2003) and that their model was solvable with same algorithms. The authors further introduced a genetic algorithm which is able to efficiently produce the trade-off curve between service and cost. Their results suggested that significant service improvements can be attained relative to the solution that yields minimum cost at relatively slight incremental cost.

Ozsen *et al.* (2008) and Ozsen *et al.* (2009) extended the model by Shen *et al.* (2003) by considering the effects of capacitated facilities which follow single sourcing and multiple sourcing. This is a departure from the traditional approach that defines capacity on the basis of maximum demand assignable to a facility. The authors imposed a limit on storage space available for holding facility inventory. They formulated a nonlinear model having nonconvex objective function which was solved using Lagrange relaxation and linear relaxation. Shen (2006a) further extends the model by relaxing the assumption of all demand being served. He presented a model which allows demand choice with flexible pricing. The results shows that flexility of demand-choice can greatly improve profitability of the supply chain this is because the firm is free to give higher service priority to more profitable customers. Shen (2006b) gives a well treated survey of these models. So far, we have only considered single-echelon problems.

Nozick and Turnquist (2001) considered locating Distribution Centers (DCs) in a two-echelon environment that holds inventory at DCs and at a plant. Their system environment is like the one considered in the conventional literature of two-echelon systems for inventory, e.g. Shebrooke (1968) and Graves (1985). The (S-1,S) policy is utilised to control inventory at the DCs at the plant. Also, the plant has unlimited production capacity. They approximated safety stock using linear approximation dependent on number of DCs. When the inventory cost is constant and incorporated into the fixed cost of location, the resultant location model has similar structure with the fixed charge problem without capacity bounds.

Candas and Kutanoglu (2007) presented an integrated inventory-location multi-commodity problem for a two-echelon system problem which optimises fill rates and stock levels in order to achieve a time driven service level for the entire system. They formulated an integer programming model which was also nonlinear and proposed a linearisation-based procedure for solving small and medium sized cases. Moreover, they compared the approach of simultaneous consideration of decisions on inventory and location in a model to the approach that determine optimal location decisions first before finding optimal inventory levels for the given facility locations. Their result showed that following the simultaneous approach results in solutions which can attain same service level with reduced cost.

All papers mentioned so far considered deterministic replenishment lead time, except that of Nozick and Turnquist (2001) who modelled the replenishment process with an $M/G/\infty$ queue. The inventory-location literature has only a handful of models that consider stochastic replenishment lead times.

Eskigun *et al.* (2007) consider supply chain network models that incorporates the consequence of choice of mode on customer satisfaction and system- wide service time. Their models permit nonidentical lead times for different modes. Moreover, the node response time is dependent on the quantity of demand processed. Their approach resulted in deterministic models and they used mean lead time from the modes as their input. Their models did not consider the effects of inventory.

Sourirajan *et al.* (2007) studied a single-item joint inventory-location model whose replenishment lead time is stochastic and which is dependent on the demand size processed by the facility. The mean replenishment lead time was represented by an approximate queue formula. They utilised stochastic lead time of the resulting model to derive the required level of safety stock pertaining to each facility. Benjaafar *et al.* (2008) studied the joint optimisation of decisions on location and inventory control for a one-echelon system and represented the replenishment process by an M/M/1 queue that considered the congestion effect.

Mak and Shen (2009) considered a two-echelon joint location-inventory problem that incorporated service consideration. They modelled the manufacturing process as a queue and formulated a mixed integer nonlinear programming model and found the solution by means of a Lagrange heuristic.

Puga and Tancrez (2016) analysed a location-inventory model with stochastic demand for large supply chain systems. They gave a continuous nonlinear formulation which integrated decisions on stocking, location and allocation, and included the transportation, inventory and facility costs. They relied on a property which made their model linear when some variables were fixed. They proposed a heuristic based algorithm which solved the resulting linear program and the solution was then utilised in improving variable estimations for the subsequent iteration. Computational experiments showed the efficiency of their heuristic algorithm for finding solutions that are fast and almost optimal for large supply chain systems. However they did not incorporate service constraints neither did they include lateral transshipment in their formulation.

Kok *et al.* (2018) very recently, carried out a comprehensive literature review of inventory location problems with demand uncertainty. They proposed a typology for inventory management in multi-echelon systems and also identified current research gaps.

2.2 Lateral transshipment

Existing LT literature have two key properties which emanate from timing of LTs. LTs can be planned to occur at scheduled times before realisation of all demand, or they can be performed at any time to salvage zero inventory situations. These two classifications of transshipments are known as proactive and reactive. Proactive transshipment models perform LTs to rebalance inventory in all locations belonging to the same echelon at preset instants in time. This is done in advance and is done in such a way that the related costs are minimal. Reactive transshipment models perform LTs in response to circumstances in which a facility faces a zero inventory level (or the likelihood of a zero inventory level) and another has enough stock available. This class of LT is appropriate for settings that have LT costs which are reasonably low when compared to inventory holding cost and penalty cost of not meeting customer demand instantly; this commonly occures in spare parts systems. Kranenburg (2006) considers a semi-conductor firm ASML under a reactive LT setting, and shows that incorporating LTs results in annual savings of up to 50% savings of total service parts inventory cost.

LTs have been expressed using different terminologies, namely, substitutions and transfer of stock, reallocation of stock, lateral resupply, etc. Permitting LTs makes the system more flexible, the implication of this being that controlling and optimising the system becomes more challenging. In addition to deciding ordering dynamics from the ’regular’ supplier, decisions timing, size, source, and destination of transshipment are also necessary. Due to this additional difficulty, LT literature is majorly limited to two echelons systems, further limitation is also observed in some contributions that consider only one echelon and/ or allow LTs among just a restricted number of facilities. However, optimal control of LTs has been studied under various differing scenarios. Some of such scenarios are; number of echelons (one or two), stocking locations, ordering policy, etc.

A major characteristic of a LT policy lies in the type of pooling employed; complete pooling or incomplete/partial pooling. Policies in which the LT facility is permitted to share all its stock are classified as complete pooling, while policies which don’t permit a facility sharing all its stock are classified as partial pooling.

The several contributions to LT literature are further categorised by certain characteristics which depend on the inventory policy and their modeling of LT specifically. Listed below are inventory system properties that help to classify the literature on LTs:

1. Number of entities; single entity or multiple entities
2. Number of echelons; multi-echelon or single echelon
3. Number of facilities; 2, 3, ...
4. Identical facilities? ; Yes, (do they have uniform cost) or no
5. Unfulfilled demands; allow lost sales or allow backorder
6. Timing of order policy; periodic or continuous
7. Ordering policy/rule; (S - 1, S), (s,S), (R,Q), General or Other
8. Transshipment type; reactive or proactive
9. Type of pooling; partial or complete
10. Type of decision makers; decentralised or centralised
11. LT cost; per transshipment, per item, both or none

This research focuses on reactive transshipments in a centralised setting. The type or transshipment considered in this study is reactive. Consequently we review reactive transshipments in the next subsection.

2.2.1 Reactive lateral transshipments

In this subsection, we review existing literature on reactive LTs under periodic review and continuous review settings.

2.2.1.1 Reactive lateral transshipments with periodic review

Here LTs are considered under the following classifications: one echelon centralised systems and two echelon centralised systems.

2.2.1.2 One echelon centralised systems

The pioneer publication on reactive transshipment was the article of Krishnan and Rao (1965). Their model followed periodic review and they obtained an expression for optimal inventory levels for several locations which allow transshipment at each replenishment period's end once demand is known. They assumed same transshipment costs across all locations, and showed that it is sufficient for the transshipment rule to select one location with an excess to satisfy the stockout in another location.

Continuing with this approach, Robinson (1990) considered a problem that allow multiple locations and multiple periods with non-uniform costs and demand distributions and multiple periods. They established the optimality of the (S-1,S)/ basestock policy and demonstrated the stationarity of the optimal order-up-to point. The optimal solution only exists for scenarios that have two non-identical facilities or for systems that have multiple identical facilities. Authors suggest an approximation based solution that uses Monte Carlo sampling to avoid the complexities associated with the problem for more general multi-location environments. However, the precision of this method is dependent on demand sample, and there is no guarantee of convergence to optimality. Nonas and Jornsten (2007) found a different 'greedy transshipment rule' which solves to optimality a three location scenario while the multi-location scenario can only be solved given certain conditions.

Yang and Qin (2007) consider a similar model from another perspective, which is consideration of 'virtual' transshipments. Their study is domiciled in the energy industry and its possible for transshipments to occur even when both locations have negative stock, thereby redistributing backorders. This underscores one of the various means of diversifying transshipment policies. Another case of diversification is the work by Kochel (1996) who consider the likelihood of stock sales outside the network before demand arrival and then performing transshipment after demand arrival.

A similar model to Robinson (1990) is the one by Herer *et al.* (2006), who examined a more widespread cost mechanism and utilise LP in addition to a network flow structure to develop a procedure that is more robust than the procedure of Robinson (1990). This was further enhanced by Ozdemir *et al.* (2006) with the introduction of capacity constraints to the model. They observed that the constraints alter the distribution of stock in the system's inventory and also result in an increase in total cost.

Hu *et al.* (2005) adopt a different method to improve the problem for multiple locations. They formulate a simplified model that is able to approximate order policies for a system with few stocking points and low transshipment cost in comparison with holding cost and cost of stockout. The defining property of the model is the assumption of free and instantaneous transshipments.

Moinzadeh and Lee (1986) also presents same pattern of reactive transshipment after demand. He considered a model having two locations along with negligible transshipment times and lead times. Tagaras and Cohen (1992) later incorporated positive lead times. These authors treated complete and partial pooling, and decisions on stocking were computed by means of an approximation heuristic. It was shown that for this system complete pooling performed best. Tagaras (1999) more comprehensively consider a similar three-location network, he found that the exact transshipment does not greatly affect the results in a complete pooling environment. He showed that increasing the number of locations resulted to increase in the advantage of transshipments.

Further work on systems with multiple locations was done in Archibald *et al.* (2009). They considered the real life scenario of a tyre dealer with a wide location network. Archibald *et al.* (2009) reduced the difficulty associated with dimensionality for systems of this type by getting an approximate value for the dynamic programming function. This was done using pairwise decomposition, which considered two locations per time and was shown to be an improvement on previous heuristics proposed by Archibald (2007) in a complete pooling environment. One constraint of the model is that the review period is same for all locations. Archibald *et al.* (2009) slackened this constraint by utilising a two-step heuristic which first computed a static policy to ascertain which location satisfied a particular demand, and then applied dynamic programming to improve the policy.

A different research direction is the consideration of systems that have dynamic deterministic demand. Herer and Tzur (2001) considered a problem with two locations and developed a solution for such problem. Seeking to determine optimal transshipment and ordering decisions within a finite frame, they considered key characteristics of the system. These characteristics form a framework which makes it possible to obtain solutions for models of this type of model in polynomial time. Herer and Tzur (2003) extended this problem and considered a setting with multiple locations.

Finally, Herer *et al.* (2002) take a more general look at the importance of transshipments under the label 'legility' which intends to create an inventory system that is lean and agile inventory. By examining a number of the models previously discussed, they showed that transshipment helps to improve the system performance under the two criteria and produced a procedure for analysing this information.

2.2.2 Centralised two-echelon systems

For systems having two echelons stockouts can be satisfied via several means. One possibility is LT but there could arise situations for which it is more beneficial to have emergency shipments from a central warehouse. Wee and Dada (2005) considered this problem using five diverse combinations of emergency shipment, transshipment, and no movements at all. He devised a procedure for determining the optimal setup with respect to a particular model description. His research helped to establish emergency stock movement structure.

Dong and Rudi (2004) examined the LT benefits for a manufacturer supplying a set of retailers. Comparing the case in which the price leader is the manufacturer, with the case having exogenous prices, they found that for exogenous prices, the retailers benefitted more when demand within the entire network is uncorrelated. They used a Stackelberg game to model endogenous price and found that the manufacturer took advantage of his price leadership to achieve increase in his benefits, this was worse for retailers who chose to use transshipments. Their results are valid for normal distributed demand, Zhang (2005) extended the results in include general demand distributions.

A case study mainly based on retail was considered by Bendoly (2004) who studied a model having store and Internet based customers. Bendoly utilised LT ideas to show how a system's performance can be improved via partial pooling of items. The model considered a retail environment in which stores operate side by side with internet channels.

2.2.3 Reactive lateral transshipments with continuous review

Reactive LT models look to transship anytime a stockout occurs or whenever there is a likely stockout. Reactive LT models can use either partial or complete pooling. This research focuses on complete pooling, hence only literature on complete pooling alone is considered. Complete pooling is frequently used in environments such as, service parts, that have typically large holding and backorder costs when compared with transshipment cost. METRIC is the basic multi-echelon model for repairable service parts Shebrooke (1968). In this model, damaged parts are taken to a central base for repairs. This base then supplies the individual bases with the repaired items. A (S-1,S) ordering system is used by these bases to resupply their stocks.

Lee (1987) considered LT in such model. Lee divided stock locations into various pooling groups, and focused on one of such groups. He assumed identical locations with Poisson demand. Damaged parts are taken for repairs at the designated central repair facility having infinite repair capacity with positive and probabilistic repair times. Lee tested three possible rules for determining transshipment source: maximum on hand stock, random selection, minimum number of orders outstanding , and random selection. For all rules, he derived approximations for the fill rate service level which is used for cost minimisation under some service level constraint. Lee found that for rules tested, using emergency LT led to substantial savings due to less stock requirement at the bases. Also, for all rules tested, Lee found no substantial difference in performance. Axsater (1990) relaxed the assumption made concerning identical locations. He also included stock holding for the central depot and presented improved methods for the approximation of service level.

Kukreja *et al.* (2001) studied a similar model and the lower echelon was their only focus. They utilised a different rule to select transshipment source: transship from location with on hand stock and the minimum transshipment cost. Kukreja *et al.* (2001) found fill rate approximations which they applied to a heuristic for determining optimal location stock levels. Kukreja and Schmidt (2005) extended this model to consider systems controlled with (s,S) policies and have demand processes that are compound Poisson. They selected the transshipment source using a dynamic programming rule and proposed a simulation based approach for finding the optimal values of s and S. Huo and

Li (2007) considered a different order policy in similar setting. They considered the lower echelon of a system controlled with a (R,Q) rule and their approximations were similar to those obtained by Axsater (1990). A model that was similar to the model of Axsater (1990) was considered by Jung *et al.* (2003). Their facilities had finite repair capacities with fill rate approximations which were used in an algorithm for finding optimal stock level. Sherbrooke (1992) considered same model as Axsater (1990), but differed by evaluating expected backorder. He determined stock levels using the VARI-METRIC model by Sherbrooke (1986) and evaluated backorder decrease due to the application of LT. He concluded that the LT has the greatest impact on parts possessing low demand rates. For a system with one echelon, Wong *et al.* (2006) and Yanagi and Sasaki (1992) determined the downtime due to waiting for a LT or backorders and thus particularly considered non-zero lead time. Wong *et al.* (2006) built on a previous model having negligible transshipment lead time Wong *et al.* (2005), and derived exact service level expressions (demands fraction fulfilled without using backorder), the expected downtime due to LT, and the expected value of LT. They derived an approximation for expected downtime due to backorders. Yanagi and Sasaki (1992) focused on the development of approximations for the average number of failed items in addition to the probability of a backorder occurrence.

An important feature also considered in spare parts modelling is time-based service level. Lee (1987) and Kutanoglu (2008) considered this sort of system. The requirement to satisfy fractions of demand within a given time interval is known as time-based service level. The latter article examined cost and service level in networks with two or three locations. This allowed the achievement of time-based service level with particular sensitivity to demand changes. This model pointed out important insights such as how service requirements that are time-based are more important in spare parts environments than fill rate, and how emergency LT results in improvement in response time performance. The former paper considered the determination of appropriate stocking levels so as to minimise cost. This was achieved by using an enumeration algorithm.

Although many of the models previously discussed are similar, there are slight differences between these systems that can significantly change which policy is best for the system. Utterbeeck *et al.* (2009) considered this problem and proposed a procedure for determining the most efficient from six viable network structures for a particular system. They considered single and double echelon networks with available options being both

LT and emergency shipments. They determined the best structure by utilising an optimisation setup with guided local search. This provided a new transshipment problem feature which can be studied and optimised along with the transshipment and ordering policies.

Tiacci and Saetta (2011) examined the relative efficiency of two lateral shipments methods in minimising mean supply delay of a consumable item, regarding a classical policy with no lateral shipment. They implemented a two-echelon simulation model and performed an experiment by varying various parameters, such as number of warehouses, lead time of supply from the central facility, demand uncertainty for an item, and warehouse size variability. In almost all newtwork configurations, their results showed appreciable reductions in mean supply delay when lateral shipment is allowed with regard to the particular classical policy.

Paterson *et al.* (2012) proposed an enhanced reactive method such that individual transshipments are seen as an opportunities for proactive redistribution of stock. They adopted a quasi-myopic procedure to the develop an enhanced reactive transshipment rule that performed strongly. Comparing their results in a transshipment approach that is fully reactive, their procedure resulted to highly improved service levels with safety stock reduction and reduced total costs, especially for large systems. They also determined an optimal policy for small networks and showed that the enhanced reactive policy significantly closes the optimality gap.

Yang *et al.* (2013) considered a service parts inventory location problem with lateral transshipment and flexible replenishment stock. They proposed a customer oriented service measure and provide and approximation optimizing inventory allocation subject to the this measure. However their lead time was deterministic.

Most papers that considered service measures used the fill rates as service measure, for example Yang *et al.* (2013), Kutanoglu and Mahajan (2009), and Kutanoglu (2008). Only very few have considered a response time threshold as the service constraint, for example Caglar (2001), Caglar *et al.* (2004), and Mak and Shen (2009). So far researchers on joint inventory location have not considered problems with lateral transshipment and service constraints; this is the gap this study intends to fill.

2.3 Some previous results

In this section, we take a close look at some techniques for determining optimal inventory policies and optimal stocking levels for our two-echelon system. The system considered in this study involves demand arrivals that occur one at a time. Firstly, basic concepts comprising of definitions and some existing results on Poisson process relevant to this study were presented. Afterwards, relevant concepts from nonlinear constrained optimisation such as the Lagrangian dual problem and convexity were presented. Thereafter, the inventory policy used in the study was highlighted and it's effects on performance measures and inventory levels were discussed. In addition, the determination of inventory levels for systems with lateral transshipment was dealt with. Finally, some closely related models were presented.

2.3.1 Basic concepts

This study dwells on spare parts supply chain for expensive parts with low failure rate. Consequently, demand arrival at a SVC occurs due to service part failure, which is categorised as a rare event. The counting process which is used to model scenarios with this property is the Poisson process, Shebrooke (1968), Caglar *et al.* (2004), Mak and Shen (2009), Yang *et al.* (2013) and Riaz (2013) are some authors who have modelled demand arrival using Poisson process in a two-echelon spare parts system.

Definition 2.3.1.1: Counting process

If $\{\bar{N}(t), t \geq 0\}$ is a stochastic process and total event occurrences at time t is denoted by $\bar{N}(t)$. $\bar{N}(t)$ is called a counting process if it satisfies the following conditions:

1. $\bar{N}(t) \geq 0$.

2. $\bar{N}(t)$ is integer valued.

3. $s < t$ implies that $\bar{N}(s) \leq \bar{N}(t)$.

4. If $s < t$, then $\bar{N}(t) - \bar{N}(s)$ represents the total events occurrence in (s, t].

Definition 2.3.1.2: Poisson process

The process $\{\bar{N}(t), t \geq 0\}$ is said to be Poisson with rate λ, where, λ is positive, if

1. $\bar{N}(0) = 0$

2. The counting process posseses increments which are independent and stationary.

3. For all $s, t \geq 0$

$$P\{\bar{N}(t+s) - \bar{N}(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, n = 0, 1, \cdots$$

4. Not more than one event occurs at a time.

Given a Poisson process which has time of it's first event represented by $\bar{T}_1$. Furthermore, for $n > 1$, let $\bar{T}_n$ represent the time in between the $(n-1)^{th}$ and the n^{th} event. Then $\{\bar{T}_n, n = 1, 2, ...\}$ is known as the interarrival times sequence.

Proposition 2.3.1.1 (Ross (2010) p. 317)
$\bar{T}_n, n = 1, 2, ...$, are idd exponential random variables having mean $\frac{1}{\lambda}$.

The stochastic process above $(\bar{N}(t))$ is a Markov process. We can therefore say, that a Poisson process with rate λ is a counting process $\{\bar{N}(t), t \geq 0\}$ whose interarrival times $\bar{T}_1, T_2, ...$ have identical exponential distribution functions
$P\{T_n \leq x\} = 1 - P\{T_n > x\} = 1 - e^{-\lambda x}, x \geq 0$.

2.3.1.1 Merging and splitting of Poisson processes

Lots of scenarios abound which require the merging and splitting of Poisson processes for various purposes. The next theorem shows that merging and splitting of Poisson processes also result in Poisson processes.

Theorem 2.3.1.1(Tijms (2003) p. 6)

1. Let $\{\bar{N}_1(t), t \geq 0\}$, $\{\bar{N}_2(t), t \geq 0\}$ be independent Poisson processes having rates λ_1 and λ_2 respectively, where $\{\bar{N}_i(t)\}$ denotes type i arrivals. Let $\bar{N}(t) = \bar{N}_1(t) + \bar{N}_2(t), t \geq 0$. Then the merged process $\{\bar{N}(t), t \geq 0\}$ is Poisson having rate $\lambda = \lambda_1 + \lambda_2$. In this merged process, let $\bar{Z}_k$ denote the time from the (k-1)th arrival to the kth arrival. Also, let $\bar{I}_k = i$ if the k^{th} arrival for the merged process happens to be of type i, then for any k = 1, 2, . . . ,

$$P\{\bar{I}_k = i | \bar{Z}_k = t\} = \frac{\lambda_i}{\lambda_1 + \lambda_2}, i = 1, 2 \qquad (2.3.1)$$

independently of t.

2. Given a Poisson process $\{\bar{N}(t), t \geq 0\}$ with rate λ, let each arrival be categorised as either an arrival of type 1 or an arrival of type 2 arrival having probabilities $\bar{p}_1$ and $\bar{p}_2$, respectively, independent of any other arrival. Let $\bar{N}_i(t)$ denote number of type i arrivals at time t . Then $\{\bar{N}_1(t)\}$ and $\{\bar{N}_2(t)\}$ are two independent Poisson processes having respective rates $\lambda\bar{p}_1$ and $\lambda\bar{p}_2$.

Most two echelon inventory problems exploit this property when determining steady state levels at both echelons; Graves (1985), Caglar *et al.* (2004), Mak and Shen (2009), Yang *et al.* (2013) and Riaz (2013). In this study, this property was exploited in the determination of SVC, pool and plant inventory levels.

2.3.2 Nonlinear constrained optimisation

In this section, the necessary tools from nonlinear programming needed for this study are presented. The ideas presented follow those given by Boyd and Vandenberghe (2004) and Gowers *et al.* (2008). For $n, m, p > 0$, a nonlinear minimisation problem is an optimisation problem which is expressible as:

$$\begin{aligned} \min\ & g_0(y) \\ \text{subject to } & g_b(y) \leq 0 \text{ for each } b \in \{1, \dots, m\} \\ & h_c(y) = 0 \text{ for each } c \in \{1, \dots, p\} \end{aligned} \tag{2.3.2}$$

where $y \in \mathbb{R}^n$ is the optimisation decision variable, $g_0 : \mathbb{R}^n \to \mathbb{R}$ is called the objective function or cost function. The inequalities $g_b(y) \leq 0$ are known as inequality constraints, while the corresponding functions $g_b : \mathbb{R}^n \to \mathbb{R}$ are called inequality constraint functions. The equations $h_c(y) = 0$ are known as the equality constraints, while the functions $h_c : \mathbb{R}^n \to \mathbb{R}$ are known as the equality constraint functions.

The set of points for which the objective and all constraint functions are defined,

$$\mathcal{D} = \bigcap_{b=0}^{m} \textbf{dom}\, g_b \cap \bigcap_{c=1}^{p} \textbf{dom}\, h_c$$

is called the domain of the optimisation problem.

A point $y \in \mathcal{D}$ is said to be feasible if it satisfies the constraints $g_b(y) \leq 0, b = 1, ..., m$, and $h_c(y) = 0, c = 1, ..., p$. The problem is said to be feasible if there exists at least a feasible point, and infeasible otherwise. A constraint set or a feasible set is the collection of all feasible points. The optimal value d' of the problem (2.3.2) is defined as

$$d' = \min\{g_0(y)|g_b(y) \leq 0, b = 1, ..., m, h_c(y) = 0, c = 1, ..., p\}$$

We say y^* solves the problem (2.3.2), if y^* is feasible and $g_0(y^*) = d'$. The set of all optimal points is the optimal set, denoted

$$Y_{opt} = \{y|g_b(y) \leq 0, b = 1, ..., m, h_c(y) = 0, c = 1, ..., p, g_0(y) = d'\}$$

If an optimal point exists for the problem (2.3.2), we say the optimal value has been attained or achieved, and the problem is solvable. If Y_{opt} is empty, we say the optimal value is not attained or not achieved (for minimisation problems this always happens when the problem is unbounded below).

2.3.3 The Lagrange dual function

The Lagrangian takes the constraints in the program (2.3.2) and integrates the constraints into the objective function. The Lagrangian $\bar{L}$: $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with this optimisation problem is

$$\bar{L}(y, \lambda_b, \nu_c) = g_0(y) + \sum_{b=1}^{m} \lambda_b g_b(y) + \sum_{c=1}^{p} \nu_c h_c(y) \tag{2.3.3}$$

λ_b is the Lagrange multiplier corresponding to the bth inequality constraint $g_b(y) \leq 0$; similarly ν_c is the Lagrange multiplier corresponding to the cth equality constraint $h_c(y) = 0$.

The Lagrange dual function $k : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is the minimum value of the Lagrangian over y: for $\lambda_b \in \mathbb{R}^m, \nu_c \in \mathbb{R}^p$,

$$k(\lambda_b, \nu_c) = \min_{y \in \mathcal{D}} \bar{L}(y, \lambda_b, \nu_c) = \min_{y \in \mathcal{D}} \left(g_0(y) + \sum_{b=1}^{m} \lambda_b g_b(y) + \sum_{c=1}^{p} \nu_c h_c(y) \right) \tag{2.3.4}$$

The dual function gives lower bounds to the optimal value d' of the problem (2.3.2): for any $\lambda_b \geq 0$ and any ν_c we have

$$k(\lambda_b, \nu_c) \leq d' \tag{2.3.5}$$

The Lagrange dual to the optimisation program (2.3.2) is

$$\max_{\lambda_b \in \mathbb{R}^m; \nu_c \in \mathbb{R}^p} k(\lambda_b, \nu_c), \text{ subject to } \lambda_b \geq 0 \tag{2.3.6}$$

The dual optimal value d^* is

$$d^* = \max_{\lambda_b \geq 0; \nu_c} k(\lambda_b; \nu_c) = \max_{\lambda_b \geq 0; \nu_c} \min_{y \in \mathcal{D}} \bar{L}(y, \lambda_b, \nu_c)$$

Since $k(\lambda_b, \nu_c) \leq d'$, we know that $d^* \leq d'$. The quantity $d' - d^*$ is called the duality gap. If $d' = d^*$, then the primal and its dual exhibits strong duality. Also, if $d' \neq d^*$, then the primal and its dual exhibits weak duality.

2.3.4 Convex optimisation

A convex optimisation problem is a problem of the form

$$\begin{aligned} &\min g_0(y) \\ &\text{subject to } g_b(a) \leq 0 \text{ for each } b \in \{1, \cdots, m\} \end{aligned} \tag{2.3.7}$$

where the functions $g_0, g_1, \cdots, g_m$: $\mathbb{R}^n \rightarrow \mathbb{R}$ are convex, that is, they satisfy

$$g_b(\alpha x + \beta y) \leq \alpha g_b(x) + \beta g_b(y), b \in \{0, \cdots, m\} \tag{2.3.8}$$

for all $x, y \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbb{R}$, with $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$.

Theorem 2.3.4.1 (Winston (2004) p. 632)
Consider a nonlinear programming problem defined as (2.3.7). Any local minimum for (2.3.7) is a global minimum.

2.3.5 Mixed integer nonlinear problem

Mixed-Integer Nonlinear Problem (MINLP) combines the difficulty of optimising over sets of discrete variables with the complexities associated with handling nonlinear functions, Bonami *et al.* (2012). MINLPs are easily expressed as (2.3.2) with an integrality constraint.

$$\begin{aligned}
&\text{minimize } g_0(y) \\
&\text{subject to } g_b(y) \leq 0 \text{ for each } b \in \{1, \ldots, m\} \\
&\qquad h_c(y) = 0 \text{ for each } c \in \{1, \ldots, p\} \\
&\qquad y \in X \subset \Re^n \\
&\qquad y_i \in Z, \forall i \in I
\end{aligned} \tag{2.3.9}$$

I is the index set of discrete variables. A MINLP is said to be convex if it's continuous relaxation is convex.

2.3.6 (S-1,S) Inventory policy under Poisson demand

In this study, inventory is controlled using a (S-1,S) policy which has proven to be most suitable for items whose demand arrival occur one at a time, Sherbrooke (1992), Candas and Kutanoglu (2007), Mak and Shen (2009). The (S-1,S) policy is also called order-up-to policy or basestock policy. The inventory position of a given facility is its stock level, S. **Following a (S-1,S) policy implies that a replenishment order is made immediately, anytime a demand occurs for an item or more. The order size matches the the demand size**. The amount in replenishment at any random time is an important random variable in studying the characteristics of systems controlled with a (S-1,S) policy. Once the stationary distribution of the amount in replenishment has been established, the stationary distributions for on-hand, backordered and transshipped inventory can be derived easily

Let λ denote the demand arrival rate of the Poisson driven customer order process and let N represent the number of orders in replenishment to the facility being considered.

Theorem 2.3.6.1 (Muckstadt (2005) p. 39)

Suppose demand rate for an item is λ, its arrival process is Poisson, and the base stock level for the item is set to S. Furthermore, suppose that $g(L)$ are density functions with mean L corresponding to the replenishment time random variables, and $G(L)$ are corresponding distribution functions . Also suppose that the replenishment times are idd from one customer order to another customer order. Then the long run probability that $\bar{s}$ units are in replenishment is given by

$$P\{N = \bar{s}\} = e^{-\lambda L} \frac{(\lambda L)^{\bar{s}}}{\bar{s}!} \tag{2.3.10}$$

Thus, the probability that we have $\bar{s}$ in the replenishment system is Poisson distributed with mean λL; that is, there is no need to know the density function for the replenishment time, but only the mean replenishment time, L .

2.3.7 Performance measures

The ready rate: Ready rate for a stock level S, $R(S)$ represents the probability that no backorder exists at any random time, it represents the probability that number of items in replenishment is equal to S or less.

$$R(S) = \sum_{\bar{s}=0}^{S} P\{N = \bar{s}\}. \tag{2.3.11}$$

The fill rate: the fill rate, $F(S)$, for a particular stock level S, represents the expected portion of demands satisfied instantly from on-hand stock. Suppose one customer order arrives, a single unit of the customer order will be satisfied if the quantity of units in replenishment is equal to S - 1 or less. Thus,

$$F(S) = \sum_{\bar{s} \le S-1} P\{N = \bar{s}\} \tag{2.3.12}$$

Thus, in this case,

$$F(S) = R(S) - P\{N = S\}$$

We see that $R(s) > F(S)$. When using either a ready rate or fill rate measure, one is not interested in the duration of occurrence of backorders. Thus, for example, having a 96% fill rate means that, on average, from every 100 ordered units 96 requests are satisfied immediately. However, the time taken to fulfill the other 4% of the ordered units is not measured. Thus, a firm which keeps high fill rates might not truly satisfy all its customers needs. Kutanoglu (2008), Kutanoglu and Mahajan (2009), Yang *et al.* (2013) utilised the fill rate as their performance measure.

Backorder rate: A third performance criterion for single-item systems determines the expected value of outstanding backorders at any random time. This measure is represented by B(S) and represents the duration of time for which backorders occur. Thus, this is a response-time based measure. B(S) is given by the product of the customer demand rate and the average demand waiting time. This follows from Little's law, $L_q = \lambda W_t$ (Little (1961)), where $B(S) = L_q$, λ is the demand rate, and W_t is the average demand waiting time. In steady state, the expected backorder level is

$$B(S) = \sum_{\bar{s}>S} (\bar{s} - S) P\{N = \bar{s}\}. \tag{2.3.13}$$

That is, $\bar{s} - S$ units are backordered if and only if $\bar{s}$ units are in replenishment, $\bar{s} > S$.

2.3.8 Mathematical properties of fill rate and ready rate

Recall that in steady state the probability of having $\bar{s}$ units in replenishment is

$$P\{N = \bar{s}\} = e^{-\lambda L} \frac{(\lambda L)^{\bar{s}}}{\bar{s}!}.$$

The fill rate for a given stock level S is

$$\begin{aligned} F(S) &= 1 - \sum_{\bar{s} \geq S} P\{N = \bar{s}\} \\ &= \sum_{\bar{s}<S} P\{N = \bar{s}\} \end{aligned} \tag{2.3.14}$$

Suppose we aim to select stock levels that maximise the average fill rate for some specific target investment threshold in inventory. An optimisation problem of this form would be easily solved if $F(S)$ happen to be discretely concave. However, it is not. The first difference is denoted by $\triangle F(S)$

$$\triangle F(S) = F(S+1) - F(S)$$

The second difference is denoted by $\triangle^2 F(S)$

$$\triangle^2 F(S) = \triangle F(S+1) - \triangle F(S).$$

Thus

$$\begin{aligned}
\triangle F(S) &= \sum_{\bar{s} \leq S} P\{N = \bar{s}\} - \sum_{\bar{s} \leq S-1} P\{N = \bar{s}\} \\
&= e^{-\lambda L} \frac{(\lambda L)^S}{S!} && (2.3.15) \\
\triangle F(S+1) &= \sum_{\bar{s} \leq S+1} P\{N = \bar{s}\} - \sum_{\bar{s} \leq S} P\{N = \bar{s}\} \\
&= e^{-\lambda L} \frac{(\lambda L)^{S+1}}{(S+1)!} && (2.3.16) \\
\triangle^2 F(S) &= e^{-\lambda L} \frac{(\lambda L)^{S+1}}{(S+1)!} - e^{-\lambda L} \frac{(\lambda L)^S}{S!} \\
&= e^{-\lambda L} \frac{(\lambda L)^S}{S!} \left\{ \frac{\lambda L}{S+1} - 1 \right\} && (2.3.17)
\end{aligned}$$

$\lambda L > S+1$ implies that $\triangle^2 F(S) > 0$ and further implies that $F(S)$ is convex whenever $\lambda L > S+1$. In fact, $F(S)$ is said to be discretely convex whenever $S < \lambda L - 1$. Thus $F(S)$ is said to be discretely concave only if $S \geq \lfloor \lambda L \rfloor$, when λL is non integer, and $S \geq \lambda L - 1$, for integer values of λL.

Next, we immediately observe that, R(S), is not also concave for all feasible values of S.

Hence, neither $R(S)$ nor $F(S)$ have the desirable feature of concavity for all $S \geq 0$. Thus, in many cases, S usually is constrained to values that are equal to or greater than $\lfloor \lambda L \rfloor$ to guarantee that ready rate or fill rate functions are indeed concave for the feasible region.

2.3.9 Mathematical properties of backorder rate

The backorder level $B(S)$ has very attractive mathematical properties.

$$B(S) = \sum_{\bar{s}>S} (\bar{s} - S)P\{N = \bar{s}\}$$

If $B(S)$ were strictly discretely convex in addition to being strictly decreasing, then

$$\triangle B(S) = B(S+1) - B(S) < 0$$

$$\triangle B(S+1) = B(S+2) - B(S+1)$$

and

$$\begin{aligned} \triangle^2 B(S) &= \triangle B(S+1) - \triangle B(S) > 0 \\ &= B(S+2) - B(S+1) - (B(S+1) - B(S)) \\ &= B(S+2) - 2B(S+1) + B(S) \end{aligned}$$

Observe that

$$\begin{aligned} \triangle B(S) &= \sum_{\bar{s}>S+1} (\bar{s} - (S+1))P\{N = \bar{s}\} - \sum_{\bar{s}>S} (\bar{s} - S)P\{N = \bar{s}\} \\ &= \sum_{\bar{s}\geq S+1} (\bar{s} - (S+1))P\{N = \bar{s}\} - \sum_{\bar{s}\geq S+1} (\bar{s} - S)P\{N = \bar{s}\} \\ &= -\sum_{\bar{s}\geq S+1} P\{N = \bar{s}\} \\ &= -(1 - \sum_{\bar{s}\leq S} P\{N = \bar{s}\}) \end{aligned} \tag{2.3.18}$$

$$
\begin{aligned}
\triangle B(S+1) &= \sum_{\bar{s}>S+2} (\bar{s}-(S+2))P\{N=\bar{s}\} - \sum_{\bar{s}>S+1} (\bar{s}-S+1)P\{N=\bar{s}\} \\
&= \sum_{\bar{s}\geq S+2} (\bar{s}-(S+2))P\{N=\bar{s}\} - \sum_{\bar{s}\geq S+2} (\bar{s}-S+1)P\{N=\bar{s}\} \\
&= -\sum_{\bar{s}\geq S+2} P\{N=\bar{s}\} \\
&= -(1-\sum_{\bar{s}\leq S+1} P\{N=\bar{s}\}) \qquad (2.3.19)
\end{aligned}
$$

and

$$
\begin{aligned}
\triangle^2 B(S) &= \triangle B(S+1) - \triangle B(S) \qquad (2.3.20) \\
&= -\sum_{\bar{s}\geq S+2} P\{N=\bar{s}\} + \sum_{\bar{s}\geq S+1} P\{N=\bar{s}\} \\
&= -\sum_{\bar{s}\geq S+2} P\{N=\bar{s}\} + \sum_{\bar{s}\geq S+2} P\{N=\bar{s}\} + P\{N=S+1\} \\
&= P\{N=S+1\} > 0 \qquad (2.3.21)
\end{aligned}
$$

Thus $B(S)$ is strictly (discretely) convex $\forall\ S \geq 0$.

In this study, we are interested in improving customer service level, hence our model has a service constraint that ensures that the customer waiting time does not exceed a given threshold. The advantage of using the backorder measure is that unlike the fill rate and ready rate, it captures all customer orders. Caglar *et al.* (2004), Mak and Shen (2009), Riaz (2013) also considered minimising backorders to improve service using a response time requirement. The two echelon structure in this study is similar to the two echelon structure found in Caglar *et al.* (2004), Mak and Shen (2009) and Riaz (2013). The structure comprises of a plant at the top echelon, multiple SVCs at the lower echelon and geographically spaced customers. It is important to note that LT has not been incorporated into the systems considered by Caglar *et al.* (2004), Mak and Shen (2009) and Riaz (2013). This is the major contribution of this study.

2.3.10 Determining inventory levels for two-echelon systems

Graves (1985) developed an exact inventory level distribution for a two-echelon system. However, this exact model happened to be computationally burdensome for real life problems with many variables and parameters. Hence, most authors use approximations to ease computational burden.

A very useful approximation method is the multi echelon technique for repairable item control (METRIC) Shebrooke (1968). It applies Palm's theorem (Palm (1938)) and approximates the distribution for inventory level and backorder level using a Poisson distribution with corresponding mean. METRIC assumes that successive lead times from the plant to SVC are independent. Whereas in reality, these successive lead times depend on the situation of inventory at the plant. Axsater (1990) noted that, METRIC approximation will perform well when each SVC demand is low compared to the total system demand. Let B_0 denote the expected plant backorder level and λ_0 the total plant demand. By Little's law, Little (1961) $Wt_0 = \frac{B_0}{\lambda_0}$, where Wt_0 is the average waiting time at the plant. From METRIC,

$$B_0 = \sum_{\bar{s}=S_0+1}^{\infty} (\bar{s} - S_0) P\{N_0 = \bar{s}\} \tag{2.3.22}$$

where

$$P\{N_0 = \bar{s}\} = e^{-\lambda_0 Wt_0} \frac{(\lambda_0 Wt_0)^{\bar{s}}}{\bar{s}!}$$

The splitting of Poisson processes imply that the plant backorder level can be split to determine backorders at the SVC. Let λ_v, S_v, and B_v represent the demand at SVC v, the stocking level for SVC v and the expected backorder level at SVC v, respectively.

$$B_v = \frac{\lambda_v}{\lambda_0}(B_0) \tag{2.3.23}$$

Caglar *et al.* (2004) considered a two-echelon inventory problem with service constraints. They utilised METRIC approximation to obtain an optimal policy for such a system and showed that

$$B_v(S_v) = I_v - S_v + E[N_v] \tag{2.3.24}$$

Mak and Shen (2009) and Riaz (2013) considered two-echelon inventory location problems. They also utilised the metric approximation to determine their inventory level.

For items controlled with a (S-1,S) policy, setting stock levels depends on given objectives and corresponding constraints.

Everett's theorem (Everett (1965))

Given an optimisation problem

$$\begin{aligned} &\min f(y) \\ &\text{subject to } g(y) \leq b \end{aligned} \tag{2.3.25}$$

where $y \in \mathbb{R}^n$ is the optimisation decision variable, $f : \mathbb{R}^n \to \mathbb{R}$ is known as the objective function or cost function. The inequality $g(y) \leq b$ is the inequality constraint, while the function $g : \mathbb{R}^n \to \mathbb{R}$ is the inequality constraint function. The constraint $g(y) \leq b$ is relaxed. Assume f and g are convex. The relaxation of Problem (2.3.25) is

$$\min_y \; [f(y) + \theta(g(y) - b)] \tag{2.3.26}$$

for $\theta \geq 0$. θ is the Lagrange multiplier linked with $g(y) \leq b$. The relationship between solutions to problem (2.3.26) and problem (2.3.25) is given by the theorem below

Theorem 2.3.10.1(Everett (1965)).

Suppose $y^0(\theta)$ solves the Problem (2.3.26) optimally with θ as the Lagrange multiplier. Let $b' = g(y^0(\theta))$. Then $y^0(\theta)$ also solves

$$\begin{aligned} &\min_y f(y) \\ &\text{subject to } g(y) \leq b' \end{aligned} \tag{2.3.27}$$

Thus, for various values of θ, we can determine optimal solutions for problems such as (2.3.27). If $b' = b$ for specific Lagrange multiplier θ, we then say that problem (2.3.25) has been solved.

2.3.11 Optimal inventory position using the (S-1,S) policy

Following ideas from Shebrooke (1968), Caglar *et al.* (2004) and Muckstadt (2005), we construct an example that illustrates how to determine the optimal inventory position S, when following the $(S-1,S)$ policy.

Suppose a company manages an item at multiple service centers in a two-echelon setting. The inventory policy for all locations is the $(S-1,S)$ policy. Hence, an order is placed for replenishment on an external source whenever fulfilling a customer's demand depletes the SVC's stock by one. The target is to choose stock levels that minimise expected number of all outstanding backorders under the constraint of investment in inventory.

Let b denote the budget threshold on expected value of inventory on hand; V is the set of SVC locations, h_v is the unit holding cost for the item in SVC v, S_v represents the item stock level for SVC v, $E[N_v]$ is the steady state expected quantity in replenishment to SVC v, and $B_v(S_v)$ is the expected value of outstanding backorders at a random time for SVC v.

We state the optimisation problem as

$$\min \textstyle\sum_{v\in V} B_v(S_v)$$

subject to

$$\textstyle\sum_{v\in V} h_v[S_v - E[N_v] + B_v(S_v)] \leq b, S_v = 0, 1, \cdots \qquad (2.3.28)$$

To solve problem (2.3.28), we utilise the method of Lagrangian relaxation. We fix the inventory position at the plant and let θ_v denote the multiplier linked to the budget constraint. The relaxed problem is

$$\min \sum_{v\in V} B_v(S_v) + \sum_{v\in V} (\theta_v[h_v(S_v - E[N_v] + B_v(S_v)) - b]) \qquad (2.3.29)$$

subject to $S_v = 0, 1, 2, \cdots$

$$= \min_{S_v=0,1,\cdots} \sum_{v\in V}[(1+\theta_v h_v)B_v(S_v) + \theta_v h_v S_v] - [\theta_v \sum_{v\in V} h_v E[N_v] + \theta_v b]$$
$$= -\sum_{v\in V}\theta_v[h_v E[N_v] + b] + \sum_{v\in V}\min_{S_v=0,1,\cdots}[(1+\theta_v h_v)B_v(S_v) + \theta_v h_v S_v]$$

Thus, the multiplier θ_v, causes the resulting relaxed optimisation problem to be separable by location. The problem that needs be solved has the same form for each location, so we temporarily drop the subscript for location. Hence, the optimisation problem is reduced to

$$-\theta[hE[N] + b] + \min_{S=0,1,\cdots}[(1+\theta h)B(S) + \theta hS] \tag{2.3.30}$$

For given values of θ , $\theta[hE[N] + b]$ can be treated as a constant. Thus the problem is further reduced to

$$\min_{S=0,1,\cdots}(1+\theta h)B(S) + \theta hS \tag{2.3.31}$$

Let

$$f(S) = (1+\theta h)B(S) + \theta hS$$

$f(S)$ is convex because B(S) has been shown to be discretely strictly convex. Define

$$\triangle f(S) = f(S+1) - f(S)$$
$$= (1+\theta h)\{B(S+1) - B(S)\} + \theta h$$

Since previously we established that

$$B(S+1) - B(S) = -(1 - \sum_{\bar{s}\leq S} P\{N = s\})$$
$$\triangle f(S) = -(1+\theta h)(1 - \sum_{\bar{s}\leq S} P\{N = \bar{s}\}) + \theta h \tag{2.3.32}$$

Since $f(S)$ is convex, our optimal stock level,for a given θ, will be smallest nonnegative integer, S^*, that gives

$$\triangle f(S) \geq 0$$

which is, the minimum value that gives

$$(1 + \theta h)(1 - \sum_{\bar{s} \leq S} P\{N = \bar{s}\}) \leq \theta h \tag{2.3.33}$$

or

$$\sum_{\bar{s} \leq S} P\{N = \bar{s}\} \geq \frac{1}{1 + \theta h} \tag{2.3.34}$$

S^* clearly depends on θ. Let

$$C(\theta') = \sum_{v \in V} h_v[S_v(\theta_v) - E[N_v] + B_v(S_v(\theta_v))] \tag{2.3.35}$$

The aim is to determine a value of θ_v for each v such that $C(\theta')$ is equal to b approximately. For each v, each value of θ gives a set of feasible stock levels, inventory cost, and minimum number of expected outstanding backorders.

Suppose for each location v, we have $\theta_1 > \theta_2 > \cdots > \theta_M$. Since $\frac{1}{1+\theta_1 h} < \frac{1}{1+\theta_2 h} < \cdots < \frac{1}{1+\theta_M h}$, $S^*(\theta_1) \leq S^*(\theta_2) \leq \cdots \leq S^*(\theta_m)$. To determine $S^*(\theta)$, find the smallest non negative integer value of S which gives

$$\sum_{\bar{s} \leq S} P\{N = \bar{s}\} \geq \frac{1}{1 + \theta h} \tag{2.3.36}$$

Hence to find $S^*(\theta_i)$, $i \in [1, M]$, we utilise

$$\sum_{\bar{s} \leq S^*(\theta_i - 1)} P\{N = \bar{s}\}$$

as an initial point for our computation. Since this value has already been computed for the determination of $S^*(\theta_i - 1)$, the computational effort needed to find $S^*(\theta_i)$ might be significantly reduced.

There exists a $\theta > 0$ such that

$$P\{N = 0\} = \frac{1}{1 + \theta h}$$

, or

$$\theta = \frac{1}{h}\left(\frac{1}{P\{N=0\}} - 1\right)$$

Let $\theta_{max} = \frac{1}{h}\left(\frac{1}{P\{N=0\}} - 1\right)$. If $\theta = \theta_{max}$, then $S^*(\theta_{max}) = 0 \; \forall$ v.

2.3.12 Determining inventory levels with lateral transshipment

Lateral Transshipment (LT) is the movement of inventory between facilities on the same echelon. LT is permitted to occur only among facilities in same pool. A pool is a collection of SVCs that satisfy a pooling criterion. A pooling criterion could be based on distance, geographical location, storage capacity, etc. In study, our pooling criterion is geographical and is such that the collection of all SVCs in a geopolitical zone forms a pool.

Let W be the set of pools, we add the pool subscript to our previous notations and also introduce some new notations

$\bar{I}_{vw}$ = Probability of satisfying demand at SVC v in pool w from on-hand stock at the SVC

T_{vw} is the steady state expected lateral transshipment level at SVC v in pool w.

$\bar{T}_{vw}$ = Probability of satisfying demand at SVC v in pool w by LT from another SVC in the same pool

$\bar{B}_{vw}$ = Probability of backordering demand at SVC v in pool w

$\bar{B}_{w}$ = Probability of backordering demand at pool w

Demand in this system must be fulfilled either from on-hand stock, lateral resupply or backorder. Thus

$$\bar{I}_{vw} + \bar{T}_{vw} + \bar{B}_{vw} = 1 \qquad (2.3.37)$$

It is obvious that if $S_{vw} = 0$, then $\bar{I}_{vw} = 0$

To show modelling ideas, we assume that there are $\bar{w}$ pools and all the service centers in same pool are identical, Lee (1987) and Muckstadt (2005). Hence we assume that SVC demand processes at all SVCs in pool w are Poisson processes with identical demand rate, λ_{vw}, and have same stock levels S_{vw}.

Also $\bar{I}_{vw}$, $\bar{T}_{vw}$, and $\bar{B}_{vw}$ are identical for all service centers. We also assume that plant to SVC replenishment time for SVCs in the pool are idd exponential random variables with mean L . $\bar{I}_{vw}$ also gives the fraction of time a SVC has positive stock. Thus, the fraction of time the SVC stock is zero or negative is

$$1 - \bar{I}_{vw} = \bar{T}_{vw} + \bar{B}_{vw}$$

There are $\hat{w}$ SVCs in pool w, thus, pool w's demand process is a Poisson process and it's rate is $\lambda_w = \hat{w}\lambda_{vw}$. Furthermore, the total plant demand processes are Poisson processes with rate $\lambda_0 = \sum_w \hat{w}\lambda_w$.

Let N_{vw} and N_w be the steady state number of items in replenishment to service center v in pool w and pool w, respectively. Lee (1987) showed that the $\bar{I}_{vw}$ is approximated by

$$\bar{I}_{vw} = \sum_{\bar{s}=0}^{S_{vw}-1} P[N_{vw} = \bar{s}] \tag{2.3.38}$$

since $S_{vw} = S$ for all service centers in pool w.

The probability that a pool's demand arrival will not be satisfied from on-hand pool stock is the probability the pool's total number of units on order is equal to or greater than the total pool stock. However, this probability is $P[N_w > \hat{w}S]$. Therefore $\bar{B}_w = P[N_w > \hat{w}S]$, because we assumed instantaneous transshipment between SVCs within same pool. Since $\bar{B}_{vw}$ and $\bar{I}_{vw}$ can be determined, we can also determine $\bar{T}_{vw}$ for a pool. We have $\bar{T}_{vw} = 1 - (\bar{I}_{vw} + \bar{B}_{vw})$. According to Lee (1987), expected value of lateral transshipments corresponding to a SVC per unit time (T_{vw}) is $\lambda_{vw}\bar{T}_{vw}$ and $\hat{w}\lambda_{vw}\bar{T}_{vw}$ for the entire pool. Lee showed that the total expected number of backorders at pool w is

$$B_w = \sum_{\bar{s}=\hat{w}S_{vw}+1}^{\infty} (\bar{s} - \hat{w}S_{vw})P\{N_w = \bar{s}\} + \hat{w}\lambda_{vw}\bar{T}_{vw} \tag{2.3.39}$$

and the backorder level at the plant is given by

$$B_0 = \sum_{\bar{s}=S_0+1}^{\infty} (\bar{s} - S_0)P\{N_0 = \bar{s}\} \tag{2.3.40}$$

The first term of equation (2.3.39) gives the expected backorder quantity at pool w which has not been satisfied by the plant. The second term gives the expected value of items

in resupply for lateral transshipments. Lee (1987) showed that these approximations are accurate for high service levels.
For this system we could choose to minimise the total costs which comprises of plant holding cost, SVC holding cost and SVC backorder cost.

$$C_{total} = h(S_0 + \sum_{w \in W} \hat{w} S_{vw}) + \bar{p} \sum_{w \in W} B_w + q \sum_{w \in W} T_w \tag{2.3.41}$$

where h is the per unit holding cost , p is the per unit backorder cost, and q is the LT cost per unit item. The first, second and third terms represent all inventory holding costs, all backorder costs and all LT costs respectively for the system.

2.3.13 Facility location model

The facility location model seeks to minimise set-up costs for SVCs or facilities, and transportation costs from SVC to customers, while determining the facilities to open and optimal customer assignment to opened facilities, Daskin (1995). The model considered here is called the uncapacitated fixed charge location model. Some very important features of this facility location problem considered are: each facility has a fixed opening cost, the facilities are assumed to be have infinite capacity, a customer can be assigned to one and only one SVC. We remove the subscript for pool in this formulation. Daskin (1995) gives a detailed treatment of facility location models. The facility location component of our model is similar to the model presented below, (Daskin (1995) p.250).

$$\min \sum_{v \in V} \left(f_v X_v + \sum_{u \in U} \lambda_u Y_{uv} d_{uv} \right) \tag{2.3.42}$$

Subject to:

$$\textstyle\sum_{v \in V} Y_{uv} = 1, \text{for each } ,u \in U \tag{2.3.43}$$

$$Y_{uv} \leq X_v, \text{ for each,} u \in U, v \in V \tag{2.3.44}$$

$$X_v \in \{0, 1\} \text{for each,} v \in V \tag{2.3.45}$$

$$Y_{uv} \in \{0, 1\} \text{for each,} u \in U \tag{2.3.46}$$

where, f_v, λ_u and d_{uv} are set up costs for SVC v, demand from customer u, and cost of transportation from customer u to SVC v, respectively. $X_v = 1$ if a SVC is set up at location v, 0 otherwise, and $Y_{uv} = 1$ if customer u is assigned to SVC v, 0 otherwise.

The objective (2.3.42) is to find the minimum sum of the fixed location costs and transportation costs. The assignment constraint (2.3.43) states that each customer's demand should be assigned to one and only one SVC. Constraints (2.3.44) require that demand assignments can be made only to a candidate location that has an open SVC and for which the resulting distance from customer is less than d_{max}. Finally, (2.3.45), and (2.3.46) are constraints on nonnegativity and integrality.

This problem is solved by making use of heuristic algorithms or Lagrangian method. The heuristic algorithms consist of the ADD and the DROP algorithms. The ADD algorithm follows a greedy procedure to add facilities to the solution until the addition of a facility decreases the cost no further. While the DROP algorithm opens all candidate facilities and the greedy procedure proceeds to drop facilities from the solution till dropping a facility can no longer decrease the cost. In this study, we explored our models' properties using the Lagrangian approach, so we lay emphasis on the Lagrange relaxation method for the uncapacitated fixed charge facility location problem.

To begin the Lagrange procedure, constraint (2.3.43) is relaxed and the corresponding Lagrange multiplier is π_u. The following Lagrange dual problem is obtained

$$\max_{\pi} \min_{X,Y} \sum_{v \in V} f_v X_v + \sum_{v \in V} \sum_{u \in U} (\lambda_u d_{uv} - \pi_u) Y_{uv} + \sum_{u \in U} \pi_u \tag{2.3.47}$$

Subject to:

$$Y_{uv} \leq X_v, \text{ for each,} u \in U, v \in V \tag{2.3.48}$$

$$X_{vw} \in \{0,1\} \text{for each,} v \in V \tag{2.3.49}$$

$$Y_{uv} \in \{0,1\} \text{for each,} u \in U \tag{2.3.50}$$

Solving the problem consists of three steps which are: determining the solution of the Lagrange relaxed problem for fixed multiplier values π_u, conversion of the relaxed solution to a primal feasible solution, and then updating the Lagrange multipliers.

The starting point is to minimise (2.3.47) for fixed values of π_u. if $\lambda_u d_{uv} - \pi_u \geq 0$, Y_{uv} can be set to 0. If $\lambda_u d_{uv} - \pi_u < 0$, Y_{uv} can be set to the maximum feasible value. Recall that $Y_{uv} \leq X_v$. Now we compute $A_v = f_v + \sum_{u \in U} \min(0, \lambda_u d_{uv} - \pi_u)$. We set

$X_v = 1$, the Lagrangian objective will change by A_v for each v. If $A_v < 0$ setting $X_v = 1$ will cause cost decrease; else we set $X_v = 0$.

Thus for given values of π we can find optimal values for the facility location variables X_v and allocation variables Y_{uv} using the following two-step algorithm:

1. For each candidate facility, compute $A_v = f_v + \sum_{u \in U} \min(0, \lambda_u d_{uv} - \pi_u)$.
   Set
   $$X_v \begin{cases} 1, & \text{if } A_v < 0 \\ 0, & \text{otherwise.} \end{cases}$$

2. Set
   $$Y_{vv} \begin{cases} 1, & \text{if } X_v = 1 \text{ and } \lambda_u d_{uv} - \pi_u < 0 \\ 0, & \text{otherwise.} \end{cases}$$

Using this two step algorithm, the evaluation of (2.3.47) for any values of π_u will give a lower bound to the problem (2.3.42). The subgradient optimisation procedure is used to find the values of π_u that maximise this bound.

The solution derived with the algorithm presented above may not satisfy some of the constraints relaxed. Particularly, it is likely some demand may not be assigned $(Y_{uv} = 0)$ and some others are assigned to two or more facilities $(\sum_{v \in V} Y_{uv} \geq 2)$. However, we can find a primal feasible solution by locating facilities at locations for which $X_v = 1$ and assigning demands to the nearest facility opened. The primal objective function obtained from this set of locations and demand allocations will provide us with the solution's upper bound. Clearly, the smallest of such values over all iterations of the Lagrangian is the best solution to use.

Daskin (1995) showed that of the Add algorithm, Drop algorithm and the Lagrange method, only the Lagrange method gave optimal solutions.

2.3.14 Two-echelon inventory model with service consideration

The two-echelon inventory with service constraints (response time requirement) is presented. Caglar *et al.* (2004) considered this model. The model presented does not have lateral transshipments. Also the subscript w is dropped because there is no pool. The model has a plant at the top echelon and a finite number of SVCs.

$$\min \sum_{v \in V} h_v I_v + h_0 I_0 \tag{2.3.51}$$

Subject to:

$$0 \leq S_v \leq C_v, \text{ for each}, v \in V \tag{2.3.52}$$

$$0 \leq S_0 \leq C_0 \tag{2.3.53}$$

$$Wt_v \leq \tau, \text{for each}, v \in V \tag{2.3.54}$$

where, Wt_v is the response time at a SVC. The objective (2.3.51) is to find the minimum sum of the plant inventory holding costs and inventory holding costs at all SVCs. Backorders at the plant are considered internal to the system hence they do not attract a monetary cost. Constraints (2.3.52) and (2.3.53) state that the SVCs, pools and plant stock levels cannot exceed the storage capacity available. The service (response) time constraints (2.3.54) require that expected response time must not be more than the required level.

Using Little's law $Wt_v = \frac{B_v}{\sum_{u \in U} \lambda_u Y_{uv}}$, (2.3.54) can be written as $\frac{B_v}{\sum_{u \in U} \lambda_u Y_{uv}} \leq \tau$ or $B_v \leq \tau \sum_{u \in U} \lambda_u Y_{uv}$. Caglar *et al.* (2004) showed that the plant backorder level B_0 and plant inventory level I_0 are

$$\begin{aligned}
B_0 &= E[N_0] - \sum_{s=o}^{S_0-1} [1 - F_0(s)] \\
I_0 &= S_0 - E[N_0] + B_0 \\
&= S_0 - E[N_0] + E[N_0] - \sum_{s=o}^{S_0-1} [1 - F_0(s)] \\
&= \sum_{s=o}^{S_0-1} F_0(s)
\end{aligned}$$

Caglar *et al.* (2004) also obtained the facility backorder level B_v and facility inventory level I_v to be:

$$B_v = E[N_v] - \sum_{s=o}^{S_v-1} [1 - F_v(s)]$$

$$I_v = S_v - E[N_v] + B_v$$

The objective function of this problem can be rewritten as

$$\min \sum_{v \in V} h_v[S_v - E[N_v] + B_v] + h_0[I_0 - \sum_{v \in V} E[N_v]]$$

Thus the problem can be rewritten as

$$\min \sum_{v \in V} h_v[S_v - E[N_v] + B_v] + h_0[I_0 - \sum_{v \in V} E[N_v]] \tag{2.3.55}$$

Subject to:

$$0 \le S_v \le C_v, \text{ for each,} v \in V \tag{2.3.56}$$

$$0 \le S_0 \le C_0 \tag{2.3.57}$$

$$B_v \le \tau \textstyle\sum_{u \in U} \lambda_u Y_{uv}, \text{for each,} v \in V \tag{2.3.58}$$

Caglar *et al.* (2004) utilised a Lagrange relaxation algorithm to solve the problem. The constraint (2.3.58) is relaxed to get the following relaxed problem with Lagrange multiplier $\theta_v > 0$

$$\sum_{v \in V} h_v[S_v + B_v] + h_0[I_0 - \sum_{v \in V} E[N_v]] + \sum_{v \in V} \theta_v[B_v - \tau \sum_{u \in U} \lambda_u Y_{uv}] \tag{2.3.59}$$

Subject to:

$$0 \le S_v \le C_v, \text{ for each,} v \in V \tag{2.3.60}$$

$$0 \le S_0 \le C_0 \tag{2.3.61}$$

A lower bound on (2.3.55) is obtained by solving (2.3.59). Caglar *et al.* (2004) solved (2.3.59) by enumerating all feasible values of S_0 and S_v.

The work by Caglar *et al.* (2004) is useful to this research because our model has similar two echelon structure with service considerations. It is important to state that the incorporation of LT into the model of Caglar *et al.* (2004) has not been considered. This presents the major difference between our work and theirs.

2.3.15 Two-echelon inventory-location model

This model was considered by Mak and Shen (2009). They merge ideas from the uncapacitated fixed charge facility location problem and the model by Caglar *et al.* (2004). The model is as presented:

$$\min \sum_{v\in V}\left(f_vX_v + h_vI_v + p_vB_v + \sum_{u\in U}\lambda_uY_{uv}d_{uv}\right) + h_0I_0 \tag{2.3.62}$$

Subject to:

$$\textstyle\sum_{v\in V} Y_{uv} = 1, \text{for each } ,u \in U \tag{2.3.63}$$

$$Y_{uv} \leq a_{uv}X_v, \text{ for each,} u \in U, v \in V \tag{2.3.64}$$

$$S_v \leq C_v, \text{ for each,} v \in V \tag{2.3.65}$$

$$S_0 \leq C_0 \tag{2.3.66}$$

$$Wt_v \leq \tau, \text{for each,} v \in V \tag{2.3.67}$$

$$S_v \geq 0, integer, \text{for each,} v \in V \tag{2.3.68}$$

$$S_0 \geq 0 \tag{2.3.69}$$

$$X_v \in \{0,1\}\text{for each,} v \in V \tag{2.3.70}$$

$$Y_{uv} \in \{0,1\}\text{for each,} u \in U \tag{2.3.71}$$

The objective (2.3.62) is to find the minimum sum of the fixed location costs, plant inventory holding costs, SVC inventory holding costs, backorder costs at SVCs, lateral transshipment costs at SVCs and transportation costs. Backorders at the plant are considered internal to the system hence they do not attract a monetary cost. Constraint (2.3.63) states that assignment of all demand from a customer should be made to one and only one SVC. The constraints (2.3.64) require that assignment of demand to any candidate location must not be initiated unless it is open and its distance from customer is less than d_{max}. The constraints (2.3.65) and (2.3.66) state that the SVC and plant base stock levels cannot be greater than the capacity available for storage. The response time requirement (2.3.67) ensure that expected response time must not exceed the required level. Finally, (2.3.68), (2.3.69),(2.3.70),and (2.3.71) are nonnegativity and integrality constraints.

This model does not consider lateral transshipment and the effect of pooling; this is the major difference between the model by Mak and Shen (2009) and our model. Mak and Shen (2009) utilised the results obtained by Caglar *et al.* (2004) for steady state expected inventory and backorder levels at SVCs and at the plant .

$$
\begin{aligned}
B_0 = E[N_0] - \sum_{s=o}^{S_0-1}[1 - F_0(s)] \\
I_0 = S_0 - E[N_0] + B_0 \\
B_v = E[N_v] - \sum_{s=o}^{S_v-1}[1 - F_v(s)] \\
I_v = S_v - E[N_v] + B_v
\end{aligned}
$$

They utilised a Lagrangian based algorithm to solve the problem. Riaz (2013) considered a similar problem to that of Mak and Shen (2009). The slight variation is that customer assignment was made based on customer's preference. It is important to state that the incorporation of LT into the model of Mak and Shen (2009) has not been considered. This presents the major difference between our work and theirs.

Chapter 3
METHODOLOGY

3.0 Introduction

In this chapter, the incorporation of Lateral Transshipments (LTs) into two-echelon systems with response time requirement and the objective of cost minimisation is considered. Firstly, a general description of the system is presented after which different model formulations depicting the system in different settings are presented. The different model formulations presented in this chapter are: two-echelon inventory model with Response Time Requirement (RTR) and Lateral Transshipment (LT), joint inventory location two-echelon inventory model with RTR and LT, model with reliable locations and model with stochastic demand. For each model, the steady state distribution of orders in replenishment and the expected levels in steady state for on-hand inventory, LT and backorder are presented. Furthermore, the system is decomposed and some properties of the models are highlighted.

3.1 Mathematical model description

3.1.1 Background

Allowable transition of inventory among locations in same level of an inventory system is known as LT. The use of LT has not been studied for two-echelon systems that jointly consider facility and inventory decisions with response time requirements. For a system that permits LT, in a case of stock-out at one facility, a demand can be satisfied by means of a stock transfer from another facility. The challenge that comes with LTs lies in deciding where and when a stock movement is beneficial. LTs may result in a reduction in the immediate shortage risk at the location receiving transshipment, but it increases the future risk at the location sending transshipment. Therefore, LT policies should seek to balance these conflicting risks and determine when transshipment cost is dominated by the expected transshipment benefit. The suitability of a particular LT policy often depends on the characteristics of the inventory system in which it is used.

Also it is common practice to see companies manage decisions on storage and distribution independently, this is partly as a result of the complexities that arise from combining them. In addition, tactical decisions involving stocking and operational decisions involving distribution are usually considered independently of the strategic decisions involving facility location and network design. It has been shown that making facility location decisions independent of inventory consideration can lead to supply chain designs that are suboptimal, Daskin *et al.* (2002), Shen *et al.* (2003), Candas and Kutanoglu (2007) and Mak and Shen (2009). Nonetheless, the fusion of facility location,inventory management and distribution decisions still remains a complex mathematical modelling task, especially with the consideration of response time requirements. Attempting to incorporate LT into such joint two-echelon systems will result in much more complex models.

In this thesis, we incorporate LT into a two-echelon service parts supply chain with response time requirement. We require that the response time for a customer order must not exceed a given threshold.

3.1.2 System description and notations

The supply chain system considered in this study comprises of a plant at the upper echelon, a set of possible SVC locations at the lower echelon in addition to a set of customers (demand nodes). We assume that each customer's demand node is also a possible SVC location, thus, the set of possible SVC locations V and the set of customers U are equivalent sets. This two echelon service parts system operates in the following manner:

1. The items are manufactured and held at the plant to satisfy Service Centre (SVC) demands. The plant resupplies the SVC within a SVC specific replenishment lead time. The items in each SVC are identical.

2. The SVCs keep inventory to satisfy orders from customers. Each customer is assigned to exactly one SVC, and customers' order process at their assigned SVC is Poisson. The assignment of a customer depends on a customer's distance to a SVC and is taken care of by an assignment constraint in our model formulation. We do not consider the case of a customer's preference of one SVC to another. Orders

placed by different customers are independent, hence the demand processes at the various SVCs are independent Poisson processes.

3. The plant and all SVCs have limited capacity for holding inventory and a continuous review $(S-1, S)$ policy is used to manage replenishment at the SVC. These items are characteristically costly for numerous service parts systems, capacity constraints can be interpreted as budget constraints.

4. When a customer order arrives at the SVC, the SVC sends a single unit of the item from its inventory on hand to the customer (if there is no stockout) and immediately places a replenishment request with the plant.

5. If the SVC has zero or negative inventory level, the customer's request will be satisfied instantly by lateral transshipment from pooled neighbouring SVCs which have stock on hand.

6. If none of the pooled neighbouring SVCs have stock on hand, demand is backordered until they are satisfied.

7. Lateral transshipment is assumed to be instantaneous

8. When a SVC replenishment request arrives the plant, the plant sends one unit of the item from its inventory on hand to the SVC (if there is no stockout), and instantly places an order for a single unit to be produced.

9. If the plant inventory level is zero, the demand at the plant is backordered till they are satisfied.

10. Finished goods produced by the plant are used to satisfy backorders or are stored as plant inventory. It is assumed that the plant possesses a single production line with exponential service rate. The plant's demand process is Poisson since, each SVC faces an aggregation of Poisson customer demand arrivals and also places orders for replenishment in a one-to-one manner. This then means that the plant production line has the properties of a Markovian queue.

11. All demands at the SVCs, replenishment orders at the plant, transshipment and backorder requests are processed in a First-Come,First-Served $(FCFS)$ manner.

12. Demand arrivals at the SVC are fulfilled through any one of the following: stock on hand, lateral transshipment or backorder.

We aim to simultaneously determine the optimal number of SVCs, customers assignment to opened SVCs, inventory levels at SVCs, the lateral transshipment levels at the SVCs and inventory levels at the plant.

The costs involved in this study are

1. The cost incurred as a result of opening the SVC known as fixed location cost.
2. The costs incurred as a result of keeping on hand inventory at the SVC and at the plant. This is known as holding costs.
3. The costs incurred when demand at a SVC is backordered. This is known as backorder costs.
4. Lateral transshipment costs at SVCs (cost incurred when demand is met via lateral transshipment).

To begin, we introduce the following notations:

Sets

U represents the set of customers. We considered three data sets comprising of 37 nodes, 109 nodes and 181 nodes. The number of nodes represents the size of U for each of the three data sets.

V represents the set of candidate SVC locations. The geographical location of each node is considered a possible location for setting up a SVC. This is taken care of by the location variable defined below. Consequently, the number of nodes represents the size of V for each of the three data sets.

W represents the set of pools. In this study, we used a geographical pooling criterion. All SVCs in a geopolitical zone form a pool

Parameters

f_{vw} is the fixed opening cost for SVC v in pool w

h_{vw} is the per unit holding cost of each unit of inventory at SVC v in pool w per unit time

h_0 is the per unit holding cost of each unit of inventory at the plant per unit time

p_{vw} is the per unit cost of backorder per unit inventory for each unit of time

q_{vw} is the LT cost per unit inventory

$\bar{d}_{uvw}$ is the distance of customer u from SVC v in pool w. This is determined by the Haversine rule using the longitudes and latitudes of different nodes from the data sets.

d_{uvw} is the transportation cost from SVC v in pool w to customer u. This is obtained by multiplying $\bar{d}_{uvw}$ by 10^{-1}

λ_u is customer u's demand rate

λ_{vw} is the SVC demand rate of a SVC in pool w demand rate $= \sum_{u \in U} \lambda_u Y_{uvw}$

λ_w is pool w's demand rate $= \sum_{v \in V} \lambda_{vw} = \sum_{v \in V} \sum_{u \in U} \lambda_u Y_{uvw}$

λ_0 is the plant demand rate $(= \sum_{w \in W} \lambda_w) = \sum_{w \in W} \sum_{v \in V} \sum_{u \in U} \lambda_u Y_{uvw}$

μ is the plant order processing rate

ρ is the plant utilisation rate $(= \frac{\lambda_0}{\mu})$

$\tau =$ response time requirement

α_w is the exact lead time from plant to pool w for any $w \in W$

d_{max} is the Maximum allowable distance between a customer and its assigned SVC

$a_{uvw} = 1$ if the distance from customer u to candidate SVC location v in pool w is not greater than d_{max}, 0 otherwise

C_{vw} is the capacity of SVC v in pool w, this is the same for all SVCs in pool w

$C_w = \hat{w} C_{vw}$ is the total space available for storage at pool w, where $\hat{w}$ is the number of SVCs in pool w

C_0 is the total space available for storage space at the plant

Binary decision variables

$X_{vw} = 1$ if a SVC is located at v in pool w, 0 otherwise

$Y_{uvw} = 1$ if customer u's demand is assigned to SVC v in pool w

Other decision variables

S_{vw} is the base stock level required at SVC v in pool w, this is uniform for all SVCs in pool w

$S_w = \hat{w} S_{vw}$ is the total base stock level required at pool w. $\hat{w}$ represents the number of SVCs in pool w.

S_0 is the base stock level required at the plant.

Service variables

I_{vw} is the expected level for on-hand inventory at SVC v in pool w in steady state

$I_w = \sum_{v\in V} I_{vw}$ is the expected level for on-hand inventory at pool w in steady state

B_{vw} is the expected level of backorder at SVC v in pool w in steady state

B_w is the expected level of backorder at pool w in steady state

T_{vw} is the expected LT level at SVC v in pool w in steady state. It describes the expected number of items from SVC v used to satisfy LT requests.

Wt_{vw} is the expected time of response at SVC v in pool w

I_0 is the expected plant inventory level in steady state

B_0 is the expected plant backorder in steady state

$N_{vw}(t)$ is the number of replenishment orders placed by SVC v in pool w which are yet to arrive at time t

N_{vw} is the steady state expected number of replenishment orders placed by SVC v in pool w which are yet to arrive

$N_w(t)$ is the number of replenishment orders placed by pool w which are yet to arrive at time t.

N_w is the steady state expected number of replenishment orders placed by pool w which are yet to arrive.

$N_0(t)$ is the number of replenishment orders placed by the plant that are yet to arrive by time t.

N_0 is the steady state expected number of replenishment orders placed by the plant that are yet to arrive by time.

3.1.3 Major assumptions

Two key assumptions are made concerning the model. Firstly, we assume that lateral transshipment times are negligible. The lead time is the time taken for orders placed at the plant to arrive at the SVC. The lead time comprises of the stochastic waiting time (delay) at the plant and the deterministic plant to SVC transportation time. The consequence of the assumption is that the deterministic plant to SVC transportation time is identical for all SVCs in a pool. Secondly, we assumed that SVC basestock level is identical for all SVCs in a pool.

3.2 Two-echelon inventory model with response time requirement and lateral transshipment (model I)

We now present a two-echelon inventory model with RTR and LT. To the best of our knowledge, researchers have not considered the incorporation of LT into two-echelon inventory systems with RTR, a plant at the top echelon, and a finite number of SVCs at the lower echelon. This model considers inventory decisions alone without location decisions. We use results from this model to formulate our other models. We call the model, Model I, and present it below.

$$\min \sum_{w\in W}\sum_{v\in V}(h_{vw}I_{vw} + p_{vw}B_{vw} + q_{vw}T_{vw}) + h_0 I_0 \tag{3.2.1}$$

Subject to:

$$0 \leq S_{vw} \leq C_{vw}, \text{ for each } v \in V \tag{3.2.2}$$

$$0 \leq S_w \leq \hat{w}C_{vw}, \text{ for each } w \in W \tag{3.2.3}$$

$$0 \leq S_0 \leq C_0 \tag{3.2.4}$$

$$Wt_{vw} \leq \tau, \text{ for each } v \in V \tag{3.2.5}$$

$$S_{vw} \geq 0, integer, \text{ for each } v \in V \tag{3.2.6}$$

$$S_0 \geq 0 \tag{3.2.7}$$

The objective (3.2.1) is to find minimum sum of the following: fixed location costs, holding costs for inventory at plant and SVCs, backorder costs at SVCs and LT costs at SVCs. Constraints (3.2.2), (3.2.3) and (3.2.4) state that base stock level for SVCs, pools and plant cannot be greater than the storage capacity available. The constraints on response time (3.2.5) require that expected response time be no greater than the required level. Finally, (3.2.6) and (3.2.7) are nonnegativity constraints.

Each SVC has the properties of a queuing system, Kruse (1981), in which customer orders can be regarded as the items in the system. Then the amount of items in the system awaiting service represents the backorder level. Also, the time spent by the item in the system is known as the response time. By using Little's law (Little, 1961), Kruse (1981) derived the waiting time and showed that at a SVC, the expected response is given by

$$Wt_{vw} = \frac{B_{vw}}{\lambda_{vw}} \tag{3.2.8}$$

Hence,

$$B_{vw} \leq \tau \lambda_{vw} \tag{3.2.9}$$

Two very important results for all models considered in this study are the result showing the relationship between on-hand inventory, LT and backorder in steady state and the result showing their steady state expected levels. We now present steady state expected levels for on-hand inventory, LT and backorder for Model I.

3.2.1 Plant inventory level for model I

In this system each customer's demand follow a Poisson process, hence the demand at each SVC being a merger of various independent Poisson processes also is a Poisson process (Tijms (2003) p. 6). Also, each SVC operates an order-up-to or (S-1,S) policy implying that a replenishment request is placed on the plant by a SVC immediately a demand occurs. Hence, the demand process at the plant is a merger of independent Poisson processes, thus, it is also a Poisson process. The plant possesses a single production line and with plant service rate being exponential, the plant possesses the characteristics of a queuing system with demands considered as arrivals to the system and items in replenishment are considered to be in service in the queuing system.

Graves (1985) developed an exact distribution for inventory levels in a two-echelon system. However, this exact model happened to be computationally burdensome for the large problems encountered in practice. Thus, most authors use approximations to reduce the computational burden. A very useful approximation method is the multi echelon technique for repairable item control (METRIC), Shebrooke (1968). It applies Palm's theorem (Palm (1938)) and approximates the distribution for inventory level and backorder level using a Poisson distribution with corresponding mean. METRIC assumes that successive lead times from the plant to SVCs are independent. These lead times actually depend on the situation of plant inventory situation, so they are not independent. Axsater (1990) showed that the METRIC approximation will generally perform well when the SVC demand is low compared to the total system. B_0 denotes the expected plant backorder level and λ_0 represents total plant demand. By Little's law , Little (1961) $Wt_0 = \frac{B_0}{\lambda_0}$, where Wt_0 is the average waiting time at the plant.

Following METRIC approximation, Caglar *et al.* (2004) established the standard expression for inventory and backorder expressions. We state their result and also give a new proof that follows from the model properties.

Proposition 3.2.1.1 Caglar *et al.* (2004)

In steady state the expected plant inventory level is given by

$$I_0 = S_0 - E[N_0] + B_0. \tag{3.2.10}$$

Proof. Demand arrival at SVCs follow a Poisson hence each SVC faces an aggregation of Poisson demand. Since each SVC is controlled with the (S-1,S) policy, the plant's arrival process is Poisson. Thus the plant exhibits the property of a queue with Markov arrival process. The steady state levels for a Markovian queue imply that steady state inflow is equal to steady state outflow and gives rise to the balance equation. The steady state expected number of items in transit from the plant's production line to its storage facility, that is steady state inflow to the plant's storage facility is denoted by $E[N_0]$. From definition $S_0 \geq I_0$ and $S_0 - I_0$ represents steady state expected number of plant demand satisfied from available inventory. While B_0 represents steady state expected number of plant demand satisfied from backorder. Therefore steady state expected outflow is $S_0 - I_0 + B_0$. Hence the balance equation of this system is

$$E[N_0] = S_0 - I_0 + B_0 \tag{3.2.11}$$

Thus

$$I_0 = S_0 - E[N_0] + B_0. \tag{3.2.12}$$

□

From (Caglar *et al.* (2004) p. 10)

$$B_0 = E[N_0] - \sum_{s=0}^{S_0-1} [1 - F_0(s)], \tag{3.2.13}$$

and

$$I_0 = \sum_{s=0}^{S_0-1} F_0(s) \tag{3.2.14}$$

where

$$F_0(s) = \sum_{m=0}^{s} P\{N_0 = m\}$$

.

The expected plant inventory level and plant backorder level can be obtained easily for different manufacturing queue systems via substituting steady state probability into the above formulas, Mak and Shen (2009). For example, we have the steady state probability for the number of customers in the $M/M/k$ queue as follows:

$$P\{N_0 = n\} = \begin{cases} \left(\sum_{m=0}^{k-1} \frac{(q\rho)^n}{m!} \frac{(k\rho)^k}{k!(1-\rho)}\right)^{-1}, & \text{if}, n = 0, \\ \frac{\lambda^n}{n!\mu^n} P\{N_0 = 0\}, & \text{if}, 1 \leq n \leq k \\ \frac{\lambda^n}{k^{n-k}k!\mu^n} P\{N_0 = 0\}, & \text{otherwise} \end{cases}$$

For our system, the plant arrival process is Poisson and the plant possesses a single production line. This in addition to queue discipline of first come first serve imply that our system behaves as the M/M/1 queue; that is a queue with Poisson arrivals, exponential service times and a single server. For this queue, the result by Buzacott and Shanthikumar (1993) gives the optimal policy at the plant, showing the expected plant inventory level and the expected plant backorder level. We state their result and give a well detailed proof which shows the behaviour of the system at the plant.

Proposition 3.2.1.2 (Buzacott and Shanthikumar (1993))

1. The steady state plant backorder level is given by $B_0 = \frac{\rho^{S_0+1}}{1-\rho}$

2. The steady state plant on hand inventory is given by $I_0 = S_0 - \frac{\rho}{1-\rho}(1 - \rho^{S_0})$

3. The expected plant response time is given by $W_0 = \frac{\rho^{S_0+1}}{\lambda_0(1-\rho)}$

Proof. The number of demand at the plant waiting for service when the plant inventory level is nonpositive is known as the plant's backorder level. The expected plant backorder level is given by (3.2.13):

$$B_0 = E[N_0] - \sum_{s=0}^{S_0-1}[1 - F_0(s)],$$

where

$$\begin{aligned} F_0(s) &= \sum_{m=0}^{s} P\{N_0 = m\}. \\ &= \sum_{m=0}^{s}(1-\rho)\rho^m = 1 - \rho^{s+1} \end{aligned}$$

$E[N_0]$ denotes the steady state expected quantity of outstanding orders in the system.

$$\begin{aligned} E[N_0] &= \sum_{n=0}^{\infty} n(1-\rho)\rho^n, \\ &= (1-\rho)\rho \sum_{n=0}^{\infty} \frac{d}{d\rho}(\rho^n) \\ &= (1-\rho)\rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\ &= \frac{\rho}{1-\rho} \end{aligned}$$

$$
\begin{aligned}
B_0 =& \frac{\rho}{1-\rho} - \sum_{s=0}^{S_0-1} \big(1 - (1-\rho^{s+1})\big) \\
=& \frac{\rho}{1-\rho} - \sum_{s=0}^{S_0-1} \rho^{s+1} \\
=& \frac{\rho}{1-\rho} - \frac{\rho(1-\rho^{S_0})}{(1-\rho)} \\
=& \frac{\rho}{1-\rho} - \frac{\rho-\rho^{S_0+1}}{1-\rho}
\end{aligned}
$$

Therefore

$$B_0 = \frac{\rho^{S_0+1}}{1-\rho} \tag{3.2.15}$$

$$I_0 = S_0 - \frac{\rho}{1-\rho}(1-\rho^{S_0}) \tag{3.2.16}$$

Applying Little's law, the expected plant response time is obtained:

$$Wt_0 = \frac{B_0}{\lambda_0} = \frac{\rho^{S_0+1}}{\lambda_0(1-\rho)} \tag{3.2.17}$$

□

3.2.2 Inventory level at pool for model I

In this subsection, we treat each pool as a single facility. Demand faced by each pool is satisfied either through pool inventory on hand or through backorders. Backorder describes a scenario in which demand not fulfilled instantly, wait in queue until they are satisfied. We assume that base stock level S_{vw} is identical for all SVCs in same pool. Hence the pool stock level is given by $S_w = \hat{w}S_{vw}$, where $\hat{w}$ represents the number of SVCs in pool w. Following the (S-1,S) policy implies that if s units are in replenishment to pool w, then the inventory on hand is given by $\hat{w}S_{vw} - s$. The (S-1,S) ensures that the sum of the number of items in replenishment to a facility and its inventory level is equal to its base stock level. That is $I_w = \hat{w}S_{vw} - s$ iff $N_w = s$. This implies that $P\{I_w = \hat{w}S_{vw} - s\}$ is equal to $P\{N_w = s\}$. In steady state the distribution of outstanding orders in pool w is $P\{N_w = s\}$. By definition, it then follows that in steady state the expected pool inventory level for pool w is:

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} (\hat{w}S_{vw} - s)P\{N_w = s\}$$

or

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) \tag{3.2.18}$$

where

$$F_w(s) = \sum_{m=0}^{s} P\{N_w = m\}$$

The following proposition establishes the expected pool backorder level in steady state.

Proposition 3.2.2.1

In steady state the expected pool backorder level is

$$B_w = \lambda_w L_w - \hat{w}S_{vw} + \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) \tag{3.2.19}$$

Proof. The number of demand at a pool waiting in queue for service when the pool's inventory level is nonpositive is known as the pool's backorder level. By the model assumptions, backorders can only occur in Pool w if all SVCs in that pool are out of stock. Thus in steady state the expected backorder level for each pool is

$$\begin{aligned} B_w &= E[N_w] - \sum_{s=0}^{\hat{w}S_{vw}-1} [1 - F_w(s)] \qquad (3.2.20) \\ &= E[N_w] - \sum_{s=0}^{\hat{w}S_{vw}-1} 1 + \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) \\ &= E[N_w] - \hat{w}S_{vw} + \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) \end{aligned}$$

At any given time t, the total outstanding orders from pool w comprises of:

(1) the backorders at the plant due to Pool w at time $t - \alpha_w$ (these orders will not reach pool w before t because they had a backorder status at time $t - \alpha_w$ and as such were not shipped immediately) and

(2) the size of new order arrivals during the interval $(t - \alpha_w, t)$.

Suppose the plant processes orders using a First Come First Served (FCFS) approach, we can split plant backorders randomly Graves (1985). This suggests that the probability that a plant backorder emanates from a particular pool is proportional to the pool demand size. In steady state the expected value of the plant backorder due to pool w for each $w \in W$, is given by $(\frac{\lambda_w}{\lambda_0})B_0$. The expected value of new arrivals in a time interval of length α_w is $\lambda_w \alpha_w$. Therefore in steady state the expected value of N_w is given by:

$$\begin{aligned} E[N_w] &= \frac{\lambda_w}{\lambda_0} B_0 + \lambda_w \alpha_w = \frac{\lambda_w}{\lambda_0} \frac{\rho^{S_0+1}}{1-\rho} + \lambda_w \alpha_w \\ &= \lambda_w W_0 + \lambda_w \alpha_w = \lambda_w (W_0 + \alpha_w) = \lambda_w L_w \end{aligned} \tag{3.2.21}$$

Where $L_w = W_0 + \alpha_w$ is the plant to pool lead time. Hence

$$B_w = \lambda_w L_w - \hat{w} S_{vw} + \sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) \tag{3.2.22}$$

□

In this study, the pooling criterion is geopolitical. That is, a pool is the collection of all SVCs in a particular geopolitical zone. Consequently, we have six pools in Nigeria.

3.2.3 Inventory level at SVCs for model I

Demand faced at a SVC is satisfied instantly through on hand inventory if the inventory level is positive. When the inventory level is zero, the demand arrival at the SVC is satisfied instantly via LT from any other available SVC in same pool without zero inventory level. This is a consequence of negligible LT times. In the event that all SVCs in the pool have zero inventory level, then the demand is backordered. So SVC demand is satisfied from any one of inventory on hand, LT, and backorder.

We begin by establishing a result which shows the relationship between steady expected levels for inventory, LT and backorder at the SVC. This result builds on work done by Caglar *et al.* (2004) and Buzacott and Shanthikumar (1993).

Proposition 3.2.3.1

1. In steady state the expected SVC inventory level at each SVC_v in pool w is

$$I_{vw} = S_{vw} - E[N_{vw}] + T_{vw} + B_{vw} \tag{3.2.23}$$

2. In steady state the expected SVC backorder level at each SVC_v in pool w is

$$B_{vw} = I_{vw} - S_{vw} + E[N_{vw}] - T_{vw} \tag{3.2.24}$$

3. In steady state the expected lateral transshipment level at each SVC_v in pool w is

$$T_{vw} = I_{vw} - S_{vw} + [N_{vw}] - B_{vw} \tag{3.2.25}$$

Proof.

$$
\begin{aligned}
I_{vw} &= \sum_{s=0}^{S_{vw}-1} (S_{vw} - s) P\{N_{vw} = s\} \\
&= S_{vw} \sum_{s=0}^{S_{vw}-1} P\{N_{vw} = s\} - \sum_{s=0}^{S_{vw}-1} sP\{N_{vw} = s\} \\
&= S_{vw} (\sum_{s=0}^{\infty} P\{N_{vw} = s\} - \sum_{s=S_{vw}}^{\infty} P\{N_{vw} = s\}) \\
&- (\sum_{s=0}^{\infty} sP\{N_{vw} = s\} - \sum_{s=S_{vw}}^{\infty} sP\{N_{vw} = s\}) \\
&= S_{vw}(1 - \sum_{s=S_{vw}}^{\infty} P\{N_{vw} = s\}) - (\sum_{s=0}^{\infty} sP\{N_{vw} = s\} - \sum_{s=S_{vw}}^{\infty} sP\{N_{vw} = s\}) \\
&= S_{vw} - S_{vw} \sum_{s=S_{vw}}^{\infty} P\{N_{vw} = s\} - (E[N_{vw}] - \sum_{s=S_{vw}}^{\infty} sP\{N_{vw} = s\}) \\
&= S_{vw} - E[N_{vw}] + \sum_{s=S_{vw}}^{\infty} (s - S_{vw}) P\{N_{vw} = s\} \\
&= S_{vw} - E[N_{vw}] + \sum_{s=S_{vw}}^{\hat{w}S_{vw}} (s - S_{vw}) P\{N_{vw} = s\} \\
&+ \sum_{s=\hat{w}S_{vw}+1}^{\infty} (s - \hat{w}S_{vw}) P\{N_{vw} = s\}
\end{aligned}
$$

Therefore

$$I_{vw} = S_{vw} - E[N_{vw}] + T_{vw} + B_{vw}$$

Where

$$T_{vw} = \sum_{s=S_{vw}}^{\hat{w}S_{vw}} (s - S_{vw}) P\{N_{vw} = s\}$$

and

$$\sum_{s=\hat{w}S_{vw}+1}^{\infty} (s - \hat{w}S_{vw}) P\{N_{vw} = s\}$$

□

The next proposition gives the optimal policy for a given SVC.

Proposition 3.2.3.2

1. In steady state the expected SVC backorder level at each SVC is

$$B_{vw} = \lambda_{vw} L_w + \frac{\lambda_{vw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \tag{3.2.26}$$

2. In steady state the expected lateral transshipment level at each SVC is

$$T_{vw} = \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \frac{\lambda_{vw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \tag{3.2.27}$$

Proof. By definition, in steady state the expected SVC inventory level is given by

$$I_{vw} = \sum_{s=0}^{S_{vw}-1} (S_{vw} - s) P\{N_{vw} = s\} \tag{3.2.28}$$

By the model assumptions, backorders can only occur if all SVCs in a pool are out of stock. Recall, $S_w = \hat{w}S_{vw}$ is the total base stock level of the Pool w. Then in steady state the expected SVC backorder level at each SVC in pool w is

$$B_{vw} = \left(\frac{\lambda_{vw}}{\lambda_w} \right) B_w \tag{3.2.29}$$

(3.2.29) follows from the splitting property of Poisson processes. Using (3.2.19) in (3.2.29)

$$B_{vw} = \left(\frac{\lambda_{vw}}{\lambda_w} \right) \left(\lambda_w L_w + \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \right)$$

hence

$$B_{vw} = \lambda_{vw} L_w + \frac{\lambda_{vw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \tag{3.2.30}$$

note that,

$$T_{vw} + B_{vw} + S_{vw} - I_{vw} = E[N_{vw}]$$

The size of outstanding orders for a SVC at a given point in time t, is the sum of:

(1) the backorders at the plant due to the SVC at time $t - \alpha_w$ (this will not get to the SVC before t) and

(2) the amount of new order arrivals to the SVC during the interval $(t - \alpha_w, t)$.

The splitting of Poisson processes imply that the probability a backorder emanates from a particular SVC is proportional to demand size at the SVC. In steady state the expected SVC backorder level is given by $\frac{\lambda_{vw}}{\lambda_0} B_0$

The expected value of new orders that arrive during a time interval with length α_w is $\lambda_{vw}\alpha_w$. Therefore in steady state, the expected value of N_{vw} is given by:

$$E[N_{vw}] = \lambda_{vw} L_w \tag{3.2.31}$$

By (3.2.25)

$$T_{vw} = E[N_{vw}] + I_{vw} - S_{vw} - B_{vw}$$

Therefore,

$$
\begin{aligned}
T_{vw} =& \lambda_{vw} L_w + \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \lambda_{vw} L_w - \left(\frac{\lambda_{vw}}{\lambda_w}\right) \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \\
=& \frac{\lambda_{vw}}{\lambda_w} \sum_{s=0}^{\hat{w}S_{vw}-1} [1 - F_w(s)] - \sum_{s=0}^{S_{vw}-1} [1 - F_{vw}(s)] \\
=& \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \left(\frac{\lambda_{vw}}{\lambda_w} (\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw}) \right)
\end{aligned}
\tag{3.2.32}
$$

□

Following METRIC method (Shebrooke (1968)), the distributions for outstanding orders at the pool and SVCs are

$$
P[N_w = m] = \frac{e^{-\lambda_w L_w} (\lambda_w L_w)^m}{m!} \tag{3.2.33}
$$

and

$$
F_w(s) = \sum_{m=0}^{s} \frac{e^{-\lambda_w L_w} (\lambda_w L_w)^m}{m!} \tag{3.2.34}
$$

In the above, L_w is the expected replenishment lead time which is made up of expected plant response time and delivery lead time.

$$
L_w = W_0 + \alpha_w = \frac{\rho^{S_0+1}}{\lambda(1-\rho)} + \alpha_w \tag{3.2.35}
$$

Similarly for SVCs

$$
P[N_{vw} = m] = \frac{e^{-\lambda_{vw} L_w} (\lambda_{vw} L_w)^m}{m!} \tag{3.2.36}
$$

and

$$
F_{vw}(s) = \sum_{m=0}^{s} \frac{e^{-\lambda_{vw} L_w} (\lambda_{vw} L_w)^m}{m!} \tag{3.2.37}
$$

Note that L_w here is same as that for the pool; this is because it is assumed that lateral transshipment between SVCs in a pool is instantaneous.

3.2.4 Model I properties

In this subsection we highlight properties of Model I. We begin by substituting the expressions for I_{vw}, B_{vw} and T_{vw} into our model to obtain the following reformulation

$$\min \sum_{w\in W}\sum_{v\in V}\left\{(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)-q_{vw}S_{vw}+\lambda_{vw}\left(p_{vw}L_w\right)\right.$$
$$\left.+(p_{vw}-q_{vw})\frac{\lambda_{vw}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)-\hat{w}S_{vw}\right)\right\}+h_0[S_0-\frac{\rho}{1-\rho}(1-\rho^{S_0})] \tag{3.2.38}$$

Subject to

$$S_{vw}\leq C_{vw},\ \text{for each,}v\in V \tag{3.2.39}$$

$$S_w\leq C_w=\hat{w}C_{vw},\ \text{for each,}w\in W \tag{3.2.40}$$

$$S_0\leq C_0 \tag{3.2.41}$$

$$[L_w-\tau]\,\lambda_{vw}\leq\frac{\lambda_{vw}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}[1-F_w(s)] \tag{3.2.42}$$

$$S_{vw},S_w,S_0\geq 0\text{ integer , for each }v\in V \tag{3.2.43}$$

$$\tag{3.2.44}$$

3.2.4.1 Upper bound for plant basestock level

The existence of capacity constraint in addition to our assumption of low system demand, indicate that the stock levels required to ensure the satisfaction of a desired service level lies within a small range which has the capacity as its upper bound. The storage capacity is usually small because the item considered in this study is slow moving (has low demand rate). Thus, storing a large number of item might lead to obsolescence and higher holding costs.Candas and Kutanoglu (2007), Mak and Shen (2009) and Riaz (2013) exploited similar properties to develop solution algorithms for determining optimal stocking levels. In this study we used 10 as our capacity.

For this model, this property implies that a number of problems can be solved when the basestock level at the plant is fixed to each feasible value. The solution with the least cost is the original problem's optimal solution. For fixed values of S_0, all terms dependent on just S_0 are treated as constants. Thus the only complicating constraint we have is the response time constraint.

3.2.5 Lagrange relaxation for model I

The reformulation of Model I does not give any obvious clue on the properties or structure of the model. Thus there is need to utilise a decomposition technique to decompose the model. We make use of Lagrange relaxation. For fixed values of S_0, we relax the service (response time) constraints (3.2.42) in the restricted problem in order to decompose the model and exploit the problem structure. Using γ_{vw} to denote the corresponding dual multiplier for (3.2.42), we obtain the Lagrangian Dual problem as:

$$\max_{\gamma\geq 0}\min \sum_{w\in W}\sum_{v\in V}\Bigg\{(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)-q_{vw}S_{vw}+(p_{vw}-q_{vw}+\gamma_{vw})\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s) \\ -\frac{\lambda_{vw}}{\lambda_w}(p_{vw}-q_{vw}+\gamma_{vw})\hat{w}S_{vw}+((p_{vw}+\gamma_{vw})L_w-\gamma_{vw}\tau)\lambda_{vw}\Bigg\} \tag{3.2.45}$$

Subject to

$$0\leq S_{vw}\leq C_{vw} \text{ for each,} v\in V, w\in W \tag{3.2.46}$$

$$S_{vw}\geq 0 \text{ integer , for each,} v\in V, w\in W \tag{3.2.47}$$

Lagrange relaxation decomposes the problem by SVCs and associated pools. The decomposed problem is

$$\max_{\gamma\geq 0}\min_{S}\sum_{s=0}^{S_{vw}-1}[(h_{vw}+q_{vw})F_{vw}(s)-q_{vw}]+(p_{vw}-q_{vw}+\gamma_{vw})\frac{\lambda_{vw}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}(F_w(s)-1) \\ +((p_{yz}+\gamma_{vw})L_w-\gamma_{vw}\tau))\,\lambda_{vw} \tag{3.2.48}$$

Subject to

$$0 \leq S_{vw} \leq C_{vw} \text{ for each}, v \in V, w \in W \tag{3.2.49}$$

$$S_{vw} \geq 0 \text{ integer , for each}, v \in V, w \in W \tag{3.2.50}$$

Next, we show that our objective function is strictly discretely convex with respect to S_{vw} and $S_w = \hat{w}S_{vw}$. Let $K(S_{vw}, S_w)$ represent the terms depending entirely on S_{vw} and S_w. $K(S_{vw}, S_w)$ is strictly discretely convex if the determinant of its Hessian matrix is positive definite. The Hessian matrix of $K(S_{vw}, S_w)$ is the matrix of second differences of $K(S_{vw}, S_w)$.

$$K(S_{vw}, S_w) = (h_{vw}+q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw}S_{vw} + (p_{vw}-q_{vw}+\gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (\sum_{s=0}^{S_w-1} (F_w(s)-1)) \tag{3.2.51}$$

$$\triangle_{S_{vw}} K(S_{vw}, S_w) = K(S_{vw}+1, S_w) - K(S_{vw}, S_w) = (h_{vw}+q_{vw})F_{vw}(S_{vw}) - q_{vw} \tag{3.2.52}$$

$$\begin{aligned} \triangle^2_{S_{vw}} K(S_{vw}, S_w) &= \triangle_{S_{vw}} K(S_{vw}+1, S_w) - \triangle_{S_{vw}} K(S_{vw}, S_w) \\ &= (h_{vw}+q_{vw})(F_{vw}(S_{vw}+1) - F_{vw}(S_{vw})) \\ &= (h_{vw}+q_{vw})(\sum_{m=0}^{S_{vw}+1} P\{N_{vw} = s\} - \sum_{m=0}^{S_{vw}} P\{N_{vw} = s\}) \\ &= (h_{vw}+q_{vw})P\{N_{vw} = S_{vw}+1\} \end{aligned} \tag{3.2.53}$$

$$\begin{aligned}\triangle_{S_w} K(S_{vw}, S_w) &= K(S_{vw}, S_w + 1) - K(S_{vw}, S_w) \\ &= (p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (F_w(S_w) - 1) \\ &= -(p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} [1 - F_w(S_w)] < 0 \end{aligned} \tag{3.2.54}$$

$$\triangle_{S_{vw}} (\triangle_{S_w} K(S_{vw}, S_w)) = 0 \tag{3.2.55}$$

$$\begin{aligned}\triangle^2_{S_w} K(S_{vw}, S_w) &= \triangle_{S_w} K(S_{vw}, S_w + 1) - \triangle_{S_w} K(S_{vw}, S_w) \\ &= (p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (F_w(S_w + 1) - F_w(S_w)) \\ &+ (p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (-1 + 1) \\ &= (p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (\sum_m^{S_w+1} P[N_w = m] - \sum_m^{S_w} P[N_w = m]) \\ &= (p_{vw} - q_{vw} + \gamma_{vw}) \frac{\lambda_{vw}}{\lambda_w} (P[N_w = S_w + 1]) > 0 \end{aligned} \tag{3.2.56}$$

The Hessian Matrix of the problem $Hess(S_{vw}, S_w)$ is

$$Hess(S_{vw}, S_w) = \begin{pmatrix} \triangle^2_{S_{vw}} K(S_{vw}, S_w) & \triangle_{S_w} (\triangle_{S_{vw}} K(S_{vw}, S_w)) \\ \triangle_{S_{vw}} (\triangle_{S_w} K(S_{vw}, S_w)) & \triangle^2_{S_w} K(S_{vw}, S_w) \end{pmatrix} \tag{3.2.57}$$

Since $\triangle^2_{S_{vw}} K(S_{vw}, S_w) > 0$, $\triangle^2_{S_w} K(S_{vw}, S_w) > 0$ and $\det(Hess(S_{vw}, S_w)) > 0$, we say that the problem is strictly discretely convex with respect to S_{vw} and S_w.

An important property of convex problems is that every local minimum is a global minimum (Winston (2004) p. 632).

3.2.6 Characteristics of the optimal solution for model 1

We decomposed the model by means of Lagrange relaxation. The Lagrangian dual problem gives a lower bound solution to the model's optimal solution. Consequently the SVC lower bound solution may not satisfy the service constraint. Having exploited the properties of Model I by means of Lagrange relaxation, we now proceed to determine the nature of our optimal solution. Here, our optimal solution is the basestock level that gives minimum cost and also satisfies the service or response time constraint.

Let the terms of the objective function of the primal problem (3.2.38) that depend on S_{vw} be given as:

$$H(S_{vw}) = (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw}S_{vw} + (p_{vw} - q_{vw})\frac{\lambda_{vw}}{\lambda_w}(\sum_{s=0}^{\hat{w}S_{vw}-1} (F_w(s) - 1)) \tag{3.2.58}$$

To determine optimal S_{vw} without response time constraint, let $\triangle H(S_{vw})$ be the change in the objective function value at the increase of base stock level from S_{vw} to $S_{vw} + 1$. Then

$$\begin{aligned}\triangle H(S_{vw}) =& H(S_{vw} + 1) - H(S_{vw}) \\ &= (h_{vw} + q_{vw})F_{vw}(S_{vw}) - q_{vw} + (p_{vw} - q_{vw})\frac{\lambda_{vw}}{\lambda_w}\left(\sum_{s=\hat{w}S_{vw}}^{\hat{w}S_{vw}+\hat{w}-1} (F_w(s) - 1)\right)\end{aligned} \tag{3.2.59}$$

$$\triangle H(S_{vw}) = F_{vw}(S_{vw}) - \frac{(q_{vw} + p_{vw})\lambda_{vw} \sum_{s=\hat{w}S_{vw}}^{\hat{w}S_{vw}+\hat{w}-1}[1 - F_w(s)] + \lambda_w q_{vw}}{\lambda_w(h_{vw} + q_{vw})} \tag{3.2.60}$$

The optimal S_{vw} without response time constraint can be found as follows:
Fix $S_0 = C_0$ and follow the following steps to determine a local minimum cost for $S_0 = C_0$.

1. If $\triangle H(C_{vw}) \leq 0$ then $S_{vw} = C_{vw}$ remains.

2. If $\triangle H(C_{vw}) > 0$ select S_{vw} as the largest integer such that $\triangle H(S_{vw}) \leq 0$.

Decrease the value of S_0 by one and follow the steps above. To get all local minimum solutions, the process is repeated until S_0 reaches zero. We pick the minimum of all local solutions, this becomes the global minimum solution.

The optimal S_{vw} with response time constraint can be found as follows:
Fix $S_0 = C_0$ and follow the following steps to determine a local minimum cost for $S_0 = C_0$.

1. If $\triangle H(C_{vw}) \leq 0$ and $[L_w - \tau]\,\lambda_{vw} \leq \frac{\lambda_{vw}}{\lambda_w} \sum_{s=0}^{\hat{w}C_{vw}-1}[1 - F_w(s)]$ then $S_{vw} = C_{vw}$ remains.

2. If $\triangle H(C_{vw}) > 0$ select S_{vw} as the smallest integer such that $\triangle H(S_{vw}) > 0$ and $[L_w - \tau]\,\lambda_{vw} \leq \frac{\lambda_{vw}}{\lambda_w} \sum_{s=0}^{\hat{w}S_{vw}-1}[1 - F_w(s)]$

Decrease the value of S_0 by one and follow the steps above. To get all local minimum solutions, the process is repeated until S_0 reaches zero. We pick the minimum of all local solutions, this becomes the global minimum solution.
Remark: We utilise Lagrange relaxation to decompose the model so as to enable us highlight the model properties. We showed that the decomposed problem is convex. The solution of the dual problem is only a lower bound. We proceed to show that Model I is convex.

Proposition 3.2.6.1
Model I is a convex optimisation problem.

Proof. We showed convexity of the dual problem for fixed multiplier values. Thus, the dual objective is convex for $\gamma_{vw} = 0$. Also the dual objective is equal to the primal objective when $\gamma_{vw} = 0$. Hence the objective function of the model is convex. The inequality constraint in Model I which is our response time constraint depends only on the variable $S_w = \hat{w}S_{vw}$ and can be written as

$$L_w - \tau + \frac{1}{\lambda_w} \sum_{s=0}^{S_w - 1} (F_w(s) - 1) \leq 0$$

Let

$$\bar{J}(S_w) = L_w - \tau + \frac{1}{\lambda_w} \sum_{s=0}^{S_w-1} (F_w(s) - 1)$$

$$\begin{aligned}
\triangle \bar{J}(S_w) &= \bar{J}(S_w + 1) - \bar{J}(S_w) \\
&= \frac{1}{\lambda_w}(F_w(S_w) - 1) \\
&= -\frac{1}{\lambda_w}(1 - F_w(S_w - 1)) < 0
\end{aligned}$$

where $\triangle \bar{J}(S_w)$ is the first difference of $\bar{J}(S_w)$.

$$\begin{aligned}
\triangle^2 \bar{J}(S_w) &= \triangle \bar{J}(S_w + 1) - \triangle \bar{J}(S_w) \\
&= \frac{1}{\lambda_w}(F_w(S_w + 1) - F_w(S_w)) \\
&= \frac{1}{\lambda_w}\left(\sum_{m}^{S_w+1} P[N_w = m] - \sum_{m}^{S_w} P[N_w = m] \right) \\
&= \frac{1}{\lambda_w}\left(P[N_w = S_w + 1]\right) > 0
\end{aligned}$$

where $\triangle^2 \bar{J}(S_w)$ is the second difference of $\bar{J}(S_w)$. The objective function and the inequality constraint are convex. Thus, Model I is a convex optimisation problem. □

The optimal solution to Model I can be obtained using GAMS software, thus there is no urgency to immediately develop any specialised heuristics for solving it.

3.3 Joint location and two echelon inventory model with response time requirement and lateral transshipment (model II)

Here we introduce facility location decisions into Model I and present the joint location-inventory model with service consideration and lateral transshipment. This model extends the model of Mak and Shen (2009) by incorporating LT. The model's basic formulation is presented below.

$$\min \sum_{w\in W}\sum_{v\in V}\left(f_{vw}X_{vw} + h_{vw}I_{vw} + p_{vw}B_{vw} + q_{vw}T_{vw} + \sum_{u\in U}\lambda_u Y_{uvw}d_{uvw}\right) + h_0 I_0 \tag{3.3.1}$$

Subject to:

$$\sum_{v\in V} Y_{uvw} = 1, \text{for each } ,u \in U \tag{3.3.2}$$

$$Y_{uvw} \leq a_{uvw}X_{vw}, \text{ for each,} u \in U, v \in V \tag{3.3.3}$$

$$S_{vw} \leq C_{vw}, \text{ for each,} v \in V \tag{3.3.4}$$

$$S_w \leq \hat{w}C_{vw}, \text{ for each,} w \in W \tag{3.3.5}$$

$$S_0 \leq C_0 \tag{3.3.6}$$

$$Wt_{vw} \leq \tau, \text{for each,} v \in V \tag{3.3.7}$$

$$S_{vw} \geq 0, integer, \text{for each,} v \in V \tag{3.3.8}$$

$$S_0 \geq 0 \tag{3.3.9}$$

$$X_{vw} \in \{0,1\} \text{for each,} v \in V \tag{3.3.10}$$

$$Y_{uvw} \in \{0,1\} \text{for each,} u \in U \tag{3.3.11}$$

The objective (3.3.1) is to find the minimum sum of the fixed location costs, plant inventory holding costs, SVC inventory holding costs, backorder costs at SVCs, LT costs at SVCs and transportation costs. We treat backorders at the plant as internal to the system hence they do not have monetary cost. Constraint (3.3.2) states that all demand should be assigned to SVCs. Constraints (3.3.3) states that demand cannot be assigned to any candidate location unless there is an open SVC whose resulting distance from customer is less than d_{max}. Constraints (3.3.4), (3.3.5) and (3.3.6) state that base stock levels for SVCs, pools and plant cannot be greater than the storage capacity available. The response time or service time constraints (3.3.7) require that expected response time (the time interval between when an order arrives and when it is shipped) cannot be greater than the required level. Finally, (3.3.8), (3.3.9),(3.3.10),and (3.3.11) are nonnegativity and integer constraints.

In this model, our response time requirement is that average time between order arrival and shipping must not be greater than the service guarantee in the customer contract. This is taken care of by (3.3.7). In this two-echelon system, stock-outs at a SVC are satisfied instantly when any of its pooled SVCs have inventory on hand. In a situation in which all SVCs in a pool are also out of stock, response time would be short if the plant has on-hand inventory and ships instantly. A longer response time would likely be observed if there is a stock-out situation at the plant. It is appropriate to impose a response time requirement because we are mainly concerned with the design of an inventory system for spare parts.

Each SVC possesses properties of a queuing system Kruse (1981) in which the customer orders can be regarded as the items in the system. Then the quantity of items in the system awaiting service (in line or in queue) represents the backorder level, while, the time the item spends in the system is known as the service time. Kruse (1981) derived the waiting time and with Little's law showed that the expected SVC response time is given by

$$Wt_{vw} = \frac{B_{vw}}{\sum_{u \in U} \lambda_u Y_{uvw}} \tag{3.3.12}$$

We may replace (3.3.7) with the following:

$$B_{vw} \leq \tau \sum_{u \in U} \lambda_u Y_{uvw} \tag{3.3.13}$$

3.3.1 Inventory levels for model II

Obtaining the steady state levels for Model II follows the procedure used for Model I. Model II differs from Model I with the inclusion of location decisions such as customer assignment. However, this difference only has notable effect at the level of SVCs. Thus, for Model II, in steady state, the plant and pool expected levels remain same as Model I. The only difference comes from the inclusion of the customer assignment variable which changes the definition of λ_0, λ_w and λ_{vw}. For Model II, $\lambda_{vw} = \sum_{u \in U} \lambda_u Y_{uvw}$, $\lambda_w = \sum_{v \in V} \lambda_{vw}$ and $\lambda_0 = \sum_{w \in W} \lambda_w$. Thus, the following hold for Model II

1. The steady state plant backorder level is given by $B_0 = \frac{\rho^{S_0+1}}{1-\rho}$
2. The steady state plant on hand inventory is given by $I_0 = S_0 - \frac{\rho}{1-\rho}(1 - \rho^{S_0})$
3. The expected plant response time is given by $W_0 = \frac{\rho^{S_0+1}}{\lambda_0(1-\rho)}$

4. in steady state the expected pool inventory level for pool w is:

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} (\hat{w}S_{vw} - s)P\{N_w = s\}$$

or

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s)$$

5. In steady state the expected pool backorder level is

$$B_w = \lambda_w L_w - \hat{w}S_{vw} + \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s)$$

For Model II, the following gives the on-hand inventory level, LT level and backorder level for SVCs.

Proposition 3.3.1.1

The optimal SVC policy for Model II is:

1. In steady state the expected SVC inventory level at each SVC in pool w is

$$I_{vw} = \sum_{s=0}^{S_{vw}-1} (S_{vw} - s)P\{N_{vw} = s\} \tag{3.3.14}$$

2. In steady state the expected SVC backorder level at each SVC in pool w is

$$B_{vw} = \sum_{u \in U} \lambda_u Y_{uvw} L_w + \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \tag{3.3.15}$$

   where

$$F_w(s) = \sum_{m=0}^{s} P\{N_w = m\}$$

3. In steady state the expected lateral transshipment level at each SVC in pool w is

$$T_{vw} = \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \frac{\sum_{u\in U} \lambda_u Y_{uvw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \quad (3.3.16)$$

where

$$F_{vw}(s) = \sum_{m=0}^{s} P\{N_{vw} = m\}$$

Also,

$$P[N_w = m] = \frac{e^{\lambda_w L_w} (\lambda_w L_w)^m}{m!} \quad (3.3.17)$$

and

$$F_w(s) = \sum_{m=0}^{s} \frac{e^{\lambda_w L_w} (\lambda_w L_w)^m}{m!} \quad (3.3.18)$$

Similarly

$$P[N_{vw} = m] = \frac{e^{-\sum_{u\in U} \lambda_u Y_{uvw} L_w} (\sum_{u\in U} \lambda_u Y_{uvw} L_w)^m}{m!} \quad (3.3.19)$$

and

$$F_{vw}(s) = \sum_{m=0}^{s} \frac{e^{-\sum_{u\in U} \lambda_u Y_{uvw} L_w} (\sum_{u\in U} \lambda_u Y_{uvw} L_w)^m}{m!} \quad (3.3.20)$$

3.3.2 Model II reformulation

With the results shown above, Model II is reformulated as:

$$\min \sum_{w\in W} \sum_{v\in V} \Bigg\{ f_{vw} X_{vw} + (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw}$$
$$+ \sum_{u\in U} ((p_{vw} L_w + d_{uvw}) \lambda_u Y_{uvw})$$
$$+ (p_{vw} - q_{vw}) \frac{\sum_{u\in U} \lambda_u Y_{uvw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \Bigg\} + h_0[S_0 - \frac{\rho}{1-\rho}(1-\rho^{S_0})]$$
$$(3.3.21)$$

Subject to

$$\sum_{v \in V} Y_{uvw} = 1, \text{for each } , u \in U \tag{3.3.22}$$

$$Y_{uvw} \le a_{uvw} X_{vw}, \text{ for each}, u \in U, v \in V \tag{3.3.23}$$

$$S_{vw} \le C_{vw}, \text{ for each}, v \in V \tag{3.3.24}$$

$$S_w \le C_w = \hat{w} C_{vw}, \text{ for each}, w \in W \tag{3.3.25}$$

$$S_0 \le C_0 \tag{3.3.26}$$

$$[L_w - \tau] \sum_{u \in U} \lambda_u Y_{uvw} \le \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \sum_{s=0}^{\hat{w} S_{vw} - 1} [1 - F_w(s)] \tag{3.3.27}$$

$$S_{vw}, S_w, S_0 \ge 0 \text{ integer , for each } v \in V \tag{3.3.28}$$

$$X_{vw} \in \{0,1\} \text{for each}, v \in V \tag{3.3.29}$$

$$Y_{uvw} \in \{0,1\} \text{for each}, u \in U \tag{3.3.30}$$

Remark 3.3.2.1

A close inspection reveals that Model II is a Mixed Integer Nonlinear Programming Problem (MINLP). This follows because the objective function has terms that are dependent on F_{vw} and F_w which are both sums of nonlinear random variables, our response time constraint (3.3.27) is dependent on F_w, while S_{vw}, X_{vw}, and Y_{uvw} are integer variables.

We proceed to examine the properties of Model II.

3.3.3 Distribution of outstanding orders in pools and service centers for model II

From previous subsections, it is obvious that before inventory levels can be evaluated, the distribution of outstanding number of orders N_w and N_{vw} at the pool and SVCs have to be determined. In this section, a close look is taken at different approximation schemes in the literature of multi-echelon inventory management which are used to find the distribution of outstanding number of orders N_w and N_{vw} in Pools and SVCs respectively.

3.3.3.1 METRIC distribution for model II

The earliest approximation method is the multi echelon technique for repairable item control (METRIC), Shebrooke (1968). It utilises Palm's theorem Palm (1938) and also approximates outstanding order distribution by means of a Poisson distribution and its corresponding mean. Using METRIC method distributions are constructed for N_w and N_{vw} with matching means. Using the METRIC-like method, N_w is approximated with a Poisson random variable.

For model II,

$$P[N_w = m] = \frac{e^{\lambda_w L_w}(\lambda_w L_w)^m}{m!} \tag{3.3.31}$$

and

$$F_w(s) = \sum_{m=0}^{s} \frac{e^{\lambda_w L_w}(\lambda_w L_w)^m}{m!} \tag{3.3.32}$$

In the above, L_w is the expected replenishment lead time which is made up of expected plant response time and delivery lead time:

$$L_w = W_0 + \alpha_w = \frac{\rho^{S_0+1}}{\lambda_0(1-\rho)} + \alpha_w \tag{3.3.33}$$

Similarly

$$P[N_{vw} = m] = \frac{e^{-\sum_{u \in U} \lambda_u Y_{uvw} L_w}(\sum_{u \in U} \lambda_u Y_{uvw} L_w)^m}{m!} \tag{3.3.34}$$

and

$$F_{vw}(s) = \sum_{m=0}^{s} \frac{e^{-\sum_{u \in U} \lambda_u Y_{uvw} L_w}(\sum_{u \in U} \lambda_u Y_{uvw} L_w)^m}{m!} \tag{3.3.35}$$

In the above, $\sum_{u \in U} \lambda_u Y_{uvw}$ is the total demand assigned to SVC v in pool w . Note that L_w here is same as that for the pool; this is because it is assumed that lateral transshipment between SVCs in a pool is instantaneous. Thus the lateral transshipment times are negligible.

METRIC discards the dependence between successive replenishment times from Plant to SVC. These replenishment times depend on the plant inventory, hence they are not independent. Axsater (1990) showed that in general METRIC works for systems that have low SVC demand compared to overall demand. The METRIC approximation performs well in such instances mainly due to the fact that successive replenishment times to a SVC is reduced as a result of many other order arrivals at the plant from other SVCs.

Caglar *et al.* (2004) show that METRIC is a very good approximation for our system. This is because the demand occurring at each SVC is low compared to total demand.

3.3.3.2 Exact distribution for model II

Graves (1985) proposed an exact probabilistic algorithm to obtain the steady state distribution of N_{vw}. Finding the exact steady state distribution begins from finding the distribution of plant total outstanding orders, this distribution is then disaggregated into the distributions of total outstanding order at each pool which is further disaggregated into the distributions of SVC total outstanding orders at each SVC. The total outstanding orders at the plant is derived from Graves (1985) as:

$$N_0 = B_0 + D_0 \tag{3.3.36}$$

Where D_0 is the total number of new arrivals at all pools during backorder duration. In order to get the distribution of N, convolve the distribution of B_0 and D_0.

$$P(N_0 = s_0) = \sum_{l=0}^{s_0} P(B_0 = l)P(D_0 = s_0 - l)$$

Assuming an M/M/1 repair system, the distribution of B_0 is given by Buzacott and Shanthikumar (1993):

$$P(B_0 = n) = (1 - \rho)\rho^{n+S_0}, n = 0, 1, 2, 3, ... \tag{3.3.37}$$

$$P(D_0 = s_0 - l) = \frac{e^{\sum_{w\in W}\lambda_w}(\sum_{w\in W}\lambda_w)^{s_0-l}}{(s_0 - l)!}, l = 0, 1, 2, 3, ..., s_0 = 0, 1, 2, 3, ... \tag{3.3.38}$$

$$P(N_0 = s_0) = \sum_{l=0}^{s_0}(1 - \rho)\rho^{l+S_0}\frac{e^{\sum_{w\in W}\lambda_w}(\sum_{w\in W}\lambda_w)^{s_0-l}}{(s_0 - l)!}$$

$$P(N_0 = s_0) = (1 - \rho)\rho^{S_0}e^{\sum_{w\in W}\lambda_w}\sum_{l=0}^{s_0}\frac{\rho^l(\sum_{w\in W}\lambda_w)^{s_0-l}}{(s_0 - l)!} \tag{3.3.39}$$

Recall that $\sum_{w\in W}\lambda_w = \lambda_0$

Once $P(N_0 = s_0)$ is found , it is disaggregated into the distributions of outstanding orders for each pool.The plant fulfils backorder by FCFS principle, hence the binomial distribution is used for the conditional distribution $P[N_w = s_w|N_0 = s_0]$.

$$P[N_w = s_w|N_0 = s_0] = \sum_{s_0=s_w}^{\infty} \binom{s_0}{s_w} \left[\frac{\lambda_w}{\lambda_0}\right]^{s_w} \left[1 - \frac{\lambda_w}{\lambda_0}\right]^{s_0-s_w}$$

$$P[N_w = s_w] = \sum_{s_0=s_w}^{\infty} P[N_w = s_w|N_0 = s_0]P[N_0 = s_0]$$

$$P[N_w = s_w] = \sum_{s_0=s_w}^{\infty} \binom{s_0}{s_w} \left[\frac{\lambda_w}{\lambda_0}\right]^{s_w} \left[1 - \frac{\lambda_w}{\lambda_0}\right]^{s_0-s_w} P[N_0 = s_0] \quad (3.3.40)$$

$P[N_w = s_w]$ is further disaggregated for each pool into distributions of outstanding orders for each SVC. Applying same argument as above yields

$$P[N_{vw} = s_{vw}] = \sum_{s_w=s_{vw}}^{\infty} \binom{s_w}{s_{vw}} \left[\frac{\sum_{u\in U} \lambda_u Y_{uvw}}{\lambda_w}\right]^{s_{vw}} \left[1 - \frac{\sum_{u\in U} \lambda_u Y_{uvw}}{\lambda_w}\right]^{s_w-s_{vw}} P[N_w = s_w] \quad (3.3.41)$$

Graves (1985) established that the exact representation is computationally burdensome for large problems encountered in practice, so we stick to the METRIC representation.

3.3.4 Waiting time probability

Our response time constraint is a major component of this study. Thus, it is necessary to know the probability that a customer's waiting time lies in a particular interval so as to set realistic response time requirements. By the model assumptions, demand satisfaction from inventory on-hand or lateral transshipment is instantaneous,hence a customer will have to wait only if his demand is backordered, that is if the pool has zero inventory. The next proposition gives the probability that a customer's waiting time lies within an interval.

Proposition 3.3.4.1

In this two-echelon environment, the probability that the waiting time of a customer in pool w lies within the interval $(0, t]$ is given by

$$P(0 < Wt_w \leq t) = \sum_{m=0}^{S_w - 1} \left(e^{-\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!} - e^{\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!} \right) \quad (3.3.42)$$

where, Wt_w is the stochastic pool waiting time.

Proof. In steady state, the replenishment request initiated by a pool demand arrival will satisfy the demand of the S_w- th future customer. The time $\bar{t}_{S_w}$ until the S_w-th future customer arrives is Erlang/ Gamma distributed with shape parameter S_w and rate λ_w (the cdf of the sum of iid $exp(\lambda)$ random variable is an Erlang/Gamma random variable). If the $S_w - th$ customer has to wait, the relationship between his waiting time and $\bar{t}_{S_w}$ is given by

$$\bar{t}_{S_w} + Wt_w = L_w \quad (3.3.43)$$

where $L_w = W_0 + \alpha_w$ is stochastic as a result of the properties of W_0.

Thus

$$
\begin{aligned}
&P(0 < Wt_{vw} \leq t) = P(L_w - t \leq \bar{t}_{S_w} < L_w) \\
&= \left[\sum_{m=S_w}^{\infty} e^{-\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!}\right] - \left[\sum_{m=S_w}^{\infty} e^{\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!}\right] \\
&= \left[\sum_{m=0}^{\infty} e^{-\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!} - \sum_{m=0}^{S_w - 1} e^{-\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!}\right] \\
&- \left[\sum_{m=0}^{\infty} e^{-\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!} - \sum_{m=0}^{S_w - 1} e^{\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!}\right] \\
&= \left[1 - \sum_{m=0}^{S_w - 1} e^{-\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!}\right] - \left[1 - \sum_{m=0}^{S_w - 1} e^{-\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!}\right] \\
&= \sum_{m=0}^{S_w - 1} \left(e^{-\lambda_w (L_w - t)} \frac{(\lambda_w (L_w - t))^m}{m!} - e^{-\lambda_w L_w} \frac{(\lambda_w L_w)^m}{m!}\right)
\end{aligned}
$$

□

3.3.5 Upper bound for plant basestock level

The following model characteristic gives an upper bound to plant's maximum stock level, it describes the model when backorder cost is set to zero and the response time threshold never gets less than the plant to pool lead times.

Proposition 3.3.5.1

Given that the objective function (3.3.21) is strictly increasing in S_0, if $p_{vw} = 0$, $L_{w_{max}} = W_0 + \max_{w \in W}\{\alpha_w\}$, and $L_{w_{max}} \leq \tau$, an upper bound of S_0 exists, which is denoted by S_0^{max}.

$$
S_0^{max} = \min\{S_0 \geq 0 : L_{w_{max}} \leq \tau\} \tag{3.3.44}
$$

Proof. Our response time constraint (3.3.27) can also be written as can also be written as

$$
[L_w - \tau] \leq \frac{\sum_{s=0}^{\hat{w}S_{vw}-1}[1 - F_w(s)]}{\lambda_w}.
$$

The requirement $L_{w_{max}} \leq \tau$ suggest that for any value of S_{vw} , LHS of (3.3.27) will be nonpositive. By definition, $0 \leq F_w(s) \leq 1$ and $1 - F_w(s) \geq 0$, therefore at S_0^{max}, $\frac{\sum_{s=0}^{\hat{w}S_{vw}-1}[1-F_w(s)]}{\lambda_w} \geq 0$ always hold, implying that the service constraint is always satisfied at S_0^{max}. Thus, the only constraints on S_0 are the response time constraint and requirement of nonnegativity, hence the feasibility of S_0^{max} holds with respect to other decision variables. A consequence of our objective function (3.3.21) being strictly monotone increasing in S_0, is that any solution having $S_0 > S_0^{max}$ will obviously be suboptimal. □

The model characteristic shown above, the existence of capacity constraint, in addition to the assumption of Poisson arrivals, suggest that the stock levels to be considered in order to attain a given service level lies within a small range which has plant capacity as its upper bound. Similar properties have also been used by Candas and Kutanoglu (2007) and Mak and Shen (2009) to develop solution algorithms.

For this model, the capacity constraint on plant implies a number of problems can be solved by fixing S_0 to each feasible value. The best cost from these solutions give the optimal solution of the initial problem. For fixed values of S_0 we treat all terms depending on S_0 alone as constants.

3.3.6 Lagrange relaxation for model II

At first view, Model II looks complex and it is difficult to determine its properties. Thus, in order to explore the model properties, there is a need to utilise a technique which can decompose the model and also take care of its complicating constraints. In this study, we make use of the technique of Lagrange relaxation to decompose our models. The complicating constraints in Model II are the assignment constraint and the response time constraints. We decide to relax our assignment constraints (3.3.22) and the response time constraints (3.3.27) in the reduced problem. Using π_u and θ_{vw} to denote the corresponding dual multipliers for constraints (3.3.22) and (3.3.27) respectively, the Lagrangian Dual problem is obtained:

$$\max_{\theta\geq 0,\pi}\min_{X,Y,S}\sum_{w\in W}\sum_{v\in V}\Bigg\{f_{vw}X_{vw}+(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)-q_{vw}S_{vw}$$
$$+(p_{vw}-q_{vw}+\theta_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)$$
$$-\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}(p_{vw}-q_{vw}+\theta_{vw})\hat{w}S_{vw}+(p_{vw}+\theta_{vw})L_w\sum_{u\in U}\lambda_u Y_{uvw}$$
$$+\sum_{u\in U}\left((d_{uvw}-\theta_{vw}\tau)\lambda_{uvw}-\pi_u\right)Y_{uvw}\Bigg\}+\sum_{u\in U}\pi_u \tag{3.3.45}$$

Subject to

$$Y_{uvw}\leq a_{uvw}X_{vw},\ \text{for each,}u\in U,v\in V \tag{3.3.46}$$
$$S_{vw}\leq C_{vw}X_{vw},\ \text{for each,}v\in V,w\in W \tag{3.3.47}$$
$$S_0\leq C_0 \tag{3.3.48}$$
$$X_{vw}\in\{0,1\},\ \text{for each,}v\in V \tag{3.3.49}$$
$$Y_{uvw}\in\{0,1\},\ \text{for each}u\in U \tag{3.3.50}$$
$$S_{vw},S_w,S_0\geq 0\ \text{integer , for each,}v\in V,w\in W \tag{3.3.51}$$

If $X_{vw}=0$, the solution is trivial. Thus, we fix all the X_{vw} to be 1 $\forall\ v\in V$, the problem further reduces to

$$\max_{\theta\geq 0,\pi}\min_{Y,S}\sum_{w\in W}\sum_{v\in V}\Bigg\{f_{vw}+(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)-q_{vw}S_{vw}$$
$$+(p_{vw}-q_{vw}+\theta_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)$$
$$-\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}(p_{vw}-q_{vw}+\theta_{vw})\hat{w}S_{vw}+(p_{vw}+\theta_{vw})L_w\sum_{u\in U}\lambda_u Y_{uvw}$$
$$+\sum_{u\in U}\left((d_{uvw}-\theta_{vw}\tau)\lambda_{uvw}-\pi_u\right)Y_{uvw}\Bigg\}+\sum_{u\in U}\pi_u \tag{3.3.52}$$

Subject to

$$Y_{uvw} \leq a_{uvw}, \text{for each}, u \in U \tag{3.3.53}$$

$$0 \leq S_{vw} \leq C_{vw} \text{ for each}, v \in V \tag{3.3.54}$$

$$Y_{uvw} \in \{0, 1\} \text{ for each}, u \in U \tag{3.3.55}$$

$$S_{vw} \text{ integer for each}, v \in V \tag{3.3.56}$$

Lagrange relaxation decomposes the problem by SVC and pool. For specific values of Lagrange multipliers, the model decomposes by candidate SVC locations and associated pools into subproblems of the following form:

$$\begin{aligned}
\min_{Y,S} (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} + (p_{vw} - q_{vw} + \theta_{vw}) \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) \\
- \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} (p_{vw} - q_{vw} + \theta_{vw}) \hat{w} S_{vw} + (p_{vw} + \theta_{vw}) L_w \sum_{u \in U} \lambda_u Y_{uvw} \\
+ \sum_{u \in U} \left((d_{uvw} - \theta_{vw} \tau) \lambda_u - \pi_u \right) Y_{uvw}
\end{aligned} \tag{3.3.57}$$

Subject to

$$Y_{uvw} \leq 1, \text{for each}, u \in U \tag{3.3.58}$$

$$0 \leq S_{vw} \leq C_{vw} \text{ for each}, v \in V \tag{3.3.59}$$

$$Y_{uvw} \in \{0, 1\} \text{ for each}, u \in U \tag{3.3.60}$$

$$S_{vw} \text{ integer for each}, v \in V \tag{3.3.61}$$

The solution of the above subproblem depends on the fact that SVCs have limited storage capacity. Hence the feasible values of S_{vw} lies between 0 and C_{vw} in the optimal solution. Therefore our solution approach to is to fix S_{vw} and S_w to each feasible value. Furthermore, the assignment variable Y_{uvw} is relaxed and allowed to be continuous in the interval $[0, 1]$. This continuous relaxation gives a lower bound solution when using Lagrangian relaxation.

With the values of S_{vw} and S_w fixed, our continuous subproblem is then reduced to a nonlinear problem of the form:

$$\min g(\sum_{u\in U}\lambda_u Y_{uvw}) + \sum_{u\in U} r_{uvw} Y_{uvw} \tag{3.3.62}$$

$$\text{subject to } 0 \leq Y_{uvw} \leq 1 \tag{3.3.63}$$

where

$$\begin{aligned} g\left(\sum_{u\in U}\lambda_u Y_{uvw}\right) =&(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1} F_{vw}\left(s, \sum_{u\in U}\lambda_u Y_{uvw}\right) \\ &+ (p_{vw}-q_{vw}+\theta_{vw})\frac{\sum_{u\in U}\lambda_u Y_{uvw}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s,\lambda_w) \end{aligned}$$

(3.3.62) has similar structure as a continuous (0,1) knapsack problem.

$$F_{vw}(s, \sum_{u\in U}\lambda_u Y_{uvw}) = \sum_{m=0}^{s} \frac{e^{-\sum_{u\in U}\lambda_u Y_{uvw} L_w}(\sum_{u\in U}\lambda_u Y_{uvw}\bar{L}_w)^m}{m!}$$

$$F_w(s,\lambda_w) = \sum_{m=0}^{s} \frac{e^{-\lambda_w L_w}(\lambda_w L_w)^m}{m!}$$

and

$$\begin{aligned} r_{uvw} =&(p_{vw}+\theta_{vw})L_w\lambda_u + (d_{uvw}-\theta_{vw}\tau)\lambda_u \\ &- \pi_u - \frac{\lambda_u}{\lambda_w}(p_{vw}-q_{vw}+\theta_{vw})\hat{w}S_{vw} \end{aligned}$$

let $\sum_{u\in U}\lambda_u Y_{uvw} = a$. The solution of the subproblem is dependent on the properties of $g(a)$ which are discussed below.

We begin by establishing convexity of the subproblem.

Proposition 3.3.6.1

$g(a)$ is convex in a when $a \geq 0$.

Proof. $\sum_{u \in U} \lambda_u Y_{uvw} = a$ is the total demand assigned to the SVC with base-stock level S_{vw}, and λ_w is the total demand at pool w.

$$g(a) = (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) + (p_{vw} - q_{vw} + \theta_{vw}) \frac{a}{\lambda_w} \sum_{s=0}^{S_w - 1} F_w(s)$$

Taking the derivatives of $g(a)$ with respect to a

$$\frac{d}{da} g(a) = (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} \frac{d}{da} F_{vw}(s) + (p_{vw} - q_{vw} + \theta_{vw}) \frac{\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s)}{\lambda_w}$$

where

$$F_{vw}(s) = \sum_{m=0}^{s} \frac{e^{-aL_w} (aL_w)^m}{m!}$$

$$\frac{d}{da} g(a) = (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} \sum_{m=0}^{s} \frac{1}{m!} \left(-L_w e^{-aL_{vw}} (aL_w)^m + e^{-aL_w} m(a)^{m-1} L_w^m \right)$$
$$+ (p_{vw} - q_{vw} + \theta_{vw}) \frac{\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s)}{\lambda_w}$$

$$\frac{d}{da} g(a) = -(h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} \frac{L_w e^{-aL_{vw}} (aL_w)^s}{s!} + (p_{vw} - q_{vw} + \theta_{vw}) \frac{\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s)}{\lambda_w} \tag{3.3.64}$$

$$\frac{d^2}{da^2} g(a) = (h_{vw} + q_{vw}) \frac{L_w^2 (aL_w)^{S_{vw}-1} e^{-aL_w}}{(S_{vw} - 1)!} \geq 0 \tag{3.3.65}$$

This establishes the convexity of g(a) with respect to assigned demand. □

Proposition 3.3.6.2

The subproblem (3.3.62) is a convex optimisation problem.

Proof. $\sum_{u \in U} r_{uvw} Y_{uvw}$ is linear in $\sum_{u \in U} \lambda_u Y_{uvw}$. A linear function is both convex an concave. The result follows from the fact that the sum of convex functions is also a convex function. □

From proposition 3.3.6.1 and proposition 3.3.6.2, we can solve our continuous subproblem using standard optimisation solvers. Hence, our model can be solved by optimisation solvers. We can obtain the solution to this model using GAMS, hence developing specialised heuristics for solving our model is not an urgent need at this point.
We go further to highlight other properties of the problem with the following propositions.

Proposition 3.3.6.3

For each SVC satisfying $X_{vw} = 1$, if all customers that satisfy $a_{uvw} = 1$ are assigned in a greedy order, $d_{1vw} \leq d_{2vw} \leq \cdots \leq d_{mvw}$, then the optimal assignment strategy for the subproblem (3.3.62) has the following properties.

1. if $Y_{kvw} > 0$ for some $1 \leq k \leq m$, then $Y_{u^*vw} = 1$ for all $\{1 \leq u \leq k - 1\}$
2. One and only one assignment variable Y_{u^*vw} takes on a strictly fractional value.

Proof. Property 1 follows from the order of assignment for customers in an open SVC.

$$d_{1vw} \leq d_{2vw} \leq \cdots \leq d_{mvw}$$

$Y_{kvw} > 0$ implies that all or part of customer k's demand have been assigned. From our order of assignment it follows that if $Y_{kvw} > 0$ for some $1 \leq k \leq m$, then $Y_{u^*vw} = 1$ for all $\{1 \leq u \leq k - 1\}$.

To prove property two, we recall that the total demand assigned to a SVC is constrained by base stock level which is also constrained by capacity. This means that the fact that a customer falls within d_{max} distance of a SVC does not guarantee it's assignment to that SVC. Let $D_{1vw} = \lambda_1$ and $D_{jvw} = D_{(j-1)vw} + \lambda_j, j = 2, \cdots m$ be the total demand assigned to SVC v with the addition of Customer j's demand. Also let S_{jvw} be base stock level required to satisfy D_{jvw}. If $S_{jvw} \leq S_{vw}$, then $Y_{jvw} = 1$. If $S_{jvw} > S_{vw}$, $S_{(j-1)vw} \leq S_{vw}$ and there exists $\bar{D}_{jvw}$ where $D_{(j-1)vw} < \bar{D}_{jvw} < D_{jvw}$ such that $\bar{S}_{jvw} = S_{vw}$, then $Y_{jvw} > 0$. If no such $\bar{D}_{jvw}$ exists, then $Y_{jvw} = 0$. $Y_{jvw} > 0$, implies that only a fraction of the customer's demand is assigned to SVC v. Our order of assignment implies that a fractional assignment can occur only at the upper boundary of the base stock level. Therefore there exists the possibility of one fractional assignment.

Suppose that there are two fractional values; $Y_{u'vw} > 0$ and $Y_{u"vw} > 0$. Sorting them according to our order gives either of $d_{u'vw} \leq d_{u"vw}$ or $d_{u"vw} \leq d_{u'vw}$. If $d_{u'vw} \leq d_{u"vw}$ and $Y_{u"vw} > 0$, then our order of assignment and property 1 imply that $Y_{u'vw} = 1$. If $d_{u"vw} \leq d_{u'vw}$ and $Y_{u'vw} > 0$, then $Y_{u"vw} = 1$. This also follows from property 1 and our sorting order. In both cases the number of fractional variables happens to be one, which is a contradiction. Hence there exists one and only fractional variable. □

In order to find an optimal solution to the reduced subproblem, it is necessary to check for the existence of an assignment variable that has fractional value. If the existence of a fractional assignment variable is established, there is a need to determine its optimal value. The next proposition establishes optimal conditions for such assignment variable.

Proposition 3.3.6.4

A solution such that there is exactly one u^* where $1 \leq u^* \leq m$ where $0 < Y_{u^*vw} < 1$ is optimal if the conditions below are satisfied:

$$\begin{aligned}\frac{r_{u^*vw}}{\lambda_u^*} =&(h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} \frac{L_{vw}^{s+1}}{s!} (\sum_{u \in U} \lambda_u Y_{uvw})^s e^{-\sum_{u \in U} \lambda_u Y_{uvw} L_w} \\ &- \frac{(p - q + \theta_w)}{\lambda_w} \sum_{s=0}^{S_w - 1} F_w(s, \lambda_w) \end{aligned} \tag{3.3.66}$$

Proof. The Lagrangian function of subproblem (3.3.62) is

$$L_{vw}(Y,\beta_u,\zeta_u) = g(\sum_{u\in U}\lambda_u Y_{uvw}) + \sum_{u\in U} r_{uvw}Y_{uvw} - \sum_{u\in U}\beta_u Y_{uvw} + \sum_{u\in U}\zeta_u(Y_{uvw}-1) \tag{3.3.67}$$

Where β and ζ are the Lagrange multipliers associated with $Y_{uvw} > 0$ and $Y_{uvw} < 1$ respectively.

For subproblem (3.3.62) KKT conditions are:

$$\begin{aligned}\frac{dL_w(Y,\beta_u,\zeta_u)}{dY_{u^*vw}} &= \frac{dg(\sum_{u\in U}\lambda_u Y_{uvw})}{dY_{u^*vw}} + \frac{d\sum_{u\in U} r_{uvw}Y_{uvw}}{dY_{u^*vw}} - \frac{d\sum_{u\in U}\beta_u Y_{uvw}}{dY_{u^*vw}} \\ &+ \frac{d\sum_{u\in U}\zeta_u(Y_{uvw}-1)}{dY_{u^*vw}} \\ &= 0\end{aligned}$$

by chain rule

$$\begin{aligned}\frac{dg(\sum_{u\in U}\lambda_u Y_{uvw})}{dY_{u^*vw}} &= \frac{dg(\sum_{u\in U}\lambda_u Y_{uvw})}{d\sum_{u\in U}\lambda_u Y_{uvw}}\frac{d\sum_{u\in U}\lambda_u Y_{uvw}}{dY_{u*vw}} \\ &= \frac{dg(\sum_{u\in U}\lambda_u Y_{uvw})}{d\sum_{u\in U}\lambda_u Y_{uvw}}\lambda_{u^*}\end{aligned}$$

In the convexity result, we showed that

$$\begin{aligned}\frac{dg(\sum_{u\in U}\lambda_u Y_{uvw})}{d\sum_{u\in U}\lambda_u Y_{uvw}} &= -(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}\frac{L_w e^{-\sum_{u\in U}\lambda_u Y_{uvw}L_w}(\sum_{u\in U}\lambda_u Y_{uvw}L_w)^s}{s!} \\ &+ (p_{vw}-q_{vw}+\theta_{vw})\frac{\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s,\lambda_w)}{\lambda_w}\end{aligned}$$

Therefore,

$$\frac{dL_w(Y,\beta_u,\zeta_u)}{dY_{u^*vw}} = -(h_{vw}+q_{vw})\lambda_u^* \sum_{s=0}^{S_{vw}-1} \frac{L_w^{s+1}}{s!}(\sum_{u\in U}\lambda_u Y_{uvw})^s e^{-\sum_{u\in U}\lambda_u Y_{uvw} L_w} + \frac{(p_{vw}-q_{vw}+\theta_w)\lambda_u}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s,\lambda_w) + r_{uvw} - \beta_u + \zeta_u = 0 \tag{3.3.68}$$

(3.3.68) is the gradient condition.

$$\frac{dL_w(Y,\beta_u,\zeta_u)}{dY_{u^*vw}} = -(h_{vw}+q_{vw})\lambda_{u^*} \sum_{s=0}^{S_{vw}-1} \frac{L_w^{s+1}}{s!}\lambda_{u^*}(\sum_{u\in U}\lambda_u Y_{uvw})^s e^{-\sum_{u\in U}\lambda_u Y_{uvw} L_w} + \frac{(p_{vw}-q_{vw}+\theta_w)\lambda_{u^*}}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) + r_{u^*vw} - \beta_u^* + \zeta_u^* = 0 \tag{3.3.69}$$

$$\beta_u Y_{uvw} = 0, \text{for each}, u \in U \tag{3.3.70}$$

$$\zeta_u(Y_{uvw}-1) = 0, \text{for each}, u \in U \tag{3.3.71}$$

$$0 \le Y_{uvw} \le 1, foreach, u \in U \tag{3.3.72}$$

$$\beta_u, \zeta_u \ge 0, \text{for each} u \in U \tag{3.3.73}$$

Equations (3.3.70) and (3.3.71) give the complementary slackness conditions. The feasibility condition is given by (3.3.72) and (3.3.73).

If $Y_{uvw} \in (0,1)$ then by (3.3.70) and (3.3.71),$\beta_u = \zeta_u = 0$ and (3.3.68) becomes (3.3.66). The subproblem is convex, thus the KKT solution above is a global minimum.

□

3.3.7 Upper bound solution for model II

The previous section gives the method of finding the lower bound to our problem for a given Lagrangian multipliers set. A lower bound solution is a system where service constraints (response time requirement) and constraints on demand assignment are relaxed. This section exploits our lower bound solution to determine some properties of the problem's optimal solution. The decision variable affected by the service constraint is S_{vw}. So a promising approach to obtain the optimal solution from our lower bound is to enumerate over all values of S_{vw} that satisfy the response time requirement. This is possible because the feasible range of S_{vw} is small. This also follows for S_0.

3.3.8 Determining S_0, S_w and S_{vw}

In this subsection we determine a means of determining the base stock levels. For each SVC in this system, we can easily obtain the optimal stock level and it's associated cost by the procedure presented below.

Proposition 3.3.8.1

In the unconstrained problem, given any feasible demand assignment and plant stock level S_0, the optimal stock level at any service center is given by

$$S_{vw} = \min \Bigg\{ S_{vw} \geq 0 : F_{vw}(S_{vw}) > \frac{(q_{vw} - p_{vw}) \sum_{u \in U} \lambda_u Y_{uvw} \left[(\hat{w} S_{vw} + \hat{w}) - \sum_{s=\hat{w}S_{vw}}^{\hat{w}S_{vw}+\hat{w}-1} F_w(s) \right] + \lambda_w q_{vw}}{\lambda_w (h_{vw} + q_{vw})} \Bigg\} \tag{3.3.74}$$

Proof. For a given demand allotment, the objective function terms which depend on S_{vw} are:

$$(h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} + (p_{vw} - q_{vw}) \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) \tag{3.3.75}$$

Let the part of our objective function dependent on S_{vw} be denoted by:

$$
\begin{aligned}
H(S_{vw}) =& (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} \\
&+ (p_{vw} - q_{vw}) \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \Big(\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) - \hat{w} S_{vw} \Big)
\end{aligned}
$$

Let $\triangle(S_{vw})$ denote change in the value of the objective if the stock level at SVC v in pool w is increased from S_{vw} to $S_{vw}+1$. Then

$$
\begin{aligned}
\triangle H(S_{vw}) &= H(S_{vw}+1) - H(S_{vw}) \\
&= (h_{vw} + q_{vw}) F_{vw}(S_{vw}) - q_{vw} \\
&+ (p_{vw} - q_{vw}) \frac{\sum_{u \in U} \lambda_u Y_{uvw}}{\lambda_w} \left(\sum_{s=\hat{w} S_{vw}}^{\hat{w} S_{vw} + \hat{w} - 1} F_w(s) - \hat{w} \right) \\
&= F_{vw}(S_{vw}) - \frac{(q_{vw} + p_{vw}) \sum_{u \in U} \lambda_u Y_{uvw} [\hat{w} - \sum_{s=\hat{w} S_{vw}}^{\hat{w} S_{vw} + \hat{w} - 1} F_w(s)] + \lambda_w q_{vw}}{\lambda_w (h_{vw} + q_{vw})}
\end{aligned}
\tag{3.3.76}
$$

When $\triangle(S_{vw}) < 0$, increasing S_{vw} by 1 will cause a decrease in cost. Also, by definition $F_{vw}(S_{vw})$ is monotone increasing in S_{vw} and lies in the interval [0, 1]. Hence the unconstrained optimal S_{vw} can be found as follows: Beginning from the smallest feasible value of S_{vw}, increase S_{vw} by 1 as long as $\triangle(S_{vw}) \leq 0$. □

However, our unconstrained optimal basestock level may be infeasible under service and capacity constraints. A better option is to use the exact method for determining basestock levels proposed below. This procedure is different from that of Riaz (2013) because our procedure begins from $S_0 = S^{maz}$ while their's begin at $S_0 = C_0$. Also their's does not consider lateral transshipment while the presence of lateral transshipments and pooling conditions imply that we simultanously determine SVC basestock level and pool basestock level.

3.3.9 Algorithm for determining S_0, S_w and S_{uvw}

An exact method for finding optimal S_{vw} is proposed. The algorithm begins at $S_0 = S_0^{max}$ and for each pool finds all feasible S_w that satisfy (3.3.25) and (3.3.27) (recall that $S_w = \hat{w}S_{vw} \leq \hat{w}C_{vw}$). Since the feasible range for S_{vw} is small and S_{vw} is identical for service centers in a pool the model can be solved by enumerating all feasible points of S_{vw} in the interval $[0, C_{vw}]$ (fixing the basestock level for one SVC fixes the basestock level for all SVCs in same pool, thereby fixing the pool basestock level). The problem is solved for all feasible values of S_{vw} and S_w. The value that results in the minimum objective function value gives the local optimal values of S_{vw} and S_w for the initial value of S_0. However, if no feasible S_w is found for a given value of S_0, then that instance is not feasible. Once a local minimum cost solution is found for the initial value of S_0, S_0 is decreased by one and the procedure is repeated. This continues till S_0 reaches zero or the solution is infeasible for some S_0. After finding all the local minimum cost solutions, the minimum is picked, this becomes the global minimum cost for SVC basestock level of S_{vw}, pool base stock level of S_w, and plant basestock level of S_0. The optimal solution in this case comprises of the values of S_0 and S_{vw} which give minimum cost

3.3.10 Infeasibility check

A procedure of determining infeasibility is proposed, which is typical of most MINLP models. The search for a solution is terminated immediately the problem is found to be infeasible.

For every iteration arising from the Lagrangian relaxation method, a lower bound of the objective value is obtained. If an instance happens to be feasible, the following observation shows that a loose upper bound (UB_f) is available:

Proposition 3.3.10.1

Given a feasible instance, the following gives an upper bound of the objective value:

$$UB_f = \sum_{w \in W} \sum_{v \in V} f_{vw} + h_0 C_0 + \sum_{w \in W} \sum_{v \in V} h_{vw} C_{vw} + \sum_{u \in U} \lambda_u d_{max} + q_{vw}(\hat{w}S_{vw}) + \sum_{w \in W} \sum_{v \in V} p_{vw} \lambda_w L_w \tag{3.3.77}$$

Proof. If we have a feasible instance, the first five terms in (3.3.77) are obviously upper bounds of the fixed costs, plant holding costs and SVC holding costs, transportation costs and transshipment costs. The fifth term is clearly an upper bound of the transshipment cost. The sixth term is also an upper bound of the SVC backorder costs because for each $v \in V$ and each $w \in W$:

$$\lambda_w L_w \geq \sum_{u \in U} \lambda_{uvw} Y_{uvw} L_w = E[N_{vw}] \geq E[N_{vw}] - \sum_{s=0}^{S_{vw}-1} [1 - F_{vw}(s)] = B_{vw} \quad (3.3.78)$$

□

3.3.11 Properties of the optimal solution for model II

The following proposition ensures that a demand assignment satisfies the assignment constraint, that is, a customer's total demand assignment is to one and only one SVC.

Proposition 3.3.11.1

For the optimal solution of (3.3.21) for customer $u \in U$, $Y_{uv^*w} = 1$ where $v^* = arg\min_{v \in V'}\{d_{uv^*w}\}$ and $V' = \{v \in V : a_{uvw} X_{vw} = 1\}$

Proof. For some customer $u \in U$, let v^* be an open SVC with the lowest transportation cost d_{uv^*w} among all the open SVCs. The cost associated with this assignment is given as

$$\begin{aligned}(h_{v^*w} + q_{v^*w}) \sum_{s=0}^{S_{v^*w}-1} F_{v^*w}(s) - q_{v^*w} S_{v^*w} + \left(\frac{p_{v^*w} \rho^{S_0+1}}{\lambda_0(1-\rho)} + p_{v^*w}\alpha_w + d_{uv^*w} \right) \lambda_u Y_{uv^*w} \\ + (p_{v^*w} - q_{v^*w}) \frac{\sum_{u \in U} \lambda_u Y_{uv^*w}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_w-1} F_w(s) - \hat{w} S_{vw} \right) \quad (3.3.79)\end{aligned}$$

Now assume that in the optimal solution, customer u is assigned to open SVC v^0.

The cost of the assignment is

$$(h_{v^0w}+q_{v^0w})\sum_{s=0}^{S_{v^0w}-1}F_{v^0w}(s)-q_{v^0w}S_{v^0w}+\left(\frac{p_{v^0w}\rho^{S_0+1}}{\lambda_0(1-\rho)}+p_{v^0w}\alpha_w+d_{uv^0w}\right)\lambda_u Y_{uv^*w}$$
$$+(p_{v^0w}-q_{v^0w})\frac{\sum_{u\in U}\lambda_u Y_{uv^0w}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_w-1}F_w(s)-\hat{w}S_{vw}\right) \tag{3.3.80}$$

since v^0 is the optimal SVC selected

$$(h_{v^0w}+q_{v^0w})\sum_{s=0}^{S_{v^0w}-1}F_{v^0w}(s)-q_{v^0w}S_{v^0w}+\left(\frac{p_{v^0w}\rho^{S_0+1}}{\lambda_0(1-\rho)}+p_{v^0w}\alpha_w+d_{uv^0w}\right)\lambda_u Y_{uv^*w}$$
$$+(p_{v^0w}-q_{v^0w})\frac{\sum_{u\in U}\lambda_u Y_{uv^0w}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_w-1}F_w(s)-\hat{w}S_{vw}\right)$$
$$\le(h_{v^*w}+q_{v^*w})\sum_{s=0}^{S_{v^*w}-1}F_{v^*w}(s)-q_{v^*w}S_{v^*w}+\left(\frac{p_{v^*w}\rho^{S_0+1}}{\lambda_0(1-\rho)}+p_{v^*w}\alpha_w+d_{uv^*w}\right)\lambda_u Y_{uv^*w}$$
$$+(p_{v^*w}-q_{v^*w})\frac{\sum_{u\in U}\lambda_u Y_{uv^*w}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_w-1}F_w(s)-\hat{w}S_{vw}\right) \tag{3.3.81}$$

however for $d_{uv^*w}\le d_{uv^0w}$

$$(h_{v^*w}+q_{v^*w})\sum_{s=0}^{S_{v^*w}-1}F_{v^*w}(s)-q_{v^*w}S_{v^*w}+\left(\frac{p_{v^*w}\rho^{S_0+1}}{\lambda_0(1-\rho)}+p_{v^*w}\alpha_w+d_{uv^*w}\right)\lambda_u Y_{uv^*w}$$
$$+(p_{v^*w}-q_{v^*w})\frac{\sum_{u\in U}\lambda_u Y_{uv^*w}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_w-1}F_w(s)-\hat{w}S_{vw}\right)$$
$$\le(h_{v^0w}+q_{v^0w})\sum_{s=0}^{S_{v^0w}-1}F_{v^0w}(s)-q_{v^0w}S_{v^0w}+\left(\frac{p_{v^0w}\rho^{S_0+1}}{\lambda_0(1-\rho)}+p_{v^0w}\alpha_w+d_{uv^0w}\right)\lambda_u Y_{uv^*w}$$
$$+(p_{v^0w}-q_{v^0w})\frac{\sum_{u\in U}\lambda_u Y_{uv^0w}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_w-1}F_w(s)-\hat{w}S_{vw}\right) \tag{3.3.82}$$

From (3.3.81) and (3.3.82) $v^0=v^*$ □

The next result establishes the convexity of Model II.

Proposition 3.3.11.2

Model II is a convex nonlinear mixed integer problem.

Proof. We showed convexity of the dual problem for fixed multiplier values and continuous assignment variable. Thus, the dual objective is convex when $\theta_{vw} = 0$, $\pi_u = 0$ and $0 \leq Y_{uvw} \leq$. Also the dual objective is equal to the primal objective when $\theta_{vw} = 0$, $\pi_u = 0$ and $0 \leq Y_{uvw} \leq$. A nonlinear mixed integer problem is convex if it's corresponding integer relaxed problem is convex. Hence, the objective function of Model II is convex.

The complicating constraints for Model II are the response time and assignment constraints. The response time constraint is an inequality constraint while the assignment constraint is an equality constraint.

The response time constraint in Model II is same as that of Model I and also depends only on the variable $S_w = \hat{w} S_{vw}$ and can be written as

$$L_w - \tau + \frac{1}{\lambda_w} \sum_{s=0}^{S_w - 1} (F_w(s) - 1) \leq 0$$

Thus the convexity result for the response time constraint of Model I also holds for Model II.

The equality constraint can be written as $1 - \sum_{v \in V} Y_{uvw} = 0$. $1 - \sum_{v \in V} Y_{uvw}$ is the sum of a linear function and a constant, thus it is affine.

This establishes the convexity of MODEL II □

Just like Model I, convexity implies that Model II can be solved by convex optimisation solvers. The solution to Model II can be obtained using GAMS software, thus there is no urgency to immediately develop any specialised heuristics for solving it.

3.4 Model with reliable locations (model III)

Here, we present a model which takes into account possible service center failures which we call Model III. Chen *et al.* (2011) studied a single echelon uncapacitated joint reliable inventory location model which had negligible lead times and did not consider lateral transshipments. Rui (2015) considered a capacitated reliable facility location model. We incorporate ideas from Chen *et al.* (2011) and Cohen *et al.* (1989) into Model II to consider reliability of facilities in a two echelon joint location-inventory setting with response time requirements and LT.

We assume that open SVCs have a uniform failure probability γ and that SVC failures are independent. A SVC failure means that the SVC is unable to provide any service. Thus, its assigned customers are reassigned to any of the other functioning SVCs with an assignment strategy.

The system's reliability follows from the assignment strategy adopted for each level. Let r $(r = 1, ..., \hat{w})$ denote a given customer's assignment level to a SVC in pool w. When r=1, it is the customer's primary assignment. If r=2, it is the customer's first backup assignment, and so on. If a customer's level-r assigned SVC failed, the level-(r + 1) assigned SVC serves the customer as the immediate backup. In this case our assignment strategy is pool based, that is, customers' backup SVCs are SVCs in same pool. The assignment level is determined by the capacity of open SVCs in same pools. We also assume that each pool has an emergency facility which is not subject to failure and is used to satisfy pool demand if all SVCs in the pool fail.

Whenever there is a SVC failure, it's assigned demand is then reassigned to a SVC in the same pool having sufficient capacity. If all level-r SVCs assigned to a customer fails, then the customer is assigned to the emergency pool facility and a cost ϕ is incurred. The probability that a customer gets served from it's level-r assigned facility is $(1-\gamma)\gamma^{r-1}$, that is, the probability that the customer's level-r SVC is functional and all it's lower level SVCs have failed. The probability that a customer gets served from the emergency SVC is $\gamma^{\hat{w}}$, this is also the probability that all SVCs in pool w have failed.

However, it is vital to note that the level r assignment in this model cannot be interpreted to mean that there exists r closer opened SVCs. That's because of the consideration of capacity constraints. The closest SVC among the ones still functional may not have sufficient capacity to accommodate new assignments, hence we proceed to the succeeding closest SVC and check its available capacity until some SVC satisfies the capacity constraint.

All demands assigned to a SVC at any level are met through any of the following; inventory on-hand, lateral transshipment or backorder, for as long as the SVC is operational.

The assignment variable for this model is

$$Y_{uvwr} = \begin{cases} 1, \text{customer u is assigned to,} SVC_{vw}, \text{as a level r assignment} \\ 0, otherwise. \end{cases} \tag{3.4.1}$$

Other decision variables and parameters remain same as in the previous section.

In this system, we also assume that the plant is not subject to failure. This in addition to our assumption that each pool possesses an emergency SVC which is not subject to failure imply that there will always be inventory on hand and back order at the plant and pools. Thus, in steady state, the expected plant and pool levels for on-hand inventory and backorder remain the same as that of Model I with the little difference being the values of λ_0 and λ_w and λ_{vw}. Thus, the following hold for Model III

1. The steady state plant backorder level is given by $B_0 = \frac{\rho^{S_0+1}}{1-\rho}$
2. The steady state plant on hand inventory is given by $I_0 = S_0 - \frac{\rho}{1-\rho}(1-\rho^{S_0})$
3. The expected plant response time is given by $W_0 = \frac{\rho^{S_0+1}}{\lambda_0(1-\rho)}$
4. in steady state the expected pool inventory level for pool w is:

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} (\hat{w}S_{vw} - s)P\{N_w = s\}$$

or

$$I_w = \sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s)$$

5. In steady state the expected pool backorder level is

$$B_w = \lambda_w L_w - \hat{w} S_{vw} + \sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s)$$

where,
$\lambda_{vw} = \sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}$,
$\lambda_w = \sum_{v\in V} \sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}$
and
$\lambda_0 = \sum_{w\epsilon W} \lambda_W = \sum_{w\in W} \sum_{v\in V} \lambda_{vw} = \sum_{w\in W} \sum_{v\in V} \sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}$.
Here, $\lambda_{vw} = \sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}$ is the expected demand at SVC v in pool w.

For Model III, steady state expected on-hand inventory level, LT level and backorder level for SVCs are given below.

1. In steady state the expected SVC inventory level at each SVC in w is

$$I_{vw} = \sum_{s=0}^{S_{vw}-1} (S_{vw} - s) P\{N_{vw} = s\} \tag{3.4.2}$$

2. In steady state the expected SVC backorder level at SVC v in pool w is

$$\begin{aligned} B_{vw} =& \sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr} L_w \\ &+ \frac{\sum_{u\in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) \end{aligned} \tag{3.4.3}$$

where

$$F_w(s) = \sum_{m=0}^{s} P\{N_w = m\}$$

3. In steady state the expected LT level at SVC v in pool w is

$$T_{vw} = \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \frac{\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) \quad (3.4.4)$$

where

$$F_{vw}(s) = \sum_{m=0}^{s} P\{N_{vw} = m\}$$

The model is as presented below

$$\min \sum_{w \in W} \sum_{v \in V} \left[f_{vw} X_{vw} + (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} + \sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \left(p_{vw} L_w + d_{uvw} \right) + (p_{vw} - q_{vw}) \frac{\sum_{u \in U} \lambda_u \sum_{r=1}^{\hat{w}} \gamma^{r-1}(1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) + \phi \lambda_w \gamma^{\hat{w}} \right] + h_0 \left[S_0 - \frac{\rho}{1-\rho}(1-\rho^{S_0}) \right] \quad (3.4.5)$$

Subject to

$$\sum_{v \in V} Y_{uvwr} = 1, \text{ for each}, u \in U, r \in \{1, 2, ..., \hat{w}\} \tag{3.4.6}$$

$$\sum_{r=1}^{\hat{w}} Y_{uvwr} \leq a_{uvw} X_{vw}, \text{ for each}, u \in U, v \in V, w \in W \tag{3.4.7}$$

$$\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u Y_{uvwr} \leq S_{vw} \leq C_{vw}, \text{ for each}, v \in V \tag{3.4.8}$$

$$S_w \leq C_w = \hat{w} C_{vw}, \text{ for each}, w \in W \tag{3.4.9}$$

$$S_0 \leq C_0 \tag{3.4.10}$$

$$[L_w - \tau] \sum_{u \in U} \sum_{r=1}^{w|} \lambda_u (\gamma^{r-1})(1-\gamma) Y_{uvwr}$$
$$\leq \frac{\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u (\gamma^{r-1})(1-\gamma) Y_{uvwr}}{\lambda_w} \sum_{s=0}^{\hat{w} S_{vw}-1} [1 - F_w(s)] \tag{3.4.11}$$

$$S_{vw}, S_0 \geq 0 \text{ integer , for each } v \in V \tag{3.4.12}$$

$$X_{vw} \in \{0, 1\} \text{for each}, v \in V \tag{3.4.13}$$

$$Y_{uvwr} \in \{0, 1\} \text{for each}, u \in U \tag{3.4.14}$$

(3.4.11) can also be written as

$$[L_w - \tau] \leq \frac{\sum_{s=0}^{\hat{w} S_{vw}-1} [1 - F_w(s)]}{\lambda_w}$$

Constraints (3.4.6) states that a customer should be assigned to only one SVC v in pool w. Constraints (3.4.7) require that demand assignments can be made to only open SVCs which are at a distance of dmax from the customer and that a SVC cannot serve a customer at more than one level. Constraints (3.4.8) states that the basestock level at a SVC cannot be less than the sum of all possible demand assignments and cannot be greater than the capacity. Constraints (3.4.9) and (3.4.10) states that the pool and plant basestock level cannot exceed their respective capacity. Constraint (3.4.11) is the service constraint. Finally, (3.4.12),(3.4.13),and (3.4.14) are nonnegativity and integer constraints. We examine the properties of Model III and give the expected inventory level in steady state for the case of probabilistic facility failures using the distribution of the number of in replenishment orders . The following gives the SVC inventory levels for Model III

Proposition 3.4.1

1. In steady state the expected SVC inventory level at each SVC v in pool w is

$$I_{vw} = \sum_{s=0}^{S_{vw}-1} (S_{vw} - s) P\{N_{vw} = s\} \tag{3.4.15}$$

   where

$$F_{vw}(s) = \sum_{m=0}^{s} P\{N_{vw} = m\}$$

2. In steady state the expected SVC backorder level at SVC v in pool w is

$$\begin{aligned} B_{vw} = & \sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} W_0 + \sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \alpha_w \\ & + \frac{\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) \end{aligned} \tag{3.4.16}$$

3. In steady state, the expected lateral transshipment level at SVC v in pool w is

$$\begin{aligned} T_{vw} = & \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} \\ & - \frac{\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) - \hat{w} S_{vw} \right) \end{aligned} \tag{3.4.17}$$

Proof. The proof follows from Model II. The slight difference lies in including the probability that a customer gets served from it's level-r assigned facility $(1-\gamma)\gamma^{r-1}$. Also for Model III the METRIC distribution imply that

$$P[N_{vw}=m]=\frac{e^{-\sum_{u\in U}\sum_{r=0}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w)^m}{m!}$$

and

$$F_{vw}(s)=\sum_{m=0}^{s}\frac{e^{-\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w)^m}{m!}$$

Similar to Model I

$$I_{vw}=\sum_{s=0}^{S_{vw}-1}(S_{vw}-s)P\{N_{vw}=s\} \qquad (3.4.18)$$

For the reliability model, expected number of demand assigned to a SVC is given by $\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}$. Also the splitting of property Poisson processes imply that

$$B_{vw}=\left(\frac{\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}}{\lambda_w}\right)B_w$$

therefore

$$\begin{aligned}B_{vw}=&\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}L_w\\&+\frac{\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}}{\lambda_w}\left(\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)-\hat{w}S_{vw}\right)\end{aligned} \qquad (3.4.19)$$

Similarly,

$$E[N_{vw}] = \sum_{u\in U}\sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr} W_0 + \sum_{u\in U}\sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}\alpha_w \quad (3.4.20)$$

$$= \sum_{u\in U}\sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr} L_w \quad (3.4.21)$$

From our previous results for Model I, we showed that

$$T_{vw} = E[N_{vw}] + I_{vw} - S_{vw} - B_{vw}$$

Therefore,

$$T_{vw} = \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - S_{vw} - \frac{\sum_{u\in U}\sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1}(1-\gamma) Y_{uvwr}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s) - \hat{w}S_{vw} \right) \quad (3.4.22)$$

□

3.4.1 Distribution of outstanding orders in pools and service centers for model III

From the previous subsection it is obvious that before inventory levels can be evaluated for Model III, the distribution of outstanding number of orders N_w and N_{vw} at the pool and SVCs, respectively, have to be determined. We take a close look the METRIC and exact approximations used for finding the distribution of outstanding number of orders N_w and N_{vw} in Pools and SVCs, respectively.

3.4.1.1 METRIC for model III

For Model III $P[N_w = s]$ and $F_w(s)$ are same as Model II.

In Model III,

$$P[N_{vw} = m] = \frac{e^{-\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w)^m}{m!} \tag{3.4.23}$$

and

$$F_{vw}(s) = \sum_{m=0}^{s}\frac{e^{-\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}(1-\gamma)\gamma^{r-1}L_w)^m}{m!} \tag{3.4.24}$$

3.4.1.2 Exact representation for model III

The following gives the exact representation for Model III.

$P[N_0 = s_0]$ and $P[N_w = s_w]$ remain same with model I.

$$P[N_{vw} = s_{vw}] = \sum_{s_w = s_{vw}} \binom{s_w}{s_{vw}} \left[\frac{\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}}{\lambda_w}\right]^{s_{vw}} \left[1 - \frac{\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}}{\lambda_w}\right]^{s_w - s_{vw}} P[N_w = s_w] \tag{3.4.25}$$

3.4.2 Lagrange relaxation for model III

Following same procedure as Model II, we relax the model's assignment (3.4.6) and service (3.4.11) constraints in the relaxed problem. Let π_{ur} and θ_{vw} denote dual multipliers for the constraints (3.4.6) and (3.4.11) respectively. The relaxation results in the Lagrangian Dual problem below.

$$\max_{\theta,\pi\geq 0}\min_{X,Y,S}\sum_{v\in V}\left\{f_{vw}X_{vw}+(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)-q_{vw}S_{vw}\right.$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}(p_{vw}+\theta_{vwr})L_w$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}\frac{(p_{vw}-q_{vw}+\theta_{vw})}{\lambda_w}(\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)-\hat{w}S_{vw})$$
$$\left.+\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}\left((d_{uvw}-\theta_{vwr}\tau)\right)\right\}$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}}\pi_{ur}-\sum_{u\in U}\sum_{r=1}^{\hat{w}}\pi_{ur}Y_{uvwr} \quad (3.4.26)$$

Subject to

$$\sum_{r=1}^{\hat{w}}Y_{uvwr}\leq a_{uvw}X_{vw},\ \text{for each,}u\in U,v\in V,w\in W \quad (3.4.27)$$
$$\sum_{u\in U}\sum_{r=1}^{\hat{w}}\lambda_uY_{uvwr}\leq S_{vw}\leq C_{vw},\ \text{for each,}v\in V \quad (3.4.28)$$
$$S_{vw}\ \text{integer , for each } v\in V \quad (3.4.29)$$
$$X_{vw}\in\{0,1\}\text{for each,}v\in V \quad (3.4.30)$$
$$Y_{uvwr}\in\{0,1\}\text{for each,}u\in U,v\in V \quad (3.4.31)$$

When $X_{vw}=0$, the optimal solution will be zero. Thus we consider the problem for cases where $X_{vw}=1$. The Lagrangian relaxation separates this problem into subproblems by SVC and its associated pool.

$$
\begin{aligned}
\min_{Y,S} & (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} + \sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} (p_{vw} + \theta_{vwr}) L_w \\
& + \sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \frac{(p_{vw} - q_{vw} + \theta_{vw})}{\lambda_w} (\sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) - \hat{w} S_{vw}) \\
& + \sum_{u \in U} \sum_{r=1}^{|V_w|-1} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \left((d_{uvw} - \theta_{vwr} \tau) \right) + \sum_{u \in U} \sum_{r=1}^{\hat{w}} \pi_{ur} \\
& - \sum_{u \in U} \sum_{r=1}^{\hat{w}} \pi_{ur} Y_{uvwr} \qquad (3.4.32)
\end{aligned}
$$

Subject to

$$\sum_{r=1}^{\hat{w}} Y_{uvwr} \leq a_{uvw}, \text{ for each,} u \in U, v \in V \qquad (3.4.33)$$

$$\sum_{u \in U} \sum_{r=1}^{\hat{w}} \lambda_u Y_{uvwr} \leq S_{vw} \leq C_{vw}, \text{ for each,} v \in V \qquad (3.4.34)$$

$$S_{vw} \text{ integer , for each } v \in V \qquad (3.4.35)$$

$$Y_{uvwr} \in \{0, 1\} \text{for each,} u \in U, v \in V \qquad (3.4.36)$$

We can rewrite the subproblem as

$$
\begin{aligned}
\min_{Y,S} & (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) - q_{vw} S_{vw} \\
& + \sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \frac{(p_{vw} - q_{vw} + \theta_{vw})}{\lambda_w} \sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) \\
& + \sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \left\{ \lambda_u \gamma^{r-1} (1-\gamma) \left\{ (p_{vw} + \theta_{vwr}) L_w + (d_{uvw} - \theta_{vwr} \tau) \right. \right. \\
& \left. \left. - \frac{(p_{vw} - q_{vw} + \theta_{vw})}{\lambda_w} (\hat{w} S_{vw}) \right\} - \sum_{r=1}^{\hat{w}} \pi_{ur} \right\} Y_{uvwr} \qquad (3.4.37)
\end{aligned}
$$

Subject to

$$\sum_{r=1}^{\hat{w}} Y_{uvwr} \leq a_{uvw}, \text{ for each,} u \in U, v \in V \tag{3.4.38}$$

$$\sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \lambda_u Y_{uvwr} \leq S_{vw} \leq C_{vw}, \text{ for each,} v \in V \tag{3.4.39}$$

$$S_{vw} \text{ integer , for each } v \in V \tag{3.4.40}$$

$$Y_{uvwr} \in \{0,1\} \text{for each,} u \in U, v \in V \tag{3.4.41}$$

Just like Model II, the solution of the above subproblem is dependent on the fact that the SVCs have capacity constraint. Thus, the possible values of S_{vw} in the optimal solution lies between $\sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \lambda_u Y_{uvwr}$ and C_{vw}. Therefore, the subproblem can be solved by fixing S_{vw} to each feasible value.

For fixed values of S_{vw}, the constraint (3.4.39) is satisfied and we can easily express the subproblem as:

$$\begin{aligned} &\min_{Y} (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vw}-1} F_{vw}(s) \\ &+ \sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \lambda_u \gamma^{r-1} (1-\gamma) Y_{uvwr} \frac{(p_{vw} - q_{vw} + \theta_{vw})}{\lambda_w} \sum_{s=0}^{\hat{w} S_{vw}-1} F_w(s) \\ &+ \sum_{u \in U} \sum_{r=1}^{\hat{w}-1} \left\{ \lambda_u \gamma^{r-1} (1-\gamma) \left\{ (p_{vw} + \theta_{vwr}) L_w + (d_{uvw} - \theta_{vwr}\tau) \right\} - \sum_{r=1}^{\hat{w}} \pi_{ur} \right\} Y_{uvwr} \end{aligned} \tag{3.4.42}$$

Subject to

$$\sum_{r=1}^{\hat{w}} Y_{uvwr} \leq a_{uvwr}, \text{ for each,} u \in U, v \in V, w \in W \tag{3.4.43}$$

$$Y_{uvwr} \in \{0,1\} \text{for each,} u \in U, v \in V \tag{3.4.44}$$

We consider (3.4.43) for $a_{uvwr} = 1$ and relax the integrality of Y_{uvwr}. This further simplifies the subproblem to:

$$\min(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}\frac{(p_{vw}-q_{vw}+\theta_{vw})}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s)$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\left\{\lambda_u\gamma^{r-1}(1-\gamma)\left\{(p_{vw}+\theta_{vwr})L_w+(d_{uvw}-\theta_{vwr}\tau)\right\}-\sum_{r=1}^{\hat{w}}\pi_{ur}\right\}Y_{uvwr} \tag{3.4.45}$$

Subject to

$$\sum_{r=1}^{\hat{w}}Y_{uvwr}\leq 1,\ \text{for each,}u\in U, v\in V, w\in W$$

The objective function of the above subproblem can be written as:

$$\min g_2(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})+\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\{R(vw)Y_{uvwr}\} \tag{3.4.46}$$

where

$$g_2(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})=(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}F_{vw}(s)$$
$$+\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}\frac{(p_{vw}-q_{vw}+\theta_{vw})}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1}F_w(s) \tag{3.4.47}$$

and

$$R_{vw}=\lambda_u\gamma^{r-1}(1-\gamma)\left\{(p_{vw}+\theta_{vwr})L_w+(d_{uvw}-\theta_{vwr}\tau)\right\}-\sum_{r=1}^{\hat{w}}\pi_{ur} \tag{3.4.48}$$

Proposition 3.4.2.1

$g_2(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})$ is convex in $\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}$ when $\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}\geq 0$.

Proof. Convexity has been established for the non-reliability case (Model II). The proof follows from the convexity result for Model II. In this case the total demand assigned to a SVC v in pool w is $\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}$. Therefore $g_2(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})$ is convex. □

The continuous relaxation of the assignment variable imply that any of the following could occur: $\sum_{r=1}^{\hat{w}-1}Y_{uvwr} = 0$, $\sum_{r=1}^{\hat{w}-1}Y_{uvwr} = 1$ or $0 < \sum_{r=1}^{\hat{w}-1}Y_{uvwr} < 1$. From nonlinear programming if $0 < Y_{uvwr} < 1$ and the first derivative at Y_{uvwr} is equal to zero then (3.4.46) has a local minimum (Winston (2004) p. 637).

Proposition 3.4.2.2

If $0 < \sum_{r=1}^{\hat{w}-1}Y_{uvwr} < 1$, then the subproblem (3.4.45) satisfies the KKT condtions.

Proof. The Lagrangian of (3.4.45) is

$$L(Y,\kappa_u) = g_2(\sum_{u\in U}\sum_{r=o}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}) + \sum_{u\in U}\sum_{r=1}^{\hat{w}-1}R_{uvw}Y_{uvwr}\gamma^{r-1}(1-\gamma) + \kappa_u\sum_{r=1}^{\hat{w}-1}(Y_{uvwr}-1) \quad (3.4.49)$$

Where κ is the Lagrange multipliers associated with $\sum_{r=1}^{\hat{w}-1}(Y_{uvwr}-1) \leq 0$.

For subproblem (3.4.45), KKT conditions are:

$$\frac{\partial L(Y,\varrho_u,\kappa_u)}{\partial Y_{u^*vwr}} = \frac{\partial}{\partial Y_{u^*vwr}}g_2(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr}) + \frac{\partial}{\partial Y_{u^*vwr}}\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}R_{vw}Y_{uvwr} + \frac{\partial}{\partial Y_{u^*vwr}}\sum_{r=1}^{\hat{w}-1}\kappa_u(Y_{uvwr}-1) = 0$$

by chain rule

$$\begin{aligned}
&\frac{\partial g_2(\sum_{u\in U}\sum_{r=o}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})}{\partial Y_{u^*vwr}} \\
&= \frac{\partial g_2(\sum_{u\in U}\sum_{r=o}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})}{\partial \sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)} \frac{\partial \sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)}{\partial Y_{u*vwr}} \\
&= \frac{\partial g_2(\sum_{u\in U}\sum_{r=o}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})}{\partial \sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)} \lambda u^*(\gamma^{r-1}(1-\gamma))
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial g_2(\sum_{u\in U}\sum_{r=o}^{\hat{w}-1}\lambda_u\gamma^{r-1}(1-\gamma)Y_{uvwr})}{\partial \sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)} \\
&= -(h_{vw}+q_{vw})\sum_{s=0}^{S_{vw}-1}\frac{L_w^{s+1}}{s!}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma))^s e^{-\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)L_w} \\
&+ (p_{vw}-q_{vw}+\theta_{vw})\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial L(Y,\kappa_u)}{\partial Y_{u*vwr}} \\
&= -(h_{vw}+q_{vw})\lambda u^*\gamma^{r-1}(1-\gamma) \\
&\sum_{s=0}^{S_{vw}-1}\frac{L_w^{s+1}}{s!}(\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma))^s e^{-\sum_{u\in U}\sum_{r=1}^{\hat{w}-1}\lambda_u Y_{uvwr}\gamma^{r-1}(1-\gamma)L_w} \\
&+ \frac{(p-q+\theta_w)\lambda u^*\gamma^{r-1}(1-\gamma)}{\lambda_w}\sum_{s=0}^{\hat{w}S_{vw}-1} F_w(s,\lambda_w) + R_{vw} + \kappa_u = 0 \qquad (3.4.50)
\end{aligned}$$

(3.4.50) is the gradient condition.

$$\kappa_u(\sum_{r=1}^{\hat{w}-1} Y_{uvwr} - 1) \leq 0, \text{ for each,} u \in U \tag{3.4.51}$$

$$0 \leq Y_{uvwr} \leq 1, \text{ for each,} u \in U \tag{3.4.52}$$

$$\kappa_u \geq 0, \text{ for each } u \in U \tag{3.4.53}$$

If Y_{uvwr} is fractional, $\kappa_u = 0$ by (3.4.51).

Equation (3.4.51) gives the complementary slackness conditions. The feasibility condition is given by (3.4.52) and (3.4.53).

By convexity of the subproblem, the KKT solution above is a global minimum.

□

Just like our previous models, we also utilise GAMS to get solutions for this model.

3.5 Model with stochastic demand (model IV)

Decision making in facility location is long-term. The absence of precise forecasts for demand and cost when making decisions is a challenge that arises frequently in the strategic aspect of supply chain. Since it is challenging and costly to reverse those decisions, making them robust against uncertainty is of utmost importance. Grahovac and Chakravarty (2001) reviews various location models with the purpose of producing solutions that are robust and reliable. A very common procedure for this family of models is scenario-based modelling. This procedure characterises uncertainty using a finite set consisting of discrete scenarios. We have two possible explanations of the scenario set. The explanation from a dynamic perspective is that as the business setting changes over time, scenarios and their associated probabilities characterise the steady state fraction of time that the setting is in each state. While, the static perspective is that the set denotes a range of probable outcomes, out of which one will occur and then remain fixed afterwards.

It is common practice for companies to conduct scenario-based studies to consider the performances of various projects under different probable future outcomes. In facility location modeling, one can apply stochastic optimisation or robust optimisation techniques, depending on ones' defined objective function, in order to produce one or more "good" solutions which perform relatively well for a good number of or all possible scenarios.

Models I-III considered fixed demand for customers in the system. Model IV, however, includes demand uncertainty into Model II by considering a set of discrete demand scenarios. The scenarios are such that a customer's demand rate is different for each scenario. Thus, for Model IV, we assume that the demand rate (λ_u) is uncertain. The model objective involves determining the optimal costs for a given finite set of likely demand rate scenarios. Define Z as the scenario set and γ^z as probability of scenario $z \in Z$ occuring. Each $z \in Z$ comprises of different demand rates for each number of customers. In addition, a subscript z is added to decision variable (for example Y_{uvwz}) and any parameter (for example λ_{uz}) that may assume different values for different scenarios.

1. In steady state the expected inventory level for each SVC v in pool w for scenario z is

$$I_{vwz} = \sum_{s=0}^{S_{vwz}-1} (S_{vwz} - s)P\{N_{vwz} = s\} \tag{3.5.1}$$

2. In steady state, the expected backorder level for SCV v in pool w is

$$\begin{aligned} B_{vwz} =& \frac{\sum_{u\in U} \lambda_{uz} Y_{uvwz}}{\lambda_{0z}} \frac{\rho_z^{S_{0z}+1}}{1-\rho_z} + \sum_{u\in U} \lambda_{uz} Y_{uvwz} \alpha_w \\ &+ \frac{\sum_{u\in U} \lambda_{uz} Y_{uvwz}}{\lambda_{wz}} \left(\sum_{s=0}^{\hat{w}S_{vwz}-1} F_{wz}(s) - \hat{w}S_{vwz} \right) \end{aligned} \tag{3.5.2}$$

where,

$$F_{wz}(s) = \sum_{m=0}^{s} P\{N_{wz} = m\}$$

3. In steady state, the expected lateral transshipment level at SVC v in pool w is

$$T_{vw} = \sum_{s=0}^{S_{vwz}-1} F_{vwz}(s) - S_{vwz} - \frac{\sum_{u\in U} \lambda_{uz} Y_{uvwz}}{\lambda_w} \left(\sum_{s=0}^{\hat{w}S_{vwz}-1} F_{wz}(s) - \hat{w}S_{vwz} \right) \tag{3.5.3}$$

where.

$$F_{vwz}(s) = \sum_{m=0}^{s} P\{N_{vwz} = m\}$$

The scenario based problem is as formulated below:

$$\min \sum_{w \in W} \sum_{v \in V} f_{vw} X_{vw} + \sum_{z \in Z} \gamma^z \left\{ \sum_{w \in W} \sum_{v \in V} \left\{ (h_{vw} + q_{vw}) \sum_{s=0}^{S_{vwz}-1} F_{vwz}(s) - q_{vw} S_{vwz} \right.\right.$$
$$+ \sum_{u \in U} \left(\left(\frac{p_{vw} \rho^{S_{0z}+1}}{\lambda_{0z}(1-\rho_z)} + p_{vw}\alpha_w + d_{uvwz} \right) \lambda_{uz} Y_{uvwz} \right)$$
$$\left.+ (p_{vw} - q_{vw}) \frac{\sum_{u \in U} \lambda_{uz} Y_{uvwz}}{\lambda_{wz}} \left(\sum_{s=0}^{\hat{w} S_{vwz}-1} F_{wz}(s) - \hat{w} S_{vwz} \right) \right\}$$
$$\left. + \sum_{z \in Z} \gamma^z h_{0z} \left[S_{0z} - \frac{\rho_z}{1-\rho_z}(1-\rho_z^{S_{0z}}) \right] \right\} \tag{3.5.4}$$

Subject to

$$\sum_{v \in V} Y_{uvwz} = 1, \text{for each} u \in U, z \in Z \tag{3.5.5}$$

$$Y_{uvwz} \leq X_{vw}, \text{ for each,} u \in U, z \in Z \tag{3.5.6}$$

$$0 \leq S_{vwz} \leq C_{vwz}, \text{ for each,} v \in V_w, z \in Z \tag{3.5.7}$$

$$S_{wz} \leq C_{wz} = \hat{w} C_{vwz}, \text{ for each,} w \in W, z \in Z \tag{3.5.8}$$

$$S_{0z} \leq C_{0z}, \text{ for each } z \in Z \tag{3.5.9}$$

$$\left[\frac{\rho_z^{S_{0z}+1}}{\lambda_{0z}(1-\rho_z)} + \alpha_w - \tau \right] \sum_{u \in U} \sum_{z \in Z} \lambda_{uz} Y_{uvwz} \leq \frac{\sum_{u \in U} \sum_{z \in Z} \lambda_{uz} Y_{uvwz}}{\lambda_{wz}} \sum_{s=0}^{\hat{w} S_{vwz}-1} [1 - F_{wz}(s)] \tag{3.5.10}$$

$$S_{vwz}, S_{wz}, S_{0z} \geq 0 \text{ integer , for each } v \in V, w \in W, z \in Z \tag{3.5.11}$$

$$X_{vwz} \in \{0,1\} \text{for each,} v \in V, z \in Z \tag{3.5.12}$$

$$Y_{uvwz} \in \{0,1\} \text{for each,} u \in U, z \in Z \tag{3.5.13}$$

(3.5.10) can also be written as

$$\left[\frac{\rho_z^{S_{0z}+1}}{\lambda_{0z}(1-\rho_z)} + \alpha_w - \tau\right] \leq \frac{\sum_{s=0}^{\hat{w}S_{vwz}-1}[1-F_{wz}(s)]}{\lambda_{wz}}$$

Note that our objective in the formulation above is to minimise the expected cost across all scenarios. However, optimising the mean outcome can result in solutions that do not perform well under some scenarios. We can produce solutions that are more robust by using alternative objectives which reflect risk-averseness. Some examples of objectives which reflect risk-averseness are minimax regret Serra and Marianov (1998), expected failure cost Snyder and Daskin (2005) and conditional value-at-risk or $CVaR$ Chen *et al.* (2006). Furthermore, the simultaneous consideration of both risk-averse objective and average-case objective is often more desirable. Using multi-objective optimisation methods, it is then possible to obtain an efficient set of solutions that are Pareto-optimal and depend on our chosen objectives. The decision maker may then evaluate trade-offs between objectives and then select one solution from our efficient set. This method is demonstrated by Snyder and Daskin (2005) who consider trade-offs between cost and customer coverage degree.

In this study, we apply similar technique in considering two objectives, which are: worst possible 'response time' among all given scenarios and the expected cost across all scenarios. Our intent is to minimise the longest response time among customers from all scenarios. The consideration of the worst possible response time for all possible scenarios, reflects the desire to have a robust supply chain.

Model IV examines the model under a stochastic demand environment. There are different demand scenarios that each have a probability of occurrence. In this section, we look at properties of Model IV given steady state expected levels. For Model IV, the minimum expected cost across all scenarios is evaluated.

Proposition 3.5.1

1. In steady state the expected SVC inventory level at each SVC v in pool w for each scenario $z \in Z$ is

$$I_{vwz} = \sum_{s=0}^{S_{vwz}-1} (S_{vwz} - s) P\{N_{vwz} = s\} \tag{3.5.14}$$

2. In steady state the expected SVC backorder level at SVC v in pool w for scenario z is

$$B_{vwz} = \frac{\sum_{u\in U}\lambda_{uz}Y_{uvwz}}{\lambda_{0z}}\frac{\rho_z^{S_{0z}+1}}{1-\rho_z} + \sum_{u\in U}\lambda_{uz}Y_{uvwz}\alpha_w + \frac{\sum_{u\in U}\lambda_{uz}Y_{uvwz}}{\lambda_{wz}}\left(\sum_{s=0}^{\hat{w}S_{vwz}-1} F_{wz}(s) - \hat{w}S_{vwz}\right) \tag{3.5.15}$$

where,

$$F_{wz}(s) = \sum_{m=0}^{s} P\{N_{wz} = m\}$$

3. In steady state, the expected lateral transshipment level at SVC v in pool w is

$$T_{vw} = \sum_{s=0}^{S_{vwz}-1} F_{vwz}(s) - S_{vwz} - \frac{\sum_{u\in U}\lambda_{uz}Y_{uvwz}}{\lambda_{wz}}\left(\sum_{s=0}^{\hat{w}S_{vwz}-1} F_{wz}(s) - \hat{w}S_{vwz}\right) \tag{3.5.16}$$

where,

$$F_{vwz}(s) = \sum_{m=0}^{s} P\{N_{vwz} = m\}$$

This is similar to Model III results. We only replace r with z and then include a subscript of z for any variable or parameter that depends on demand.

3.5.1 Distribution of number of outstanding orders in pools and service centers for model IV

METRIC For model IV

For model IV

$$P[N_{wz} = m] = \frac{e^{\lambda_{wz}L_{wz}}(\lambda_{wz}L_{wz})^m}{m!} \tag{3.5.17}$$

and

$$F_{wz}(s) = \sum_{m=0}^{s} \frac{e^{\lambda_{wz}L_{wz}}(\lambda_{wz}L_{wz})^m}{m!} \tag{3.5.18}$$

In the above,

$$L_{wz} = W_{0z} + \alpha_w = \frac{\rho_z^{S_{0z}+1}}{\lambda(1-\rho_z)} + \alpha_w \tag{3.5.19}$$

In model IV,

$$P[N_{vwz} = m] = \frac{e^{-\sum_{u\in U}\lambda_{uz}Y_{uvwz}L_{wz}}(\sum_{u\in U}\lambda_{uz}Y_{uvwz}L_{wz})^m}{m!} \tag{3.5.20}$$

and

$$F_{vwz}(s) = \sum_{m=0}^{s} \frac{e^{-\sum_{u\in U}\lambda_{uz}Y_{uvwz}L_{wz}}(\sum_{u\in U}\lambda_{uz}Y_{uvwz}L_{wz})^m}{m!} \tag{3.5.21}$$

Exact representation for model IV

The following shows the exact representation for Model IV.

$$P(N_{0z} = s_{0z}) = (1-\rho_z)\rho_z^{S_{0z}} e^{\lambda_{0z}\alpha_{wz}} \sum_{l=0}^{s_{0z}} \frac{\rho_z^l(\lambda_{0z}\alpha_{wz})^{s_{0z}-l}}{(s_{0z}-l)!} \tag{3.5.22}$$

$$P[N_{wz} = s_{wz}] = \sum_{s_z=s_{wz}} \binom{s_{0z}}{s_{wz}} \left[\frac{\lambda_{wz}}{\lambda_{0_z}}\right]^{s_{wz}} \left[1-\frac{\lambda_{wz}}{\lambda_{0z}}\right]^{s_{0z}-s_{wz}} P[N_{0z} = s_{0z}] \tag{3.5.23}$$

$$P[N_{vwz} = s_{vwz}] = \sum_{s_{wz}=s_{vwz}} \binom{s_{wz}}{s_{vwz}} \left[\frac{\sum_{u\in U}\lambda_{uz}Y_{uvwz}}{\lambda_{wz}}\right]^{s_{vwz}} \left[1-\frac{\sum_{u\in U}\lambda_{uz}Y_{uvwz}}{\lambda_{wz}}\right]^{s_{wz}-s_{vwz}} P[N_{wz} = s_{wz}] \tag{3.5.24}$$

Proposition 3.5.1.1

The objective function of Model IV is convex.

Proof. Model IV can be treated as solving Model II for each scenario. Model IV for a single scenario is same as Model II. We have shown that Model II is convex. Thus, Model IV being a sum of convex scenarios is also convex. □

Chapter 4
RESULTS AND DISCUSSION

4.0 Introduction

In this chapter computational experiments are designed and implemented to examine the models' properties. We use GAMS to implement the experiments. Our GAMS code can be found in appendices II- V. In this study, we have a finite number of customers with known demand rates. The customers are geographically dispersed and we consider each customer's location as a node. The customers are located in different cities and each city is treated as a node. For example, suppose we have a customer $u_1 \in U$ in Lagos and another customer $u_2 \in U$ in Ibadan. In this case, the demand emanating from Lagos is said to be the demand from customer u_1 and is denoted by λ_{u_1}, while, demand emanating from Ibadan is said to be the demand from customer u_2 and is denoted by λ_{u_2}. The collection of all customers $\{u_1, u_2, u_3 \cdots u_n\}$ gives the set U. Also, each customer's location is treated as a candidate SVC location $v \in V$. The collection of the location of all customers $\{v_1, v_2, v_3, \cdots v_n\}$ gives the set V. The sets U and V are equivalent. LT can only occur among SVCs in the same pool. We use a geopolitical pooling criterion. Thus, the collection of all SVCs in a geopolitical zone form a pool $w \in W$. Consequently, we have six pools in Nigeria. The collection of all pools $\{w_1, w_2, \cdots, w_6\}$ gives us the set W.

Three data sets are used comprising of 37 nodes, 109 nodes and 181 nodes; each node is considered as a potential SVC location and as a demand node. The 37 nodes represent the most populous cities in each of the 36 states in Nigeria and Federal Capital Territory (FCT). The 109 nodes represents the 3 most populous cities in each of the 36 states in Nigeria and the FCT. The 181 nodes denotes the 5 most populous cities in each of the 36 states in Nigeria and the FCT. The population data was obtained from the 2006 census. The costs associated with opening a SVC at a candidate SVC location are the fixed cost of setting up a SVC in that location and the transportation costs from that SVC location to its assigned customers. The fixed costs are given in the data sets. The cost of transportation between a SVC location and any of its assigned customers is derived by multiplying the distance between the SVC and the assigned customer by 10^{-1}. Demand rates are obtained by multiplying the population at a node or city by 10^{-5}. For example, Ife with a population of 643582 has demand rate of 6. The demand rate for each node is constrained to be no more than 10 for nodes with very large population. The deterministic transportation time from the plant to a SVC is obtained by dividing the distance between the plant and the SVC by 2400. α_w is set to be the maximum of the transportation times from the plant to all the SVCs in pool w. For all data sets, the plant is located in Abuja.

The data sets are given in Tables 5.1- 5.8 in Appendix I.

4.1 Computational results for model I

4.1.1 Model I performance

Here we compare expected costs obtained from our model with expected costs obtained from the model without LT. This test is conducted with the 37 node and 109 node data sets for various values of ($\rho = UR$. We take note of the objective function value of our model (OBJ LT) and the objective function value of the model without LT (OBJ WLT). We also take note the Minimum RTR (MRTR) corresponding to the objective value costs of our model (MRTR LT)and the model without LT (MRTR WLT). The model without LT follows the model by Caglar *et al.* (2004) with the pooling criterion but without LT. The pooling criterion is imposed on the model without LT because for our model the pooling criterion partitions the main problem into sub problems by geographical region. Hence, a fair comparison will be to compare also with a collection of sub problems by geographical region. The results are summarised in Table 4.1.

Table 4.1: Model I performance

S/N	NODES	UR	MRTR LT	MRTR WLT	OBJ LT	OBJ WLT
1	37	0.99	0.7	0.217	25110.1896	29643.265
2	37	0.9	0.244	0.007	41713.239	44315
3	37	0.7	0.219	0.006	43648.115	46020.384
4	37	0.5	0.216	0.004	43905.332	46249.862
5	37	0.3	0.216	0.004	43965.959	46305.816
6	37	0.1	0.216	0.004	43985.072	46324.456
7	109	0.99	0.53	0.0605	99688.493	117638.186
8	109	0.9	0.34	0.005	126235.589	142015.236
9	109	0.7	0.321	0.0039	128213.665	1435757.798
10	109	0.5	0.319	0.0038	128474.328	143990.155
11	109	0.3	0.319	0.0037	128535.506	144046.565
12	109	0.1	0.319	0.0037	128554.674	144065.252

4.1.2 Discussion of model I performance

For all instances tested, the total system cost of our LT model was lower than that of the model without LT. This illustrates the cost savings that can be achieved via the incorporation of LT. The MRTR for our model was higher than that of the model without LT. This occurs because for our model, the assumption of negligible LT times imply that the lead time L_w is identical for all SVCs in a pool. However, the focus of this experiment is to compare costs only, the effect of response time requirement on our model is examined in the next result.

4.1.3 Effect of response time requirement for model I

In this experiment we check the behaviour of the model as the response time is varied. The 37-node dataset is used with (ρ) set to 0.9. From Table 5.1, Model I gives lower costs than the model without LT for all values of ρ, thus, our choice of ρ is arbitrary. We vary response time requirement values between 0.272 and 0.668. For this model, response time values greater than 0.668 will always be feasible. Hence, there is no need to increase the value of response time beyond 0.668. This is a result of the values obtained for our deterministic transportation time from the plant to pool w, α_w. The result of this experiment is shown in Fig 4.1 below.

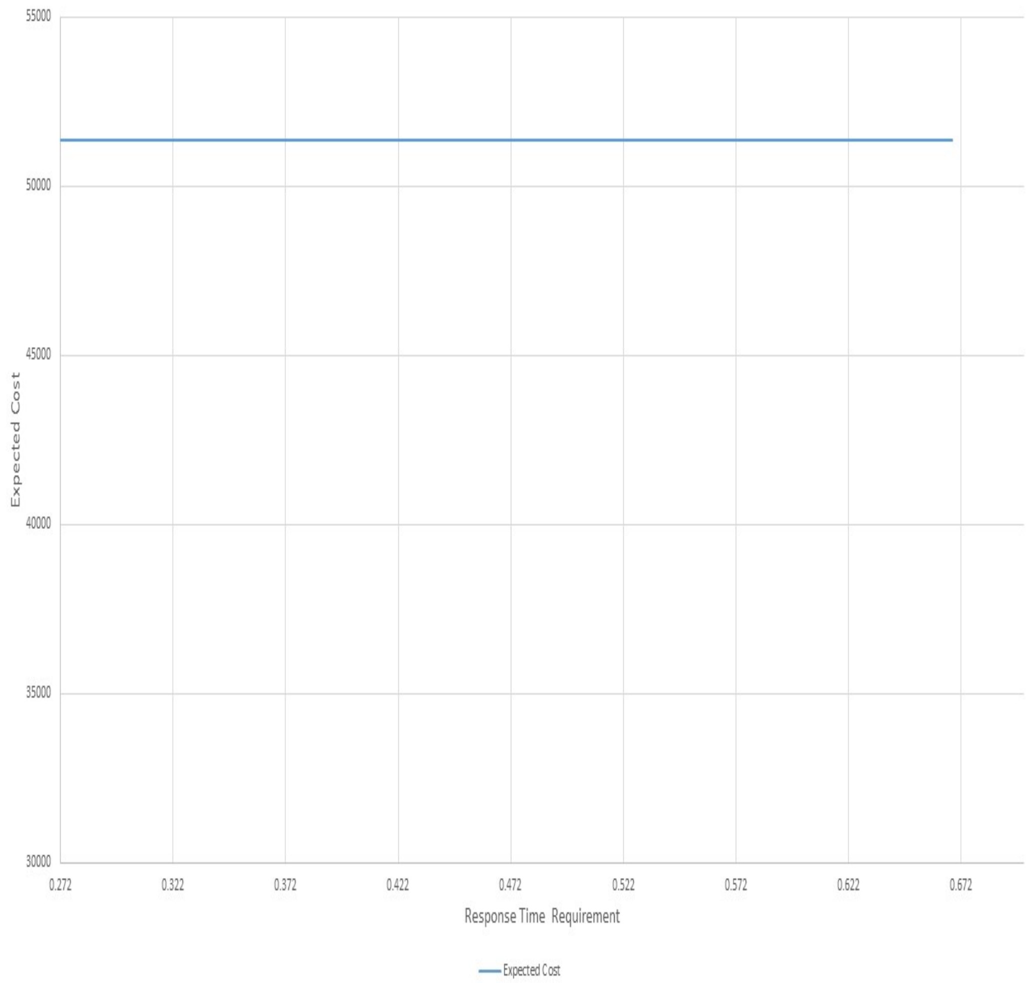

Figure 4.1: Effect of response time for model I

4.1.4 Discussion of the effect of response time requirement for model I

Figure 4.1 shows that expected cost remains stable with varying response time requirement values. This occurs as a result of lateral transshipment and pooling which ensure uniform response time constraint for all SVCs in a pool. The implication of this is that within feasible values, the decision maker can slacken or tighten the response time requirement to fit into the contract signed with a customer and this will have negligible effect on the expected cost. This is especially important, because from Table 4.1, we see that Model I gives lower costs than the model without LT. This, and the consistency of costs with varying response times, means that the decision maker is able to negotiate response times with a customer with a certain degree of certainty about the consequence of whatever agreement the make.

4.1.5 Effect of base stock level on model 1

We utilise the 37 node dataset for this experiment and set (ρ) to 0.9. From Table 4.1, Model I gives lower costs than the model without LT for all values of ρ, thus, our choice of ρ is arbitrary. In the first instance, S_{vw} is fixed at 5, while S_0 is varied between the feasible range. In the second instance S_0 is fixed at 3 and the value of S_{vw} is varied within the feasible range. In all cases the minimum feasible value of τ and the corresponding total cost are recorded. The total cost and τ are plotted against the stock level. Figure 4.2 and Figure 4.3 below, show the result of this experiment.

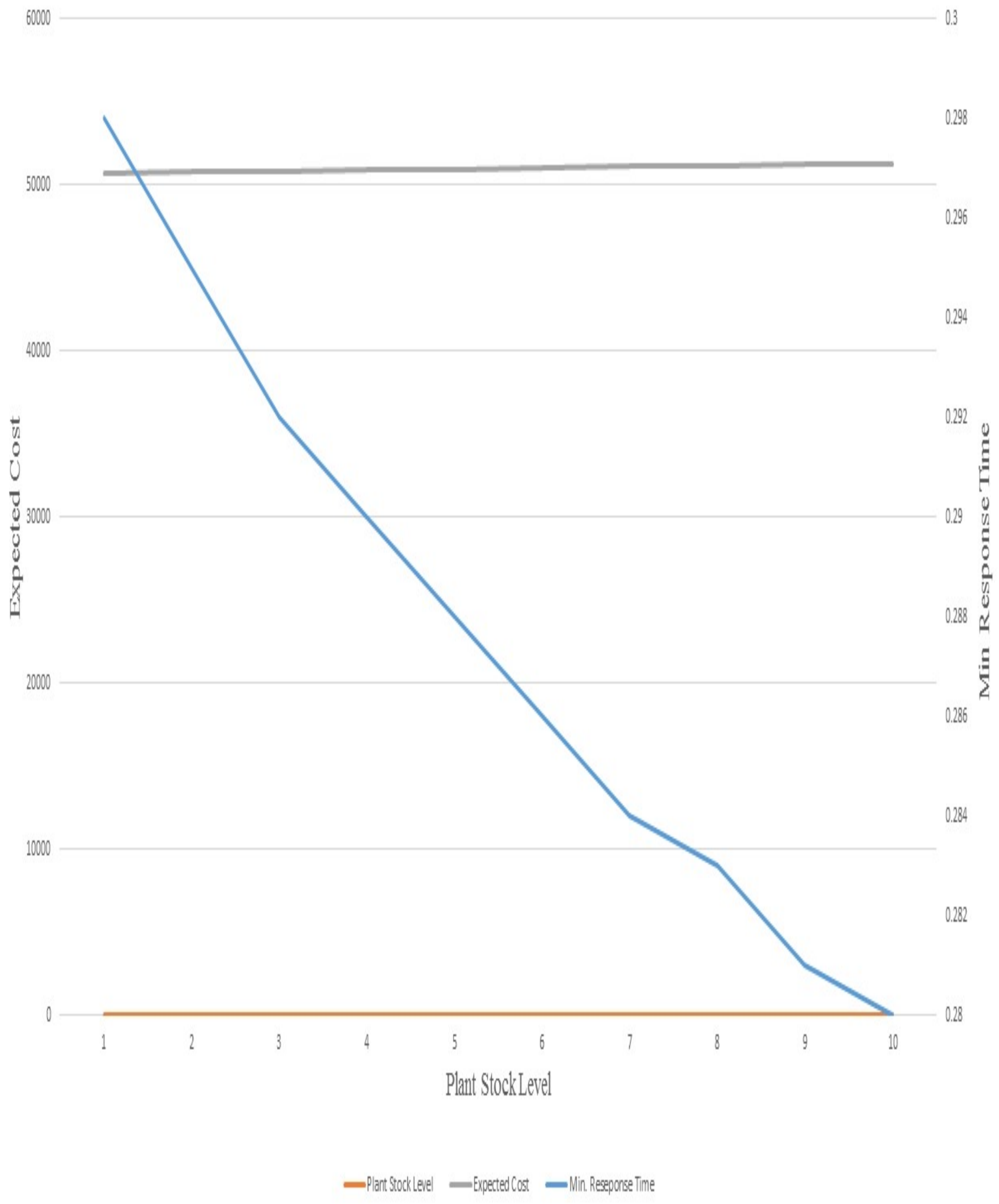

Figure 4.2: Effect of plant base stock level for model I

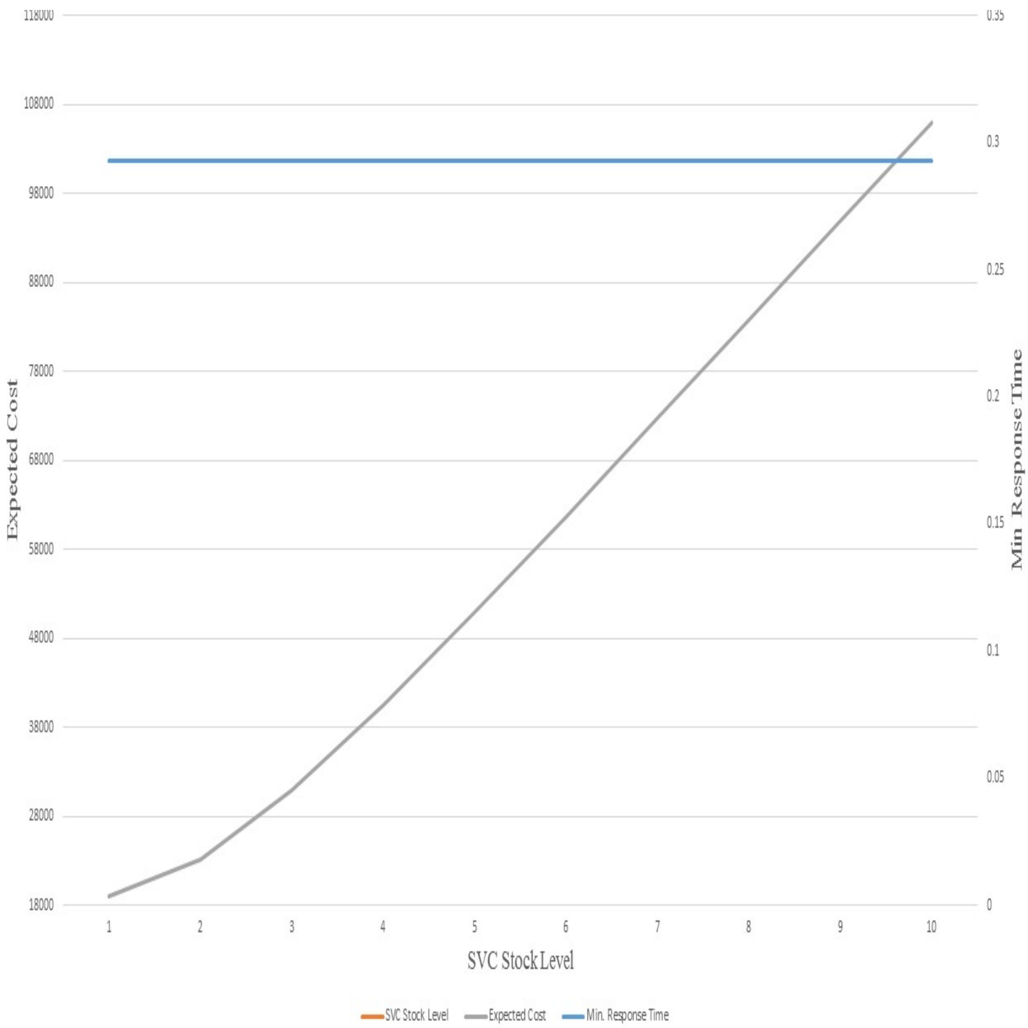

Figure 4.3: Effect of SVC base stock level for model I

4.1.6 Discussion of the effect of base stock level on model 1

Figure 4.2, shows that increasing plant base stock results in negligible increase in expected cost and causes a decrease in the minimum response time requirement. Figure 4.3, shows that increasing SVC stock level results in an increase in expected cost and causes no change in reduced minimum response time requirement. If the decision maker intends to reduce response times to customers with minimum increase in cost, she has to increase the plant base stock level. In real life, the value of S_0 is usually constrained by capacity.

4.2 Computational results for model II

In this section we utilise all three data sets. The difference between Model II and Model I is that Model II is a location-inventory model, while, Model I considers inventory alone. Consequently, all the SVC location variables such as X_{vw}, Y_{uvw}, d_{uvw} are factored into the formulation of Model II. This Model is solved using GAMS and the GAMS code for Model II can be found in Appendices IV-V.

4.2.1 Model II performance

Here we compare our model with the model without LT. This test is conducted with the three data sets for $(\rho = UR = (0.9, 0.5), \tau = RTR = (0.5, 0.3, 0.22 or 0.2), d_{max} = (150, 100)$. We take note of the objective function value of our model and the objective function value of the model without LT. The model without LT follows the model by Mak and Shen (2009) with the pooling criterion but without LT. The pooling criterion is imposed on the model without LT because for our model the pooling criterion partitions the main problem into sub problems by geographical region. Hence, a fair comparison will be to compare also with a collection of sub problems by geographical region. In this Model the plant is also located in Abuja. The results are summarised in Table 4.2.

Table 4.2: Model II performance

S/N	NODES	UR	RTR	dmax	OBJ LT	OBJ WLT	SVC LT	SVC WLT
1	37	0.9	0.5	150	306486.913	311808.6209	19	21
2	37	0.5	0.5	150	299939.9668	312063.1298	19	21
3	37	0.9	0.3	150	306486.913	311808.6209	19	21
4	37	0.5	0.3	150	299939.9668	312063.1298	19	21
5	37	0.9	0.22	150	306486.913	311808.6209	19	21
6	37	0.5	0.22	150	299939.9668	308860.178	19	21
7	37	0.9	0.5	100	383839.3325	355492.5327	28	28
8	37	0.5	0.5	100	365704.9459	355805.1491	28	28
9	37	0.9	0.3	100	383839.3325	355492.5326	28	28
10	37	0.5	0.3	100	385704.9459	355805.1491	28	28
11	37	0.9	0.2	100	383839.3325	355492.5327	28	28
12	37	0.5	0.2	100	385704.9459	355805.1491	28	28
13	109	0.9	0.5	150	413185.1743	568250.7459	25	27
14	109	0.5	0.5	150	413292.6378	568604.1851	25	27
15	109	0.9	0.3	150	413383.5425	568230.7459	25	27
16	109	0.5	0.3	150	413388.0967	565604.1851	25	27
17	109	0.9	0.2	150	413383.5425	568230.7459	25	27
18	109	0.5	0.2	150	413345.021	568604.1851	25	27
19	109	0.9	0.5	100	531475.699	640499.9813	37	38
20	109	0.5	0.5	100	531660.5781	640871.0662	37	38
21	109	0.9	0.3	100	531479.7333	640499.9813	37	38
22	109	0.5	0.3	100	531671.7892	640695.7594	37	38
23	109	0.9	0.2	100	531487.9847	640499.9813	37	38
24	109	0.5	0.2	100	531665.913	640871.0662	37	38
25	181	0.9	0.5	150	503122.2887	756096.5497	23	27
26	181	0.5	0.5	150	503121.2472	756165.2004	23	27
27	181	0.9	0.3	150	503341.0051	756096.5497	23	27
28	181	0.5	0.3	150	503193.3191	756165.2004	23	27
29	181	0.9	0.2	150	503341.0051	755797.6929	23	27
30	181	0.5	0.2	150	503421.4726	756165.2004	23	27
31	181	0.9	0.5	100	709562.535	874040.9894	44	45
32	181	0.5	0.5	100	709671.889	873901.569	44	45
33	181	0.9	0.3	100	709605.4168	873920.4	44	46
34	181	0.5	0.3	100	709597.7653	874047.5136	44	45
35	181	0.9	0.2	100	709608.2478	873971.9055	44	46
36	181	0.5	0.2	100	709629.8785	874199.5546	44	45

4.2.2 Discussion of model II performance

For the 37 node set, the entire system cost for the our LT model is lower compared to the model without LT when $d_{max} = 150$, while the model without LT performs better than our model when $d_{max} = 100$. Generally our model gives lower costs for as long as $d_{max} \geq 120$. Also in all instances tested for the 109 nodes data set and the 181 nodes data set respectively, the total system of our LT model was lower than that of the model without LT. This shows that our model is suitable for systems having many nodes.Furthermore, in most instances fewer SVCs are required for the our LT model (SVC LT) compared to the model without LT (SVC WLT).

Also, the total cost increases as d_{max} reduces from 150 to 100, this is due to the fact that more facilities need to be opened as the coverage distance reduces.

4.2.3 Effect of base stock level on response time for model II

The 37 node dataset is utilised for this experiment with (ρ) set to 0.9, d_{max} set to 150 kilometers, and S_{vw} set to 1. In the first instance, S_{vw} is fixed at 1, while S_0 is varied between the feasible range. In the second instance S_0 is fixed at 4 and the value of S_{vw} is varied within the feasible range. In both instances the minimum feasible value of τ and the corresponding total cost are recorded. The total cost and τ are plotted against base stock level. Figure 4.4 and Figure 4.5 show the results from this experiment.

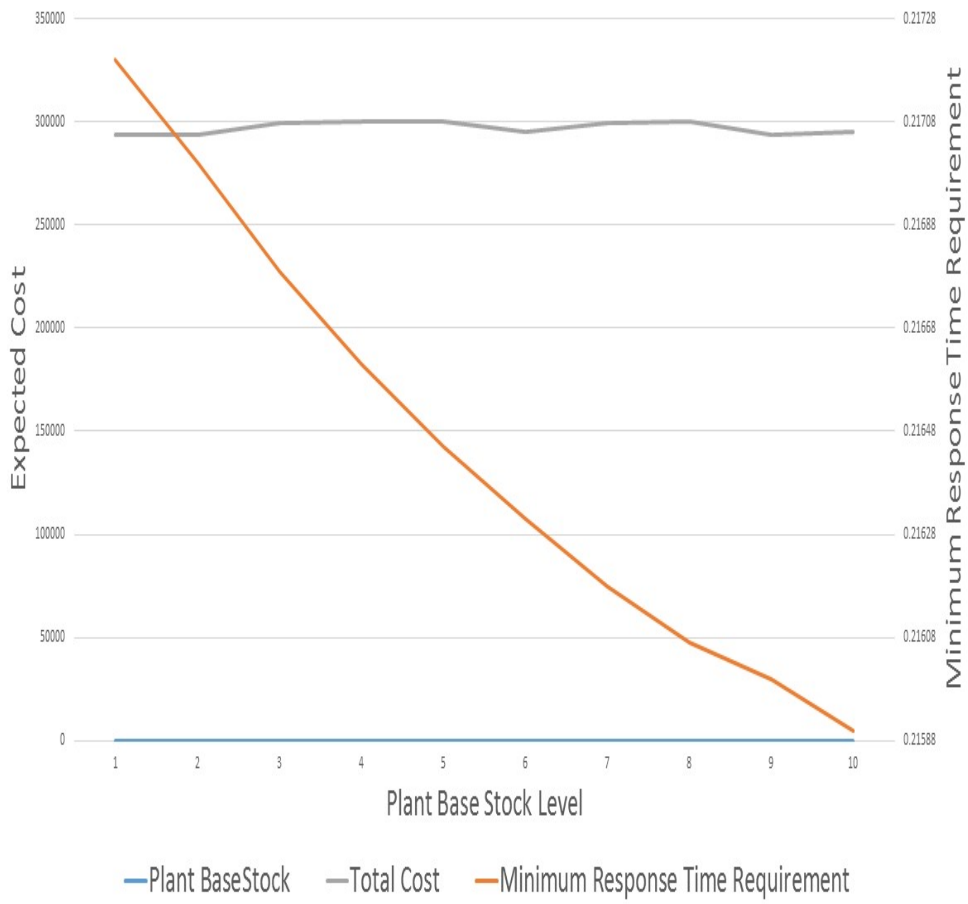

Figure 4.4: Effect of plant base stock level for model II

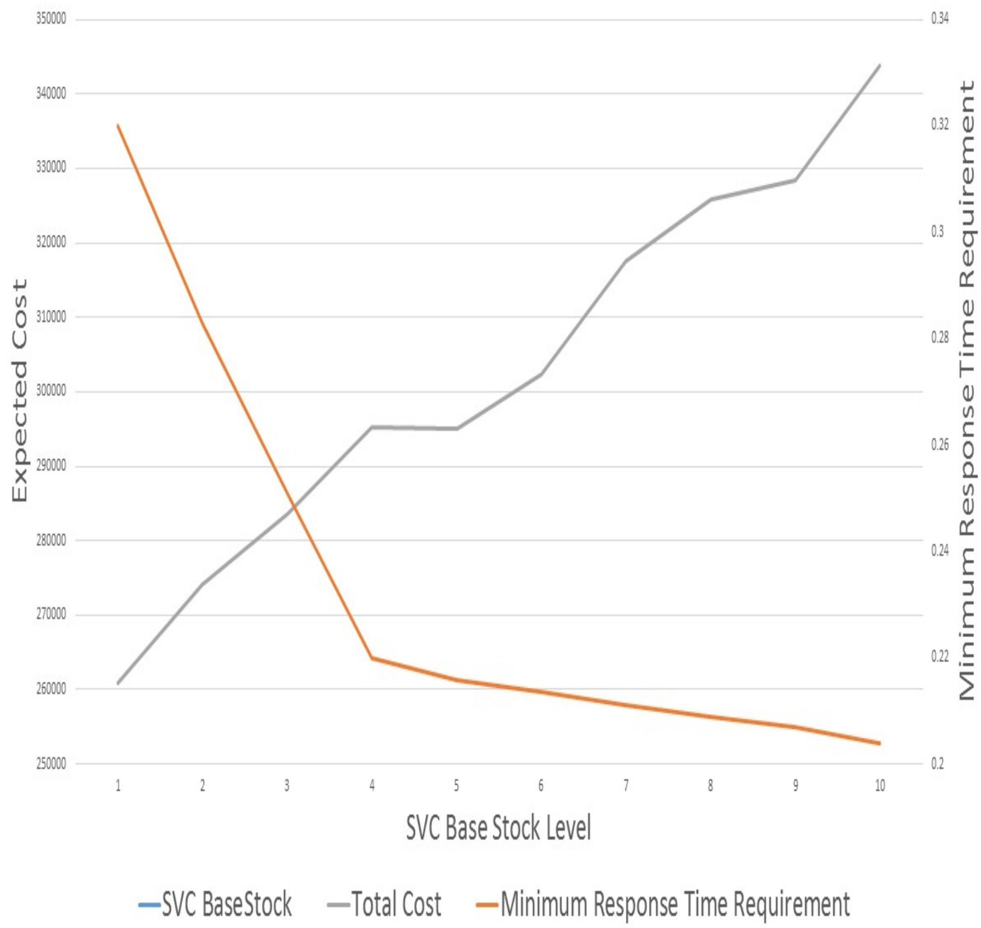

Figure 4.5: Effect of SVC base stock level for model II

4.2.4 Discussion of the effect of base stock level on response time for model II

From Figure 4.4, it is seen that an increase in plant base stock causes a decrease in minimum response time requirement and has little effect on total cost. From Figure 4.5, an increase in SVC base stock level results in reduced minimum response time requirement while the total costs increase with increase in SVC base stock level. Thus increasing SVC base stock level will lead to better response times. This however causes an increase in total cost. In real life, the value of S_{vw} is usually constrained by capacity or budget constraints.

4.2.5 Relationship between backorder cost and response time

Most literature in traditional inventory theory make use of backorder as utilisation measure. This means that in order to discourage long waiting times, penalty costs are usually imposed on backorder. That is, backorder costs are usually increased in order to discourage backorders. This experiment considers the behaviour of the model as backorder and response time are varied. The 37-node dataset is used with (ρ) set to 0.9, d_{max} set to 150 kilometers, and storage capacity set to 5. We test two sets of instances. In the first set, the penalty on backorder, p, is varied between 60 and 180 by gradually increasing 60 by $2\%, 4\%, ..., 200\%$, and for each value of p the total expected cost is observed. In the second instance, the response time requirement is varied between 0.2 and 0.6 time units by gradually increasing 0.2 by $2\%, 4\%, ..., 200\%$ and for each value, the total expected cost is also recorded. Then we plotted the total costs expected with varied response time and the total costs expected with varied backorder costs against the percentage increase. The result of this experiment is shown by Figure 4.6.

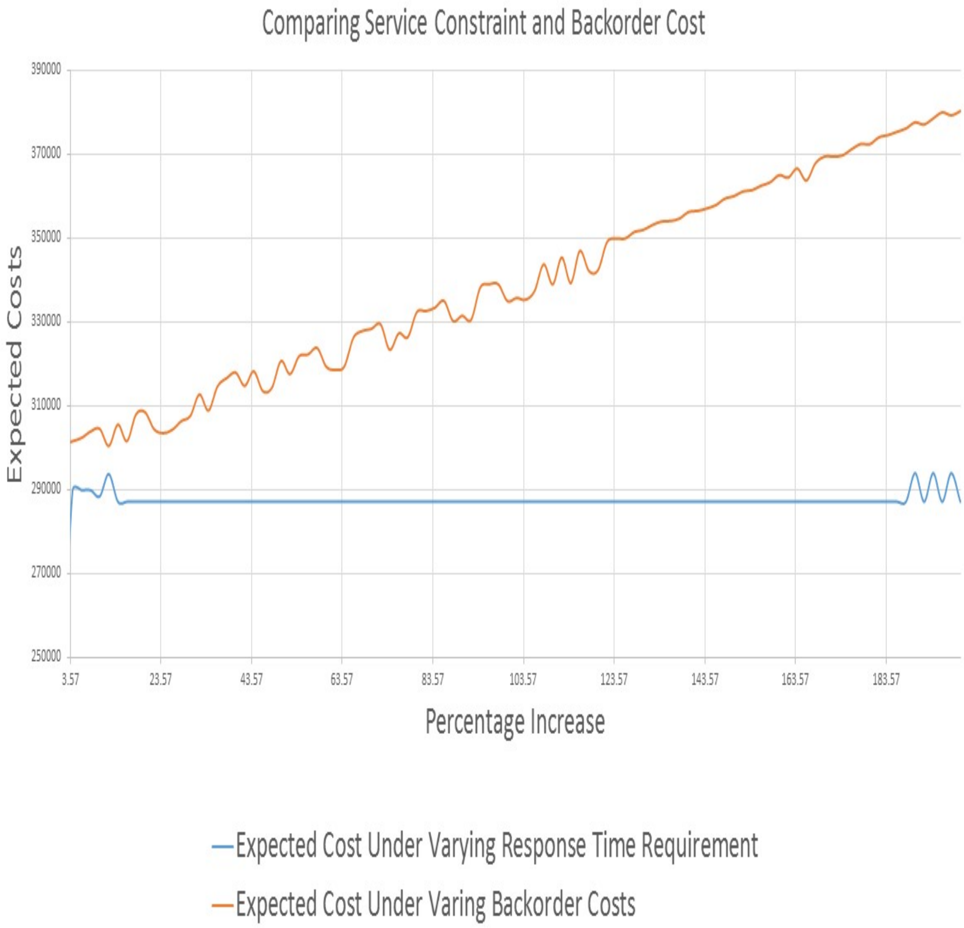

Figure 4.6: Relationship between backorder cost and response time for model II

4.2.6 Discussion of the relationship between backorder cost and response time

From Figure 4.6 it is observed that the graph for backorder is has a steeper ascent than that of response time, while the graph for response time is consistent. This implies that the change in cost observed when the response time is varied among its feasible points is minimal when compared with backorder values. Hence in terms of expected cost, the response time is a more consistent measure than backorder. Also, the response time curve is dominated by the backorder costs. This implies that using response time in place of backorder as a service measure will always result in lower costs for our system. The spikes observed on the graph for expected cost under varying response time imply that cost fluctuate when response time lies between 0.22 and 0.28. The cost also fluctuates when response time lies between 0.576 and 0.6. With these spikes the value of the expected costs under varying response require lies in the interval [287190.9578, 293965.0978]. Thus, the gap between the minimum and maximum values of the expected costs is 2.03%. Thus, fluctuations only have little effect on the expected costs.

4.2.7 Trade off between service and cost

Service consideration is very crucial in the design of efficient spare parts supply chain. However, emphasis on cost minimisation alone in supply chain design usually results in poor service delivery. Cost and service consideration can act as two divergent performance measures in supply chain design Shen and Daskin (2005). Investigating the trade-off between costs and service level is therefore critical. This computational experiment solves the problem at different levels of service requirement. Expected cost is plotted against service time requirement (τ) resulting in a trade-off curve. The same graph also shows the optimal number of SVCs against response time requirements. This is shown in Figure 4.7.

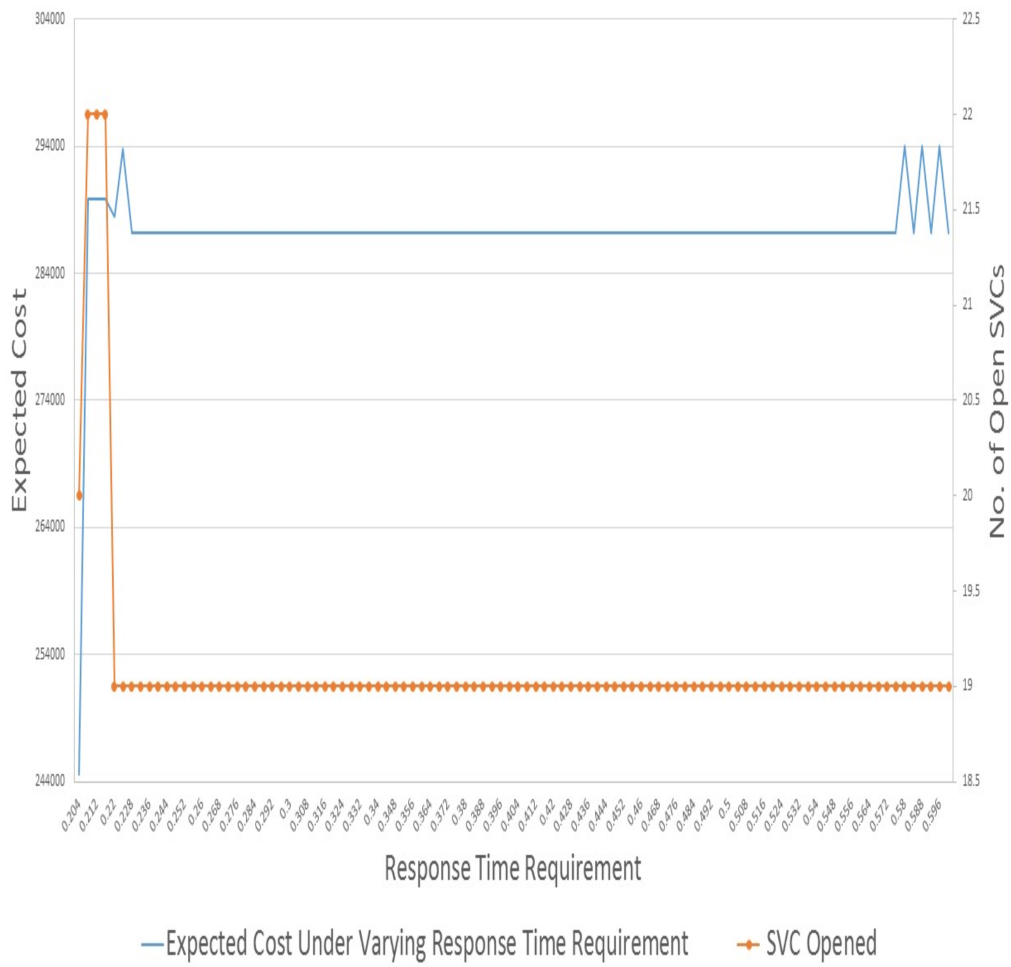

Figure 4.7: Relationship between response time, cost and open SVCs for Model II

4.2.8 Discussion of the trade off between service and cost

With a large value of τ it is optimal to open 19 facilities. As τ approaches it's lower bound, the number of open service increased. As τ is varied along its feasible range, Figure 4.7 shows that the expected costs remains stable. The reason for this is because the number of facilities required to meet demand also remains consistent with varying response time. The fluctuations observed on the graph for expected cost under varying response time imply that cost fluctuate when response time lies between 0.22 and 0.28. The cost also fluctuates when response time lies between 0.576 and 0.6. With these fluctuations the value of the expected costs under varying response require lies in the interval [287190.9578, 293965.0978]. Thus, the gap between the minimum and maximum values of the expected costs is 2.03%. Thus, fluctuations only have little effect on the expected costs. We also see that the fluctuation has no effect on the number of SVCs opened.

4.3 Computational results for model III

This model extends Model II by considering possible SVC failures. Thus, in this section we only consider the effects of falire probability and response time.

4.3.1 Effect of failure probability on model III

We examine how the model behaves as the probability of failure is varied between 0.05 and 0.5. The 37-node dataset is used with (ρ) set to 0.9, d_{max} set to 150 kilometers, and storage capacity set to 5. This result is shown in Figure 4.8.

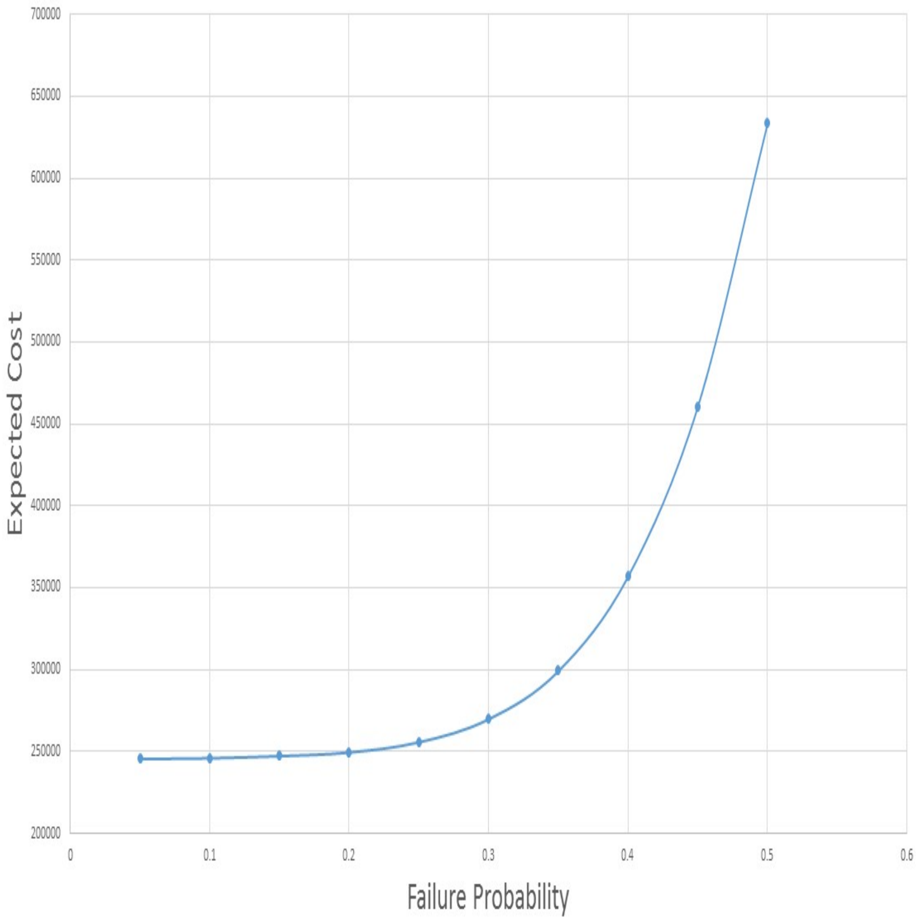

Figure 4.8: Effect of failure probability on model III

4.3.2 Discussion of the effect of failure probability on model III

Figure 4.8 shows that cost remains relatively stable when the failure probability lies between 0.05 and 0.3. As the probabily increases from 0.3, the cost rise is steeper. Thus the optimal decision will be to open service centers whose failure probability lie in the range [0.05-0.3].

4.3.3 Effect of response time on model III

The response time requirement τ is varied between 0.22 and 0.72. The expected cost in each instance is plotted against the response time in Figure 4.9.

4.3.4 Discussion of the effect of response time on model III

It is observed that the expected cost is fairly consistent as the response time varies. This is a consequence of lateral transshipment which ensures that the response time is pool specific. This also agrees with the results obtained for Model I and Model II.

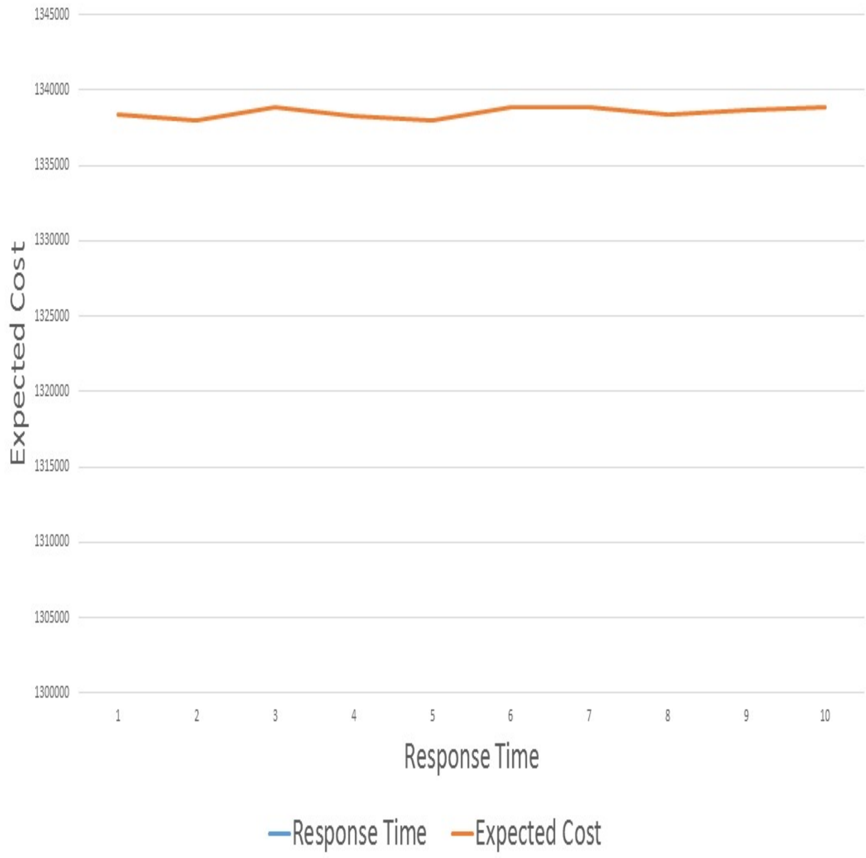

Figure 4.9: Effect of response time on model III

4.4 Computational results for model IV

4.4.1 Effect of response time and cost for model IV

We look at the model behavior as the number of demand scenarios increase from 2 to 20. Each customer has different demand rate for each scenario. For each case the model is solved for two objectives which are; the minimum expected cost and the maximum response time. The objective of maximum response time is chosen to ensure a robust solution. That is, a solution that satisfies all possible response times. The expected cost is plotted against the response time. The result of this experiment is found in Figure 4.10 below.

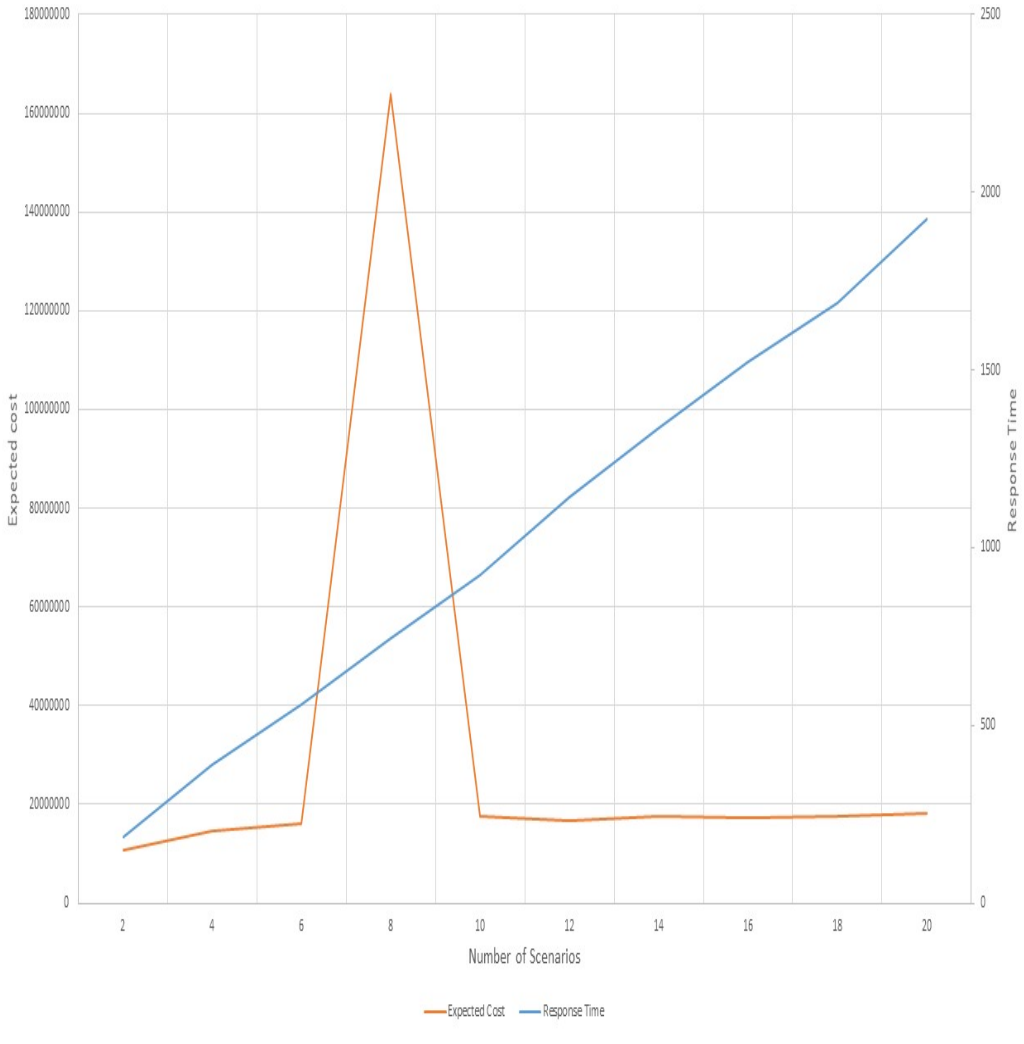

Figure 4.10: Waiting time and cost for model IV

4.4.2 Discussion of the effect of response time and cost for model IV

From Figure 4.10, maximum response time increase as the number of scenarios increase. While the expected cost experiences slight increase as the number of scenarios increase from 2 to 6. As the scenarios increase from 6 to 10 a spike in expected cost is observed. The expected cost is stable when the number of scenarios lie between 10 and 20. The variation observed occurs because of the uncertainty that comes with considering a number of likely scenarios at the same time. The consideration of a number of scenarios each having a probability of occuring, makes the system highly stochastic.

4.4.3 Effect of change in probability on model IV

In this experiment we observe the effect of change in probability on the expected cost. The result is shown in Figure 4.11.

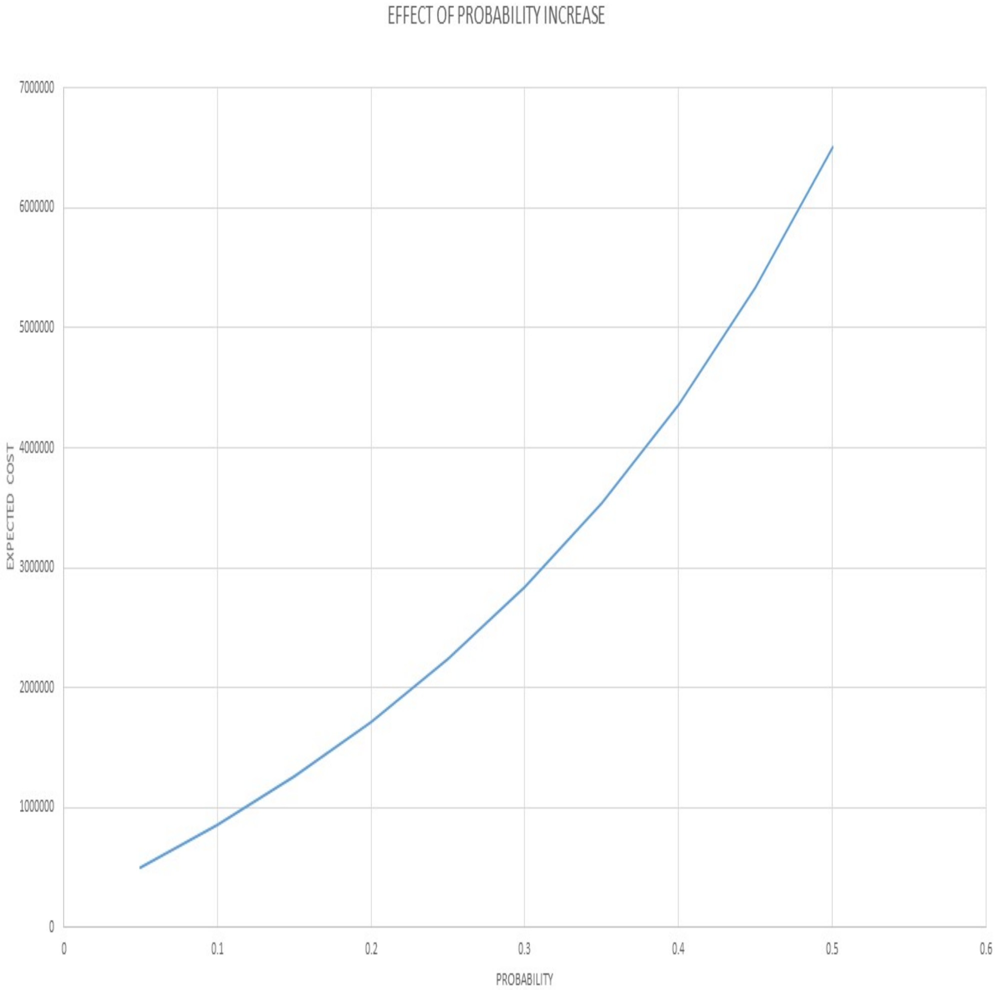

Figure 4.11: Effect of probability change on model IV

4.4.4 Discussion of effect of change in probability on model IV

From Figure 4.11, the cost rises steadily as the probability increases. This is to be expected because higher probabilities certainly mean higher inventory costs.

Chapter 5

SUMMARY AND CONCLUSIONS

5.0 Introduction

The incorporation of lateral transshipment into two-echelon location-inventory systems with response time requirement has been extensively studied in this thesis. Basic formulations for four models which looked at the incorporation of LT into different two-echelon systems were obtained. The steady state distribution for number of items in replenishment was obtained for all models. These distributions along with the properties of a Markovian queue were used to derive the steady state inventory levels for our models. These steady state levels were used to obtain full model formulation for our models which then made it possible to examine the models' properties. Convexity was established for our models using second order conditions. Computational experiments showed that the two-echelon joint location-inventory model with response time requirement and lateral transshipment resulted to lower costs when compared with the model without lateral transshipment . It was also established that lateral transshipment also resulted to stability of expected cost with varying response time requirement.

Conclusively, this study showed that lateral transshipment is an interesting and efficient tool for simultaneously reducing cost and achieving desired customer response time requirement in a two-echelon joint location-inventory system with response time requirement.

5.1 Contribution to knowledge

The contributions to knowledge of this study are enumerated below:

1. This study established a significant contribution by enhancing the literature on two-echelon systems via the incorporation of lateral transshipment into a centralised two-echelon location-inventory system with finite number of facilities at the lower

echelon and response time requirement across all facilities.

2. Steady state distributions for number of items in replenishment to SVCs and the plant were obtained for the system considered.

3. The steady state relationship between expected on-hand inventory level, expected lateral transshipment level and expected backorder level was obtained.

4. Steady state expressions for expected on-hand inventory level, expected lateral transshipment level and expected backorder level were determined.

5. This study presented four new mathematical models for the system, namely, two-echelon inventory model with response time requirement and lateral transshipment, joint two-echelon location-inventory model with response time requirement and lateral transshipment, model with reliable locations, and model with stochastic demand.

6. It was established that the objective functions of the four models and their constraints satisfied the convexity conditions. Thus, the two-echelon location-inventory system with response time requirement and lateral transshipment is a convex optimisation problem and can be solved with convex optimisation solvers.

7. It was established that incorporating lateral transshipment resulted to lower cost when compared with the model without lateral transshipment.

8. The results obtained in this study showed that lateral transshipment is an efficient tool for balancing the contrasting objectives of minimising costs and improving service in a two-echelon joint location-inventory problem with response time requirement

5.2 Recommendations

Our two-echelon system points to some directions for future research. A continuous review (S-1,S) policy was used in this study. Thus, an extension will consider the problem using other inventory policies such as batch ordering policies. Another possible extension will be the relaxation of the assumption of uniform base stock level for all SVCs in a pool. The assumption of negligible transshipment time implied that SVCs in a pool had identical lead times, consequently, the consideration of non negligible lead times will also be an interesting area for further research. Further extension will be the development of specialised heuristics for the various models and compare the solutions obtained with the heuristics with the solutions obtained with GAMS.

REFERENCES

Archibald, T. W. (2007). Modelling replenishment and transshipment decisions in periodic review multilocation inventory systems. *Journal of the Operational Research Society*, 58(7):948–956.

Archibald, T. W., Black, D., and Glazebrook, K. D. (2009). An index heuristic for transshipment decisions in multilocation inventory systems based on a pairwise decomposition. *European Journal of Operational Research*, 192:69–78.

Axsater, S. (1990). Modelling emergency lateral transshipments in inventory systems. *Management Science*, 36(11):1329–1338.

Bendoly, E. (2004). Integrated inventory pooling for firms servicing both on-line and store demand. *Computers and Operations Research*, 31:1465–1480.

Benjaafar, S., Li, Y., Xu, D., and Elhedhli, S. (2008). Demand allocation in systems with multiple inventory locations and multiple demand sources. *Manufacturing and Operations Management*, 10(1):43–60.

Bonami, P. KilinÃğ, M., Linderoth, L., Lee, J., and Leyffer, S. (2012). Mixed integer nonlinear programming. *The IMA Volumes in Mathematics and its Applications. New York: Springer-Verlag*, 154.

Boyd, S. and Vandenberghe, L. (2004). Convex optimization. *Cambridge: Cambridge University Press*.

Buzacott, J. A. and Shanthikumar, J. G. (1993). Stochastic models of manufacturing systems. *New Jersey: Prentice Hall.*

Caglar, D. (2001). A multi-echelon spare parts inventory system with emergency lateral shipments subject to a response time constraint. *Ph.D Thesis, Northwestern University, Evanston, IL.*

Caglar, D., Li, C. L., and Simchi-Levi, D. (2004). Two echelon spare parts inventory system subject to a service constraint. *IIE Transactions*, 36:655–666.

Candas, M. F. and Kutanoglu, E. (2007). Benefits of considering inventory in service parts logistic network design problems with time-based service constraints. *IIE Transactions*, 39(3):159–176.

Chen, G., Daskin, M. S., Shen, Z. J., and Uryasev, S. (2006). The α-reliable mean-excess regret model for stochastic facility location modeling. *Naval Research Logistics*, 53:617–626.

Chen, Q., Xiaopeng, L., and Quyang, Y. (2011). Joint inventory-location problem under the risk of probabilistic facility disruptions. *Transportation Research*, Part B(45):991–1003.

Cohen, M. A., Kliendorfer, P. A., and Lee, H. L. (1989). Near-optimal service constrained stocking policies for spare parts. *Operations Research*, 37:104–117.

Daskin, M. (1995). Network and discrete location: Models, algoriths, and applications. *New York: John Wiley & Sons.*

Daskin, M., Coullard, C., and Shen, Z. J. (2002). An inventory-location model: Formulation, solution algorithm and computational results. *Annals of Operations Research*, 110(1-4):83–106.

Dong, L. and Rudi, N. (2004). Who benefits from transshipment? exogenous vs. endogenous wholesale prices. *Management Science*, 50(5):645–657.

Drezner, Z. (1995). Facility location: A survey of applications and methods. *New York: Springer*.

Drezner, Z. and Hamacher, H. W. (2002). Facility location: Applications and theory. *New York: Springer*.

Eppen, G. D. (1979). NoteâĂŤeffects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5):498–501.

Eskigun, E., Uzsoy, R., Preckel, P. V., Beaujon, G., Krishnan, S., and Tew, J. D. (2007). Outbound supply chain network design with mode selection and lead time considerations. *Naval Research Logistics*, 54:282–300.

Everett, H. (1965). Generalized lagrange multiplier method for solving problems of optimal allocation of resources. *Operations Research*, 11(3):399–417.

Gowers, T., Barrow-Green, J., and Leader, I. (2008). The princeton companion to mathematics. *Princeton: Princeton University Press*.

Grahovac, J. and Chakravarty, A. (2001). Sharing and lateral transshipment of inventory in a supply chain with expensive low demand items. *Management Science*, 47(4):579–594.

Graves, S. (1985). A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science.*, 40(567-602).

Herer, Y. T. and Tzur, M. (2001). The dynamic transshipment problem. *Naval Research Logistics*, 48:386–408.

Herer, Y. T. and Tzur, M. (2003). Optimal and heuristic algorithms for the multi-location dynamic transshipment problem with fixed transshipment costs. *IIE Transactions*, 5:419–432.

Herer, Y. T., Tzur, M., and Yucesan, E. (2006). The multi-location transshipment problem. *IIE Transactions*, 38:185–200.

Herer, Y. T., Tzur, M., and Yucesanc, E. (2002). Transshipments: An emerging inventory recourse to achieve supply chain leagility . *International Journal of Production Economics.*, 80(3):201–212.

Hu, J., Watson, W., and Schneider, H. (2005). Approximate solutions for multi-location inventory systems with transshipments. *International Journal of Production Economics*, 97:31–43.

Huo, J. and Li, H. (2007). Batch ordering policy of multi-location spare parts inventory system with emergency lateral transshipments. *Systems Engineering - Theory & Practice*, 27(12):62–67.

Jung, B., Sun, B., Kim, J., and Ahn, S. (2003). Modeling lateral transshipments in multiechelon repairable-item inventory systems with finite repair channels. *Computers & Operations Research*, 30:1401–1417.

Kochel, P. (1996). On queueing models for some multi-location problems. *International Journal of Production Economics*, 45:429–433.

Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J., and Schade, K. (2018). A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3):955–983.

Kranenburg, A. A. (2006). Spare parts inventory control under system availability constraints. *PhD dissertation, Eindhoven University of Technology.*

Krishnan, K. and Rao, V. (1965). Inventory control in n warehouses. *Journal of Industrial Engineering*, XVI(3):212–215.

Kruse, K. (1981). Waiting time in continuous review (s,s) inventory systems with constant lead times. *Operations Research.*, 29:202–207.

Kukreja, A. and Schmidt, C. P. (2005). A model for lumpy demand parts in a multi-location inventory system with transshipments. *Computers & Operations Research*, 32(8):2059–2075.

Kukreja, A., Schmidt, C. P., and Miller, D. M. (2001). Stocking decisions for low-usage items in a multilocation inventory system. *Management Science*, 47(10):1371–1383.

Kutanoglu, E. (2008). Insights into inventory sharing in service parts logistics systems with time-based service levels. *Computers & Industrial Engineering*, 54:341–358.

Kutanoglu, E. and Mahajan, M. (2009). An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels. *European Journal of Operational Research*, 194:728–742.

Lee, H. (1987). A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33(10):1302–1316.

Little, J. D. (1961). A proof for the queiung formula: $l = \lambda w$. *Operations Research*, 9(3):383–387.

Mak, H. Y. and Shen, Z. J. (2009). A two echelon-inventory location problem with service considerations. *Naval Research Logistics*, 56(8).

Moinzadeh, K. and Lee, H. L. (1986). Batch size and stocking levels in multi-echelon reparable systems. *Management Science*, 32(12):1567–1581.

Muckstadt, J. (2005). Analysis and algorithm for service parts supply chain. *New York: Springer.*

Muckstadt, J. A. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory system. *Institute for Operations Research and the Management Sciences (INFORMS)*, 20(4):472–481.

Nonas, L. M. and Jornsten, K. (2007). Optimal solutions in the multi-location inventory system with transshipments. *Journal of Mathematical Modelling and Algorithms*, 6:47–75.

Nozick, L. K. and Turnquist, M. A. (2001). A two-echelon inventory allocation and distribution center location analysis. *Transportation Research Part E*, 37(6):425–441.

Ozdemir, D., Yucesan, E., and Herer, Y. T. (2006). Multi-location transshipment problem with capacitated transportation. *European Journal of Operational Research*, 175:602–621.

Ozsen, L., Daskin, M., and Coullard, C. (2008). Capacitated warehouse location model with risk pooling. *Wiley InterScience (www.interscience.wiley.com).*

Ozsen, L., Daskin, M., and Coullard, C. (2009). Facility location modeling and inventory management with multi-sourcing. *Transportation Science*, 43(4):455–472.

Palm, C. (1938). Analysis of the erlang traffic formulae for busy- signal arrangements. *Ericsson Technics*, 4:39–58.

Paterson, C., Teunter, R., and Glazebrook, K. (2012). Enhanced lateral transshipments in a multi-location inventory system. *European Journal of Operational Research*, 221:317–327.

Puga, M. S. and Tancrez, J. S. (2016). A heuristic algorithm for solving large location-inventory problems with demand uncertainty. *European Journal of Operational Research*.

Riaz, M. W. (2013). Two-echelon supply chain design for spare parts with time constraints. *Thesis, University of Waterloo, UWSpace. http://hdl.handle.net/10012/7914*.

Robinson, L. W. (1990). Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. *Operations Research*, 38(2):278–295.

Ross, S. M. (2010). Introduction to probability models. 10^{th} *ed. San Diego, California: Academic Press.*

Rui, Y. (2015). The capacitated reliable fixed-charge location problem: Model and algorithm. *Theses and Dissertations. Paper*, 1684.

Serra, D. and Marianov, V. (1998). The p-median problem in a changing network: the case of barcelona. *Location Science*, 6:383–394.

Shebrooke, C. (1968). A multi-echelon technique for recoverable item control. *Operations Research*, 16:122–141.

Shen, Z. J. (2006a). A profit-maximizing supply chain network design model with demand choice flexibility. *Operations Research Letters*, 34:673–682.

Shen, Z. J. (2006b). Integrated supply chain design models: A survey and future research directions. *Journal of Industrial and Management Optimization*, 3(1):1–27.

Shen, Z. J., Coullard, C., and Daskin, M. S. (2003). A joint location-inventory model. *Transportation Science*, 37:40–55.

Shen, Z. J. and Daskin, M. S. (2005). Trade-offs between customer service and cost in integrated supply chain design. *Manufacturing and Service Operations Management*, 7(3):188–207.

Shen, Z. J. and Qi, L. (2006). Incorporating inventory and routing costs in strategic location models. *European Journal of Operational Research*, 179:372–389.

Sherbrooke, C. (2004). Optimal inventory modeling of systems. *Norwell,MA.: Kluwer Academic Publishers.*

Sherbrooke, C. C. (1986). Vari-metric: Improved approximations for multi-indenture, multi-echelon availability models. *Operations Research*, 34(2):311–319.

Sherbrooke, C. C. (1992). Multi-echelon inventory systems with lateral supply. *Naval Research Logistics*, 39:29–40.

Shu, J., Teo, C., and Shen, Z. M. (2005). Stochastic transportation-inventory network design problem. *Operations Research*, 53:48–60.

Snyder, L. (2006). Facility location under uncertainty: a review. *IIE Transactions*, 38:537–554.

Snyder, L. and Daskin, M. S. (2005). Reliability models for facility location: The expected failure cost case. *Transportation Science*, 39(3):400–416.

Sourirajan, K., Ozsen, L., and Uzsoy, R. (2007). A single-product network design mode with lead time and safety stock considerations. *IIE Transactions*, 39(5):411–424.

Tagaras, G. (1999). Pooling in multi-location periodic inventory distribution systems. *Omega International Journal of Management Science*, 27:39–59.

Tagaras, G. and Cohen, M. A. (1992). Pooling in two-location inventory systems with non-negligible replenishment lead times. *Management Science*, 38(8):1067–1083.

Tiacci, L. and Saetta, S. (2011). Reducing the mean supply delay of spare parts using lateral transshipments policies. *Int. J.Production Economics*, 133:182–191.

Tijms, H. C. (2003). A first course in stochastic models. *The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons Ltd*.

Utterbeeck, F. V., Wong, H., Oudheusden, D. V., and Cattrysse, D. (2009). The effects of resupply flexibility on the design of service parts supply systems. *Transportation Research Part E*, 45:72–85.

Wee, K. E. and Dada, M. (2005). Optimal policies for transshipping inventory in a retail network. *Management Science*, 51(10):1519–1533.

Winston, W. L. (2004). Operations research applications and algorithms. 4^{th} *ed. United States: Thomson Brooks/Cole*.

Wong, H., Cattrysse, D., and van Oudheusden, D. (2006). Inventory pooling of repairable spare parts with nonzero lateral transshipment time and delayed lateral transshipments . *European Journal of Operational Research*, 165(1):207–218.

Wong, H., van Houtum, G. J., Cattrysse, D., and van Oudheusden, D. (2005). Stocking decisions for repairable spare parts pooling in a multi-hub system. *International Journal of Production Economics*, 93-94:309–317.

Yanagi, S. and Sasaki, M. (1992). An approximation method for the problem of a repairable-item inventory system with lateral resupply. *IMA Journal of Mathematics Applied in Business and Industry*, 3:305–314.

Yang, G., Dekker, R., Gabor, A. F., and Axsater, S. (2013). Service parts inventory control with lateral transshipment and pipeline stock flexibility. *International Journal of Production Economics*, 142:278–289.

Yang, J. and Qin, Z. (2007). Capacitated production control with virtual lateral transshipments. *Operations Research*, 55(6):1104–1119.

Zhang, J. (2005). Transshipment and its impact on supply chain's member performance. *Management Science*, 51(10):1534–1539.

APPENDICES

Appendix (I) Data sets

The data sets used in this study are provided in this section. The 37 nodes data set comprises of the most populous cities in each of the 36 state capitals in Nigeria and the Federal Capital Territory (FCT). The 109 nodes data set comprises of the 3 most populous cities in each of the 36 state capitals in Nigeria and the FCT.The 181 nodes data set comprises of the 5 most populous cities in each of the 36 state capitals in Nigeria and the FCT. The data sets are presented in Tables 6.1-6.8.

Table 5.1: 37 nodes data set

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Ado Ekiti	1	313690	3	7.61261	5.27384	10000
Ajeromi-Ifelodun	2	6873316	10	6.4555	3.3339	30000
Ifo	3	539170	5	6.78197	3.25195	12000
Akoko	4	701785	7	7.641	5.79399	11500
Ife	5	643582	6	7.47442	4.55933	12500
Ibadan	6	3034200	10	7.39639	3.91667	20000
Uyo	7	427873	4	5.033	7.917	25000
Yenagoa	8	352285	4	5.033	6.333	20000
Calabar	9	375196	4	4.95	8.325	23000
Warri	10	564658	6	5.517	5.75	18000
Benin	11	1495800	10	6.33333	5.62222	15000
Portharcourt	12	1005904	10	4.73292	6.96919	27000
Aba	13	534265	5	5.117	7.367	18000
Onitsha	14	7425000	10	6.132942	6.792399	18500
Afikpo	15	314191	3	5.89258	7.93534	15000
Enugu	16	717291	7	6.4526667	7.5103333	20000
Owerri	17	403425	4	5.485	7.035	19000
Abuja	18	776298	8	9.0765	7.3986	40000
Gboko	18	361325	4	7.325	9.005	16000
Okene	20	325623	3	7.48333	6.21667	15000
Illorin	21	777667	8	8.43005	4.426	18000
Lafia	22	329922	3	8.64007	8.69959	19000
Minna	23	348788	3	9.583555	6.546316	17000
Jos	24	900000	9	9.933	8.883	18000
Yola	25	395871	4	9.23	12.46	16000
Bauchi	26	493730	5	10.5	10	16000
Maiduguri	27	540016	5	11.833	13.15	18500
Akko	28	337435	3	10.09994	11.04833	15000
Gassol	29	245086	2	8.47269	10.5442	13000
Fune	30	301954	3	11.85212	11.44862	16500
Birnin Kudu	31	314108	3	11.49023	9.49368	13500
Kaduna	32	767306	8	10.52306	7.44028	25000
Kano	33	2828861	10	12	8.517	30000
Katsina	34	459022	5	12.983	7.6	19000
Birnin Kebbi	35	268620	3	12.45832	4.26267	14000
Sokoto	36	427760	4	13.067	5.233	18000
Gusau	37	383712	4	11.87555	6.61984	13000

Table 5.2: 109 nodes data set I

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Ado Ekiti	1	313690	3	7.61261	5.27384	10000
Ijero	2	221873	2	7.84582	5.05686	9600
Ekiti West	3	179600	2	7.68663	5.02847	9000
Ajeromi-Ifelodun	4	6873316	10	6.4555	3.3339	30000
Alimosho	4	1319571	10	6.61056	3.29583	28000
Kosofe	5	682772	7	6.598781	3.409463	20000
Ifo	6	539170	5	6.78197	3.25195	12000
Ado-Odo/Ota	7	527242	5	6.62366	3.08626	11500
Ijebu	8	458641	5	6.99551	3.97085	13500
Akoko	10	701785	7	7.641	5.79399	11500
Akure	11	491033	5	7.26382	5.31136	12000
Ondo	12	364960	4	7.09145	4.96053	11000
Ife	13	643582	6	7.47442	4.55933	12500
Illesha	14	212225	2	7.60146	4.7618	11000
Iwo	15	191348	2	7.67548	4.14127	9000
Ibadan	16	3034200	10	7.39639	3.91667	20000
Saki	17	382225	4	8.70467	3.58944	12000
Ibaraba	18	320718	3	7.43737	3.26761	10000
Uyo	19	427873	4	5.033	7.917	25000
Essien Udim	20	193257	2	5.13333	7.6833	15000
Ibiono Ibom	21	188605	2	5.23333	7.88333	12000
Yenagoa	22	352285	4	5.033	6.333	20000
Southern Ijaw	23	321808	3	4.7	5.967	12000
Ekeremor	24	269588	3	5.05	5.783	12500
Calabar	25	375196	4	4.95	8.325	23000
Akpabuyo	26	272262	3	4.880791	8.528161	18000
Odunkpani	27	192884	2	5.081238	8.349917	15000
Warri	28	564658	6	5.517	5.75	18000
Ughelli	29	533325	5	5.5	5.983	16000
Asaba	30	514679	5	6.1978417	6.7284667	17000
Benin	31	1495800	10	6.33333	5.62222	15000
Oredo	32	374515	4	6.23581	5.55114	12000
Ikpobah-Okha	33	372080	4	6.16445	5.62284	11000
Portharcourt	34	1005904	10	4.73292	6.96919	27000
Obio/Akpor	35	462350	5	4.83153	6.98906	20000
Ahoada	36	415556	4	5.0858	6.47089	17000
Aba	37	534265	5	5.117	7.367	18000
Umuahia	38	359230	4	5.524953	7.492241	16000
Isiala ngwa	39	290773	3	5.286819	7.416505	14500
Onitsha	40	7425000	10	6.132942	6.792399	18500
Idemili	41	637821	6	6.123656	6.947753	17000
Nnewi	42	391227	4	6.010519	6.910345	16300
Afikpo	43	314191	3	5.89258	7.93534	15000
Izzi	44	236679	2	6.48453	8.29468	14000
Onicha	45	236609	2	6.09899	7.84133	14000
Enugu	46	717291	7	6.4526667	7.5103333	20000
Nsukka	47	309448	3	6.85329	7.34801	17000
Igbo Eze North	48	258829	3	6.98333	7.45	14000
Owerri	49	403425	4	5.485	7.035	19000
Ideato	50	315815	3	5.88537	7.13136	16000

Table 5.3: 109 nodes data set II

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Isiala-Mbano	51	197921	2	5.66767	7.20338	15000
Abuja	52	776298	8	9.0765	7.3986	40000
Gboko	53	361325	4	7.325	9.005	16000
Makurdi	54	300377	3	7.73056	8.53611	17000
Gwer	55	290973	3	7.3	8.48333	14000
Okene	56	325623	3	7.48333	6.21667	15000
Ankpa	57	266176	3	7.44848	7.63174	12000
Dekina	58	260968	3	7.58435	7.17344	1000
Illorin	59	777667	8	8.43005	4.426	18000
Baruten	60	206679	2	9.26039	3.31607	9000
Edu	61	201642	2	8.89181	5.21012	8000
Lafia	62	329922	3	8.64007	8.69959	19000
Karu	63	216230	2	9.04694	7.76364	20000
Nasarawa	64	187220	2	8.5	8.25	15000
Minna	65	348788	3	9.583555	6.546316	17000
Mashegu	66	215197	2	9.92993	5.23385	15000
Suleja	67	215075	2	9.19085	7.15184	18000
Jos	68	900000	9	9.933	8.883	18000
Mangu	69	300520	3	9.39052	9.17968	16000
Langtang	70	247489	2	9.05651	9.82247	15000
Yola	71	395871	4	9.23	12.46	16000
Mubi	72	281471	3	10.267	13.267	13000
Fufore	73	209460	2	9.217	12.65	12000
Bauchi	74	493730	5	10.5	10	16000
Ningi	75	385997	4	11.067	9.567	14500
Toro	76	346000	3	10.059584	9.070932	12000
Maiduguri	77	540016	5	11.833	13.15	18500
Gwoza	78	276568	3	11.08611	13.69139	16000
Bama	79	270119	3	11.521831	13.688377	14500
Akko	80	337435	3	10.09994	11.04833	15000
Gombe	81	266844	3	10.28263	11.16674	16500
Yamaltu/Deba	82	255726	3	10.23814	11.44152	14000
Gassol	83	245086	2	8.47269	10.5442	13000
Wukari	84	238283	2	7.96327	9.84767	13500
Sardauna	85	224357	2	6.87463	11.21184	12000
Fune	86	301954	3	11.85212	11.44862	16500
Jakusko	87	232458	2	1245605	10.91226	14000
Potiskum	88	204866	2	11.71064	11.15721	12000
Birnin Kudu	89	314108	3	11.49023	9.49368	13500
Gwaram	90	271368	3	11.1858	9.9686	12000
Kafin Hausa	91	267284	3	12.14958	10.00247	11000
Kaduna	92	767306	8	10.52306	7.44028	25000
Zaria	93	698163	7	10.98854	7.69622	20000
Igabi	94	430753	4	10.781	7.504	18000
Kano	95	2828861	10	12	8.517	30000
Nasarawa	96	596411	6	8.5	8.25	19000
Dala	97	418759	4	12.017	8.483	15000
Katsina	98	459022	5	12.983	7.6	19000
Kankara	99	243259	2	11.97384	7.36266	14000
Funtua	100	225156	2	11.47196	7.30699	13000

Table 5.4: 109 nodes data set III

Birnin Kebbi	101	268620	3	12.45832	4.26267	14000
Wasagu/Danko	102	265271	3	11.42938	5.6749	12000
Bagudo	103	238014	2	11.34757	3.95674	11000
Sokoto	104	427760	4	13.067	5.233	18000
Gada	105	249052	2	13.74739	5.65521	16000
Gwadabawa	106	231569	2	13.44757	5.3106	15000
Gusau	107	383712	4	11.87555	6.61984	13000
Zurmi	108	293977	3	12.81147	6.7938	12000
Maru	109	293141	3	11.53807	6.27888	1000

Table 5.5: 181 nodes data set I

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Ado Ekiti	1	313690	3	7.61261	5.27384	10000
Ijero	2	221873	2	7.84582	5.05686	10000
Ekiti West	3	179600	2	7.68663	5.02847	10000
Ikole	4	170414	2	7.89347	5.5111	10000
Ekiti south west	5	165277	2	7.51325	5.05179	10000
Ajeromi-Ifelodun	7	6873316	10	6.4555	3.3339	10000
Alimosho	6	1319571	10	6.61056	3.29583	10000
Kosofe	8	682772	7	6.598781	3.409463	10000
Mushin	9	631857	6	6.535233	3.348967	10000
Oshodi-Isolo	10	629061	6	6.535498	3.308678	10000
Ifo	11	539170	5	6.78197	3.25195	10000
Ado-Odo/Ota	12	527242	5	6.62366	3.08626	10000
Ijebu	13	458641	5	6.99551	3.97085	10000
Abeokuta	14	449088	4	7.23079	3.16845	10000
Yewa(Egbado)	15	352180	4	7.14729	2.91795	10000
Akoko	16	701785	7	7.641	5.79399	10000
Akure	17	491033	5	7.26382	5.31136	10000
Ondo	18	364960	4	7.09145	4.96053	10000
Illaje	19	289838	3	6.20323	4.76788	10000
Okitipupa	20	234138	2	6.54797	4.69894	10000
Ife	21	643582	6	7.47442	4.55933	10000
Illesha	22	212225	2	7.60146	4.7618	10000
Iwo	23	191348	2	7.67548	4.14127	10000
Ede	24	159307	2	7.7069	4.50922	10000
Osogbo	25	155507	2	7.75963	4.57625	10000
Ibadan	26	3034200	10	7.39639	3.91667	10000
Saki	27	382225	4	8.70467	3.58944	10000
Ibaraba	28	320718	3	7.43737	3.26761	10000
Ona-Ara	29	265571	3	7.28333	4.03333	10000
Oyo	30	260552	3	7.87878	4.02132	10000
Uyo	31	427873	4	5.033	7.917	10000
Essien Udim	32	193257	2	5.13333	7.6833	10000
Ibiono Ibom	33	188605	2	5.23333	7.88333	10000
Eket	34	172856	2	4.65	7.933	10000
Etinan	35	168924	2	4.85	7.83333	10000
Yenagoa	36	352285	4	5.033	6.333	10000
Southern Ijaw	37	321808	3	4.7	5.967	10000
Ekeremor	38	269588	3	5.05	5.783	10000
Sagbama	39	186869	2	5.167	6.2	10000
Brass	40	184127	2	4.315	6.24167	10000
Calabar	41	375196	4	4.95	8.325	10000
Akpabuyo	42	272262	3	4.880791	8.528161	10000
Odunkpani	43	192884	2	5.081238	8.349917	10000
Boki	44	186611	2	6.27389	9.01	10000
Obubra	45	172543	2	6.08333	8.33333	10000
Warri	46	564658	6	5.517	5.75	10000
Ughelli	47	533325	5	5.5	5.983	10000
Asaba	48	514679	5	6.1978417	6.7284667	10000
Ethiope	49	403654	4	5.678246	5.962111	10000
Sapele	50	142652	1	5.9	5.667	10000

Table 5.6: 181 nodes data set II

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Benin	51	1495800	10	6.33333	5.62222	10000
Oredo	52	374515	4	6.23581	5.55114	10000
Ikpobah-Okha	53	372080	4	6.16445	5.62284	10000
Egor	54	340287	3	6.35748	5.57547	10000
Akoko-Edo	55	261567	3	7.34506	6.11489	10000
Portharcourt	56	1005904	10	4.73292	6.96919	10000
Obio/Akpor	57	462350	5	4.83153	6.98906	10000
Ahoada	58	415556	4	5.0858	6.47089	10000
Khana	59	292924	3	4.69962	7.42264	10000
Abua-Oduai	60	282410	3	4.82977	6.56739	10000
Aba	61	534265	5	5.117	7.367	10000
Umuahia	62	359230	4	5.524953	7.492241	10000
Isiala ngwa	63	290773	3	5.286819	7.416505	10000
Ohafia	64	245987	2	5.617	7.833	10000
Osisioma ngwa	65	220662	2	5.14972	7.33028	10000
Onitsha	66	7425000	10	6.132942	6.792399	10000
Idemili	67	637821	6	6.123656	6.947753	10000
Nnewi	68	391227	4	6.010519	6.910345	10000
Awka	69	301846	3	6.222004	7.082116	10000
Ihiala	70	302277	3	5.851644	6.851181	10000
Afikpo	71	314191	3	5.89258	7.93534	10000
Izzi	72	236679	2	6.48453	8.29468	10000
Onicha	73	236609	2	6.09899	7.84133	10000
Ikwo	74	214969	2	6.05316	8.16284	10000
Abakaliki	75	149683	1	6.32485	8.11368	10000
Enugu	76	717291	7	6.4526667	7.5103333	10000
Nsukka	77	309448	3	6.85329	7.34801	10000
Igbo Eze North	78	258829	3	6.98333	7.45	10000
Udi	79	238305	2	6.51181	7.35535	10000
Igbo-Etiti	80	208333	2	6.68151	7.41959	10000
Owerri	81	403425	4	5.485	7.035	10000
Ideato	82	315815	3	5.88537	7.13136	10000
Isiala-Mbano	83	197921	2	5.66767	7.20338	10000
Abo-Mbaise	84	194779	2	5.42549	7.2518	10000
Ahiazu-Mbaise	85	170824	2	5.54639	7.27135	10000
Abuja	86	776298	8	9.0765	7.3986	10000
Gboko	87	361325	4	7.325	9.005	10000
Makurdi	88	300377	3	7.73056	8.53611	10000
Gwer	89	290973	3	7.3	8.48333	10000
Oturkpo	90	266411	3	7.19306	8.14639	10000
Kwande	91	248642	2	6.80099	9.47021	10000
Okene	92	325623	3	7.48333	6.21667	10000
Ankpa	93	266176	3	7.44848	7.63174	10000
Dekina	94	260968	3	7.58435	7.17344	10000
Okehi	95	223574	2	7.68192	6.28543	10000
Lokoja	96	196643	2	8.20488	6.56305	10000
Illorin	97	777667	8	8.43005	4.426	10000
Baruten	98	206679	2	9.26039	3.31607	10000
Edu	99	201642	2	8.89181	5.21012	10000
Ifelodun	100	204975	2	7.91667	4.66667	10000

Table 5.7: 181 nodes data set III

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Asa	101	124668	1	8.43005	4.426	10000
Lafia	102	329922	3	8.64007	8.69959	10000
Karu	103	216230	2	9.04694	7.76364	10000
Nasarawa	104	187220	2	8.5	8.25	10000
Obi	105	148405	1	8.34738	8.72769	10000
Nasarawa Eggon	106	148405	1	8.74947	8.44035	10000
Minna	107	348788	3	9.583555	6.546316	10000
Mashegu	108	215197	2	9.92993	5.23385	10000
Suleja	109	215075	2	9.19085	7.15184	10000
Shiroro	110	235665	2	10.11543	6.68307	10000
Mokwa	111	242858	2	9.2	5.33333	10000
Jos	112	900000	9	9.933	8.883	10000
Mangu	113	300520	3	9.39052	9.17968	10000
Langtang	114	247489	2	9.05651	9.82247	10000
Shendam	115	205119	2	8.71667	9.5	10000
Quaan Pan	116	197276	2	8.76078	9.16142	10000
Yola	117	395871	4	9.23	12.46	10000
Mubi	118	281471	3	10.267	13.267	10000
Fufore	119	209460	2	9.217	12.65	10000
Song	120	195188	2	9.82444	12.625	10000
Demsa	121	178407	2	9.455541	12.152552	10000
Bauchi	122	493730	5	10.5	10	10000
Ningi	123	385997	4	11.067	9.567	10000
Toro	124	346000	3	10.059584	9.070932	10000
Alkaleri	125	328284	3	9.883	10.5	10000
Katagum	126	293020	3	12.283	10.35	10000
Maiduguri	127	540016	5	11.833	13.15	10000
Gwoza	128	276568	3	11.08611	13.69139	10000
Bama	129	270119	3	11.521831	13.688377	10000
Ngala	130	236498	2	12.342068	14.185827	10000
Damboa	131	233200	2	11.15	12.75	10000
Akko	132	337435	3	10.09994	11.04833	10000
Gombe	133	266844	3	10.28263	11.16674	10000
Yamaltu	134	255726	3	10.23814	11.44152	10000
Funakaye	135	237687	2	10.74227	11.40843	10000
Balanga	136	211490	2	9.8096	11.78623	10000
Gassol	137	245086	2	8.47269	10.5442	10000
Wukari	138	238283	2	7.96327	9.84767	10000
Sardauna	139	224357	2	6.87463	11.21184	10000
Bali	140	211024	2	8.15541	10.96853	10000
Karim Lamido	141	193924	2	9.22208	10.86536	10000
Fune	142	301954	3	11.85212	11.44862	10000
Jakusko	143	232458	2	1245605	10.91226	10000
Potiskum	144	204866	2	11.71064	11.15721	10000
Geidam	145	155740	2	12.65454	12.06733	10000
Nguru	146	150699	2	12.97973	10.39914	10000
Birnin Kudu	147	314108	3	11.49023	9.49368	10000
Gwaram	148	271368	3	11.1858	9.9686	10000
Kafin Hausa	149	267284	3	12.14958	10.00247	10000
Dutse	150	251135	3	11.80331	9.30708	10000

Table 5.8: 181 nodes data set IV

Cities	Nodes	Population	Demand	Latitude	Longitude	Fixed cost
Jahun	151	229882	2	12.07996	9.55457	10000
Kaduna	152	767306	8	10.52306	7.44028	10000
Zaria	153	698163	7	10.98854	7.69622	10000
Igabi	154	430753	4	10.781	7.504	10000
Lere	155	339740	3	10.34643	8.56734	10000
Chikun	156	372272	4	1o.315	7.274	10000
Kano	157	2828861	10	12	8.517	10000
Nasarawa	158	596411	6	8.5	8.25	10000
Dala	159	418759	4	12.017	8.483	10000
Gwale	160	357827	4	11.98211	8.49818	10000
Kumbotso	161	294391	3	11.90063	8.51633	10000
Katsina	162	459022	5	12.983	7.6	10000
Kankara	163	243259	2	11.97384	7.36266	10000
Funtua	164	225156	2	11.47196	7.30699	10000
Daura	165	224884	2	12.98195	8.2516	10000
Kafur	166	209360	2	11.67376	7.68971	10000
Birnin Kebbi	167	268620	3	12.45832	4.26267	10000
Wasagu/Danko	168	265271	3	11.42938	5.6749	10000
Bagudo	169	238014	2	11.34757	3.95674	10000
Argungu	170	200248	2	12.68795	4.40358	10000
Jega	171	197757	2	12.16397	4.48643	10000
Sokoto	172	427760	4	13.067	5.233	10000
Gada	173	249052	2	13.74739	5.65521	10000
Gwadabawa	174	231569	2	13.44757	5.3106	10000
Tambuwal	175	225917	2	12.36609	4.84882	10000
Sabon Birni	176	207470	2	13.4847	6.26409	10000
Gusau	177	383712	4	11.87555	6.61984	10000
Zurmi	178	293977	3	12.81147	6.7938	10000
Maru	179	293141	3	11.53807	6.27888	10000
Kaura Namoda	180	285363	3	12.56195	6.57617	10000
Tsafe	181	266929	3	11.88251	6.8947	10000

Appendix (II) model I GAMS code

```
$ontext
* EVALUATING GAMS
$offtext
* Turn on the default end of line comment Character !!
$onEolcom
* Define an end of line Comment Character of your choice to be used
* $eolcom %%
* Declaration Of Sets to Be used
Sets
  V     'Set of Customers' /1*37/
* V     'Set of Candidate SVC Locations' /1*37/
  W     'Set of Pools'   /w1*w6/
*DYNAMIC SETS (ACTING AS SUBSETS OF CUSTOMERS/SvCs FOR EACH POOL)
  v1(V) 'Set of cities in pool 1' /1*6/
  v2(V) 'Set of cities in pool 2' /7*12/
  v3(V) 'Set of cities in pool 3' /13*17/
  v4(V) 'Set of cities in pool 4' /18*24/
  v5(V) 'Set of cities in pool 5' /25*30/
  v6(V) 'Set of cities in pool 6' /31*37/
* THE SAME THING IS DONE FOR POOLS
  w1(w) 'Set of cities in pool 1' /w1/
  w2(w) 'Set of cities in pool 2' /w2/
  w3(w) 'Set of cities in pool 3' /w3/
  w4(w) 'Set of cities in pool 4' /w4/
  w5(w) 'Set of cities in pool 5' /w5/
  w6(w) 'Set of cities in pool 6' /w6/

sets
* Data sets containing the lattitude, longitude and demand for cities
  Data1 'Raw Data Sets To Be entered' /Latitude, Longitude, Demand/;
* Declare the set indices for U, V and W  as same as i, j and pool
*----------------------------------------------------------------------
* Declaration of Parameters to be used
* Note That Gams is not case sensitive,thus f(V,W) is same as f(v,w)
*----------------------------------------------------------------------
Parameters
```

```
 h0 'is the per unit holding cost of each unit of inventory at SV Cv
 *in Poolw per unit time'
 p(v,w)'is the per unit cost of backorder per unit inventory for each
 *unit of time'
 q(v,w)     'is the LT cost per unit inventory'
 lambu(v)   'Demand rate of customer u'
 lambw(w)   'Demand rate (Poisson) at Pool $w$'
 lamb0      'Demand rate (Poisson) at Plant'
rho        'Utilisation rate of the plant'
 tau        'Average response time requirement'
 alpha(w)   'exact lead time from plant to pool $w$'
 C(v,w)'is the space capacity of SVC $v$ in pool $w$, this is uniform
 for *all SVCs in pool $w$'
 Cw(w)  'is the total space available for storage at pool $w$'
 C0   'is the total space available for storage space at the plant';
*-------------------------------------------------------------------
* The Data Containing longitude and latitude of each City is to be
*read in as a table
*-------------------------------------------------------------------
 Table Data(v,*)  'Datasets for each City(Node)'
$ Include RawDataset371.inc
 Display Data;  !! This Line displays the input Data
*==================================================================
* Data containing distances between cities in each of the six
*pools, are imported from Excel files that were saved in csv format
*-----------------------------------------------------------------


*------------------------------------------------------------------
* These are Added too enable simple coding Parameter values
*------------------------------------------------------------------
* Theline below represents \hat w
parameter mode(w) /w1 6, w2 6, w3 5, w4 7, w5 6, w6 7 /;
*------------------------------------------------------------------
* Assigning The Parameter values
*-------------------------------------------------------------------
 h0= 50 ;
 p(v,w) = 70;
 q(v,w) = 25;
```

```
 C0 = 5 ;
 C(v,w) = C0;
*The total storage in each pool is clxulated and assigned separately
*for each pool
 Cw(w)  = sum(v, C(v,w)) ;

lambu(v) = Data(v,"Demand") ;
 lamb0  = sum(v,lambu(v));
* Total demand in each pool is calculated and assigned separately
*for each pool
 lambw('w1') = sum(v$(1<= ord(v) and ord(v)<=6),lambu(v));
 lambw('w2') = sum(v$(7<= ord(v) and ord(v)<=12),lambu(v));
 lambw('w3') = sum(v$(13<= ord(v) and ord(v)<=17),lambu(v));
 lambw('w4') = sum(v$(18<= ord(v) and ord(v)<=24),lambu(v));
 lambw('w5') = sum(v$(26<= ord(v) and ord(v)<=30),lambu(v));
 lambw('w6') = sum(v$(31<= ord(v) and ord(v)<=37),lambu(v));
display lamb0,  lambw, C, cw;
 rho = 0.9 ; !! The value of rho will be varied later on
 tau = 0.5 ; !! The value of tau will be varied later on
parameter  alpha(w)  /w1 0.125745458, w2 0.140129167,
w3 0.06223475, w4 0.216291083, w5 0.195848958,
w6 0.241284333/;
* The Distances are saved temporarily in a variable called a
for pool
*-----------------------------------------------------------------
* VARIABLE DECLARATIONS
*-----------------------------------------------------------------
*Binary Variables
*X(v,w)    'x(v,w)= 1 If SCv is open in pool $w$, 0 otherwise'
*Y(u,v,w)  'y(u,v,w) =1, if customer u's demand is allocated to
SVC $v$ in pool $w$, 0 otherwise '
*---------------------------------------------------------------
*DECLARING THE CONDITION FOR Y(u,v,w) and a(u,v,w)
*---------------------------------------------------------------


parameters
 S0,Ib0, Lb(w);
```

```
 S0  = 3;
 Ib0 = S0 - (rho/(1-rho))*(1-rho**S0);
 Lb(w) = (rho**(S0+1)/(1-rho))/lamb0 + alpha(w);
* Determining the value of F(w)
set s2 /0*175/;  !! A set to be used for the summation
* number of svc in pools used to estimate
Sw(w) /w1 6, w2 6, w3 5, w4 7, w5 6, w6 7 /
parameter
 Sw(w) / w1 30, w2 30, w3 25, w4 35, w5 30, w6 35 /
 Fw(w), Svw(v,w),  Fvw(v,w);
Svw(v,w) = Sw(w)/mode(w);
*Svw(v,w) = Sw(w)/mode(w);
*here Svw(v,w) =  5. In order to find optimal Svw(v,w) the problem
*is solved for different scenarios of Svw(v,w) from 0 to 5 0r from
*0 to 10 and the minimum is selected.
*----------------------------------------------------------------
*THESE ARE TO BE USED IN THE OBJECTIVE EQUATION TO BE OPTIMIZED
*----------------------------------------------------------------
*Y.lo(u,v,w) = 1;
Variables  OBJ;
option sysout = on;
option domlim = 2;
*$ontext
*-----------------------------------------------------------------
* EQUATION DECLARATIONS
*------------------------------------------------------------------
Equations
Eq1, Eq4(v,w),Eq5(v,w);

* THE ObJECTIVE FUNCTION  TO BE MINIMIZED
Eq1 .. OBJ =e= sum((v,w),  (h0+q(v,w))*(sum(s2$(s2.val
<=Svw(v,w)-1),
(Svw(v,w)-s2.val)*exp[-lambu(v)*Lb(w)]*
(([lambu(v)*Lb(w)]**s2.val))/fact(s2.val))) -q(v,w)*Svw(v,w) +
lambu(v)*p(v,w)*((rho**(S0+1))/(lamb0*(1 -rho)) + alpha(w))
+ (p(v,w)-q(v,w))*(lambu(v)/lambw(w))
*(sum(s2$(s2.val<=Sw(w)-1), (Sw(w)-s2.val)*exp[-lambu(v)*Lb(w)]
*(([lambu(v)*Lb(w)]**s2.val))/fact(s2.val))-Sw(w)) ) + h0*Ib0;
```

```
* SUBJECT TO
Eq4(v,w) .. Svw(v,w) =l= C(v,w);
Eq5(v,w) .. ((rho**(S0+1))/(lamb0*(1 -rho)) + alpha(w)- tau) =l=
(lambu(v)/lambw(w))*(Sw(w)-sum(s2$(s2.val<=Sw(w)-1), (Sw(w)-s2.val)
*exp[- lambu(v)*Lb(w)]*(([lambu(v)*Lb(w)]**s2.val))/fact(s2.val)));
*Eq6(v,w) .. Svw(v,w) =l= C(v,w);
* The Model Statement  assigns a name and required equations to the
*Problem to be solved

 option iterlim =50000;
option reslim =10000;
Model Optim /all/;

* The Solve Statement Then Solves the Problem

solve Optim minimizing OBJ using mip ;
*$offtext
```

Appendix (III) model II GAMS code

```
$ontext
* EVALUATING GAMS
$offtext
* Turn on the default end of line comment Character !!
$onEolcom
* Define an end of line Comment Character of your choice to be used
* $eolcom %%
* Declaration Of Sets to Be used
Sets
  U     'Set of Customers' /1*37/
* V     'Set of Candidate SVC Locations' /1*37/
  W     'Set of Pools'   /w1*w6/
* DYNAMIC SETS (ACTING AS SUBSETS OF CUSTOMERS/SvCs FOR EACH POOL)
  u1(U) 'Set of cities in pool 1' /1*6/
  u2(U) 'Set of cities in pool 2' /7*12/
  u3(U) 'Set of cities in pool 3' /13*17/
  u4(U) 'Set of cities in pool 4' /18*24/
  u5(U) 'Set of cities in pool 5' /25*30/
```

```
  u6(U) 'Set of cities in pool 6' /31*37/
* THE SAME THING IS DONE FOR POOLS
  w1(w) 'Set of cities in pool 1' /w1/
  w2(w) 'Set of cities in pool 2' /w2/
  w3(w) 'Set of cities in pool 3' /w3/
  w4(w) 'Set of cities in pool 4' /w4/
  w5(w) 'Set of cities in pool 5' /w5/
  w6(w) 'Set of cities in pool 6' /w6/
* Make the cities(nodes) associated with u1 to u6 the same with
v1 to v6
Alias (u, v), (u1, v1),(u2, v2), (u3, v3), (u4, v4),
 (u5, v5), (u6, v6);
sets
* Data sets containing the lattitude, longitude and demand for
  Data1 'Raw Data Sets To Be entered' /Latitude, Longitude, Demand/;
*------------------------------------------------------------------
* Declaration of Parameters to be used
* Note That Gams is not case sensitive,thus f(V,W) is same as f(v,w)
*-----------------------------------------------------------------
Parameters
 f(v,w)  'Fixed cost of opening SVC at location v in Pool $w$'
 h0       'is the per unit holding cost of each unit of inventory
 *at SVCv in P oolw per unit time'
 p(v,w)     'is the per unit cost of backorder per unit inventory
 * for each unit of time'
 q(v,w)     'Lateral transshipment cost per unit inventory'
 d(u,v,w)   'Distance from SVCvw to customer u'
 d1(u,v,w)   'Transportation cost from SVCvw to customer u'
 lambu(u)   'Demand rate of customer u'
 lambvw(v,w) 'Demand rate (Poisson) at SVC $v$ in Pool $w$'
 lambw(w)   'Demand rate (Poisson) at Pool $w$'
 lamb0      'Demand rate (Poisson) at Plant'
 rho        'Utilisation rate of the plant'
 tau        'Average response time requirement'
 alpha(w)   'exact lead time from plant to pool $w$'
 dmax   'is the Maximum allowable distance between a customer its
 assigned SVC'
 a(u,v,w) 'if the distance from customer u to candidate location v
```

```
 in poolw is not greater than dmax, 0 otherwise'
 C(v,w)     'is the space capacity of SVC $v$ in pool $w$, this is
 uniform for all
  SVCs in pool $w$'
 Cw(w)      'is the total space available for storage at pool $w$'
 C0, k(u,w)        'is the total space available for storage space
 *at the plant';
*-------------------------------------------------------------------
* The Data Containing longitude and latitude of each City is to be
*read in as a table
*------------------------------------------------------------------
 Table Data(u,*)  'Datasets for each  City (Node)'
$ Include RawDataset37.inc
 Display Data;  !! This Line displays the input Data
*===================================================================
* Data containing distances between cities in each of the six pools,
*are imported from Excel files that were saved in csv format
*--------------------------------------------------------------------
$ondelim    !! Turn delimeters on
Table Distp1(*,*)
$ Include 37nodesPool1Dist.csv
display Distp1 ;

Table distp2(*,*)
$ Include 37nodesPool2Dist.csv
display Distp2

Table Distp3(*,*)
$ Include 37nodesPool3Dist.csv
display Distp3

Table Distp4(*,*)
$ Include 37nodesPool4Dist.csv
display Distp4

Table Distp5(*,*)
$ Include 37nodesPool5Dist.csv
display Distp5
```

```
Table Distp6(*,*)
$ Include 37nodesPool6Dist.csv
display Distp6 ;

$offdelim
*---------------------------------------------------------------------
* These are Added too enable simple coding Parameter values
*---------------------------------------------------------------------
* Theline below represents \hat w
parameter mode(w) /w1 6, w2 6, w3 5, w4 7, w5 6, w6 7 /;
*----------------------------------------------------------------------
* Assigning The Parameter values
*----------------------------------------------------------------------
 f(v,w)= 10000;
 h0= 50 ;
 p(v,w) = 70;
 q(v,w) = 25;
 dmax = 150;
 C0 = 5 ;
 C(v,w) = C0;
*The total storage in each pool is clxulated and assigned separately
*for each pool
 Cw(w)  = sum(v, C(v,w)) ;

lambu(u) = Data(u,"Demand") ;
 lamb0  = sum(u,lambu(u));
* Total demand in each pool is calculated and assigned separately
*for each pool
 lambw('w1') = sum(u$(1<= ord(u) and ord(u)<=6),lambu(u));
 lambw('w2') = sum(u$(7<= ord(u) and ord(u)<=60),lambu(u));
 lambw('w3') = sum(u$(13<= ord(u) and ord(u)<=17),lambu(u));
 lambw('w4') = sum(u$(18<= ord(u) and ord(u)<=24),lambu(u));
 lambw('w5') = sum(u$(26<= ord(u) and ord(u)<=30),lambu(u));
 lambw('w6') = sum(u$(31<= ord(u) and ord(u)<=37),lambu(u));
display lamb0,  lambw, C, cw;
 rho = 0.9 ; !! The value of rho will be varied later on
 tau = 0.205 ; !! The value of tau will be varied later on
```

```
parameter  alpha(w)  /w1 0.2282144583, w2 0.2269349167,
w3 0.1834542083, w4 0.1869252500, w5 0.3439902500,
w6 0.2304287917 /;
* The Distances are saved temporarily in a variable called a
*for each pool
d(u1,v1,'w1') = Distp1(u1,v1) ;
d(u2,v2,'w2') = Distp2(u2,v2) ;
d(u3,v3,'w3') = Distp3(u3,v3) ;
d(u4,v4,'w4') = Distp4(u4,v4) ;
d(u5,v5,'w5') = Distp5(u5,v5) ;
d(u6,v6,'w6') = Distp6(u6,v6) ;
d1(u1,v1,'w1') = Distp1(u1,v1)/10 ;
d1(u2,v2,'w2') = Distp2(u2,v2)/10 ;
d1(u3,v3,'w3') = Distp3(u3,v3)/10;
d1(u4,v4,'w4') = Distp4(u4,v4)/10 ;
d1(u5,v5,'w5') = Distp5(u5,v5)/10 ;
d1(u6,v6,'w6') = Distp6(u6,v6) /10;
*------------------------------------------------------------
* VARIABLE DECLARATIONS
*------------------------------------------------------------
Binary Variables
 X(v,w)    'x(v,w)= 1 If SCv is open in pool $w$, 0 otherwise'
 Y(u,v,w)  'y(u,v,w) =1, if customer u's demand is allocated to
 SVC $v$ in pool $w$, 0 otherwise '
*--------------------------------------------------------------
*DECLARING THE CONDITION FOR Y(u,v,w) and a(u,v,w)
*--------------------------------------------------------------
loop((u,v,w),
if( [(d(u,v,w) > 0 and d(u,v,w)<= dmax)] ,
    a(u,v,w)= 1;
*    Y.l(u,v,w)= 1;
    k(u,w) = 1 ;
else
    a(u,v,w)= 0;
*    Y.l(u,v,w)= 0;
```

```
    k(u,w) = 0 ;
 );

);

loop((u1,v1,w1),
if ([(d(u1,v1,w1)=0) and (ord(u1)= ord(v1))],
 a(u1,v1,w1)= 1;
);
);

loop((u2,v2,w2),
if ([(d(u2,v2,w2)=0) and (ord(u2)= ord(v2))],
 a(u2,v2,w2)= 1;
);
);

loop((u3,v3,w3),
if ([(d(u3,v3,w3)=0) and (ord(u3)= ord(v3))],
 a(u3,v3,w3)= 1;
);
);

loop((u4,v4,w4),
if ([(d(u4,v4,w4)=0) and (ord(u4)= ord(v4))],
 a(u4,v4,w4)= 1;
);
);

loop((u5,v5,w5),
if ([(d(u5,v5,w5)=0) and (ord(u5)= ord(v5))],
 a(u5,v5,w5)= 1;
);
);

loop((u6,v6,w6),
if ([(d(u6,v6,w6)=0) and (ord(u6)= ord(v6))],
 a(u6,v6,w6)= 1;
```

```
);
);

*Display the parameter 'a' that contains all distances between
*cities
display d, a , lambu;

parameters
 S0, Bb0, Wt0,  Ib0, Lb(w);
 S0  = 4;
 Bb0 = rho**(S0+1)/(1-rho) ;
 Wt0 = Bb0/lamb0 ;
 Ib0 = S0 - (rho/(1-rho))*(1-rho**S0);
 Lb(w) = Wt0 + alpha(w);
* Determining the value of F(w)
set s2 /0*175/;  !! A set to be used for the summation
* number of svc in pools used to estimate Sw(w) /w1 6, w2 6,
 w3 5, w4 7, w5 6, w6 7 /
parameter
 Sw(w) / w1 60, w2 60, w3 50, w4 70, w5 60, w6 70 /
 Fw(w), Svw(v,w),  Fvw(v,w);
Svw(v,w) = Sw(w)/mode(w);
Svw(v,w) = Sw(w)/mode(w);
*here Svw(v,w) =  5. In order to find optimal Svw(v,w) the
problem is solved for different scenarios of Svw(v,w) from
0 to 5 Or from 0 to 10 and the minimum is selected.
* A Library of Function is  Loaded which contains the cdfpoisson
function
$funcLibIn stolib stodclib
Function cdfPoisson /stolib.cdfPoisson/;
* Fw(w)    =  cdfPoisson(s2.val$(s2.val<=Sw(w)), lambw(w)*Lb(w));
* Fvw(v,w) = cdfPoisson( s2.val, sum(u,lambu(u)*Y(u,v,w)*Lb(w)) );
Parameters Ibw(w), Bbw(w);
Ibw(w) = sum(s2${s2.val<=Sw(w)-1}, cdfPoisson{s2.val, lambw(w)*Lb(w)});
Bbw(w) = ((lambw(w)/lamb0)*Bb0 + lambw(w)*alpha(w)- Sw(w)+ Ibw(w));
*-----------------------------------------------------------------
*THESE ARE TO BE USED IN THE OBJECTIVE EQUATION TO BE OPTIMISED
*-----------------------------------------------------------------
```

```
*Y.lo(u,v,w) = 1;
Variables  OBJ;
option sysout = on;
option domlim = 2;
*$ontext
*-----------------------------------------------------------------
* EQUATION DECLARATIONS
*------------------------------------------------------------------
Equations
Eq1,   Eq3(u,v,w),  Eq5(v,w)
Eq21(u1,w1) , Eq22(u2,w2),Eq23(u3,w3),Eq24(u4,w4),Eq25(u5,w5),
Eq26(u6,w6) ;

* THE ObJECTIVE FUNCTION  TO BE MINIMIZED
Eq1 .. OBJ =e= sum((v,w), f(v,w)*X(v,w) + h0*(sum(s2$
(s2.val<=Svw(v,w)-1),(Svw(v,w)-s2.val)*exp[-sum{u,lambu(u)*Y(u,v,w)}
*Lb(w)]*(([sum{u,lambu(u)*Y(u,v,w)}*Lb(w)]**s2.val))/fact(s2.val))) +
p(v,w)*(( sum(u,lambu(u)*Y(u,v,w))/lambw(w) )*Bbw(w))+  q(v,w)*
(sum(s2${s2.val<=Svw(v,w)-1},(Svw(v,w)-s2.val)*exp[-sum{u,lambu(u)*
Y(u,v,w)}*Lb(w)]*(([sum{u,lambu(u)*Y(u,v,w)}*Lb(w)]**s2.val)
/fact(s2.val)) )- Svw(v,w) - (sum(u,lambu(u)*Y(u,v,w))/lambw(w))*
(Ibw(w)-Sw(w)))+ sum(u, lambu(u)*Y(u,v,w)*d1(u,v,w)) ) + h0*Ib0;
* SUBJECT TO
Eq21(u1,w1) .. sum[v1, Y(u1,v1,w1)] =e= 1;
Eq22(u2,w2) .. sum[v2, Y(u2,v2,w2)] =e= 1;
Eq23(u3,w3) .. sum[v3, Y(u3,v3,w3)] =e= 1;
Eq24(u4,w4) .. sum[v4, Y(u4,v4,w4)] =e= 1;
Eq25(u5,w5) .. sum[v5, Y(u5,v5,w5)] =e= 1;
Eq26(u6,w6) .. sum[v6, Y(u6,v6,w6)] =e= 1;

Eq3(u,v,w) .. Y(u,v,w) =l= a(u,v,w)*X(v,w) ;
*Eq4(v,w) .. Svw(v,w) =l= C(v,w);
Eq5(v,w) .. (sum(u, lambu(u)*Y(u,v,w))/lambw(w) )*Bbw(w) =l=
tau*sum(u, lambu(u)*Y(u,v,w));
* The Model Statement  assigns a name and required equations
*to the Problem to be solved

 option iterlim =50000;
```

```
option reslim =10000;
Model Optim /all/;

* The Solve Statement Then Solves the Problem

solve Optim minimizing OBJ using minlp ;
*$offtext
parameter cnt ;
cnt = sum((v,w)$(x.l(v,w)),1)
display X.l, Y.l, cnt;
```

Appendix (IV)model II GAMS code without LT

```
$ontext
* EVALUATING GAMS
$offtext
* Turn on the default end of line comment Character !!
$onEolcom
* Define an end of line Comment Character of your choice to be used
* $eolcom %%
* Declaration Of Sets to Be used
Sets
  U     'Set of Customers' /1*37/
* V     'Set of Candidate SVC Locations' /1*37/
  W     'Set of Pools'   /w1*w6/
* DYNAMIC SETS (ACTING AS SUBSETS OF CUSTOMERS/SvCs FOR EACH POOL)
  u1(U) 'Set of cities in pool 1' /1*6/
  u2(U) 'Set of cities in pool 2' /7*12/
  u3(U) 'Set of cities in pool 3' /13*17/
  u4(U) 'Set of cities in pool 4' /18*24/
  u5(U) 'Set of cities in pool 5' /25*30/
  u6(U) 'Set of cities in pool 6' /31*37/
* THE SAME THING IS DONE FOR POOLS
  w1(w) 'Set of cities in pool 1' /w1/
  w2(w) 'Set of cities in pool 2' /w2/
  w3(w) 'Set of cities in pool 3' /w3/
  w4(w) 'Set of cities in pool 4' /w4/
  w5(w) 'Set of cities in pool 5' /w5/
```

```
  w6(w) 'Set of cities in pool 6' /w6/
*Make the cities(nodes) associated with u1 to u6 the same
*with v1 to v6
Alias (u, v), (u1, v1),(u2, v2), (u3, v3), (u4, v4),
(u5, v5), (u6, v6);
sets
* Data sets containing the lattitude, longitude and demand
Data1 'Raw Data Sets To Be entered' /Latitude, Longitude, Demand/;
* Declare the set indices for U, V and W as same as i, j and pool
*------------------------------------------------------------------
* Declaration of Parameters to be used
 * Note That Gams is not case sensitive,thus f(V,W) same as f(v,w)
*-------------------------------------------------------------------
Parameters
 f(v,w)  'Fixed cost of opening SVC at location v in Pool $w$'
 h0    'is the per unit holding cost of each unit of inventory
 at SVCv in P oolw per unit time'
 p(v,w) 'is the per unit cost of backorder per unit inventory for
 each unit of time'
 q(v,w)     'Lateral transshipment cost per unit inventory'
 d(u,v,w)   'Distance from SVCvw to customer u'
 d1(u,v,w)   'Transportation cost from SVCvw to customer u'
 lambu(u)   'Demand rate of customer u'
 lambvw(v,w) 'Demand rate (Poisson) at SVC $v$ in Pool $w$'
 lambw(w)   'Demand rate (Poisson) at Pool $w$'
 lamb0      'Demand rate (Poisson) at Plant'
 rho        'Utilisation rate of the plant'
 tau        'Average response time requirement'
 alpha(w)   'exact lead time from plant to pool $w$'
 dmax       'is the Maximum allowable distance between a customer
 its assigned SVC'
 a(u,v,w)   'if the distance from customer u to candidate location
 v in pool $w$ is not
greater than dmax, 0 otherwise'
 C(v,w)     'is the space capacity of SVC $v$ in pool $w$, this is
 uniform for all SVCs in pool $w$'
 Cw(w)      'is the total space available for storage at pool $w$'
 C0, k(u,w)        'is the total space available for storage space
```

```
 at the plant'  ;
*------------------------------------------------------------------
* The Data Containing longitude and latitude of each City is to
*be read in as a table
*-----------------------------------------------------------------
 Table Data(u,*)  'Datasets for each of The Cities(Nodes)'
$ Include RawDataset371.inc
 Display Data;  !! This Line displays the input Data
*===================================================================
* Data containing distances between cities in each of the six pools,
*are imported from Excel files that were saved in csv format
*--------------------------------------------------------------------
$ondelim    !! Turn delimeters on
Table Distp1(*,*)
$ Include 37nodesPool1Dist.csv
display Distp1 ;

Table distp2(*,*)
$ Include 37nodesPool2Dist.csv
display Distp2

Table Distp3(*,*)
$ Include 37nodesPool3Dist.csv
display Distp3

Table Distp4(*,*)
$ Include 37nodesPool4Dist.csv
display Distp4

Table Distp5(*,*)
$ Include 37nodesPool5Dist.csv
display Distp5

Table Distp6(*,*)
$ Include 37nodesPool6Dist.csv
display Distp6 ;

$offdelim
```

```
*-----------------------------------------------------------------
* These are Added too enable simple coding Parameter values
*-----------------------------------------------------------------
* Theline below represents \hat w
parameter mode(w) /w1 6, w2 6, w3 5, w4 7, w5 6, w6 7 /;
*------------------------------------------------------------------
* Assigning The Parameter values
*------------------------------------------------------------------
 f(v,w)= 10000;
 h0= 50 ;
 p(v,w) = 70;
 q(v,w) = 25;
 dmax = 150;
 C0 = 5.5 ;
 C(v,w) = C0;
*The total storage in each pool is clxulated and assigned separately
*for each pool
 Cw(w)  = sum(v, C(v,w)) ;

lambu(u) = Data(u,"Demand") ;
 lamb0  = sum(u,lambu(u));
* Total demand in each pool is calculated and assigned separately
*for each pool
 lambw('w1') = sum(u$(1<= ord(u) and ord(u)<=6),lambu(u));
 lambw('w2') = sum(u$(7<= ord(u) and ord(u)<=60),lambu(u));
 lambw('w3') = sum(u$(13<= ord(u) and ord(u)<=17),lambu(u));
 lambw('w4') = sum(u$(18<= ord(u) and ord(u)<=24),lambu(u));
 lambw('w5') = sum(u$(26<= ord(u) and ord(u)<=30),lambu(u));
 lambw('w6') = sum(u$(31<= ord(u) and ord(u)<=37),lambu(u));
display lamb0,  lambw, C, cw;
 rho = 0.9 ; !! The value of rho will be varied later on
 tau = 0.11 ; !! The value of tau will be varied later on

*CALCULATION FOR THE DISTANCES BETWEEN EACH CITY
parameters
lat(v), long(v);
    lat(v)  = Data(v,'latitude');
    long(v) = Data(v,'longitude');
```

```
* The line below is Optional.
display
    lat ;
* Convert Them to radians
    lat(v)  = lat(v)*Pi/180;
    long(v) = long(v)*Pi/180;
* The line below is Optional.
display
    lat ;
Scalar
    Radi 'The   Radius  of the Earth ';
    Radi = 6371;
Parameter
dist1(u,v);
* The Distances are calculated using a the haversine formula
dist1(u,v) = 2*Radi*arcsin{ sqrt[ power( sin( (lat(u)-lat(v))/2 ),
2 ) + cos(lat(u))
*cos(lat(v)) * power( sin( (long(u)-long(v))/2 ), 2 )  ]} ;
display
    dist1 ;
* The Shipping cost is obtained via dividing the distances by 10
* The plant is assumed to be in Abuja (i.e node 37) so we have to
*find the shipping cost
* of transporting goods from the plant (Abuja) to each candidate
*facility (service centres)
* location, And  then save these values in a parameter called alpha(W).
* YOU CAN CHANGE THE 37  TO THE NUMBER  of the city where you want the
*plant to be located, e.g
* If you want the plant to be in Abia state, then you have to use
*alpha(W) = d('1',w);  to
* replace the line of code below.

*The Distances are saved temporarily in a variable called a
d(u1,v1,'w1') = Distp1(u1,v1) ;
d(u2,v2,'w2') = Distp2(u2,v2) ;
d(u3,v3,'w3') = Distp3(u3,v3) ;
d(u4,v4,'w4') = Distp4(u4,v4) ;
d(u5,v5,'w5') = Distp5(u5,v5) ;
```

```
d(u6,v6,'w6') = Distp6(u6,v6) ;
d1(u1,v1,'w1') = Distp1(u1,v1)/10 ;
d1(u2,v2,'w2') = Distp2(u2,v2)/10 ;
d1(u3,v3,'w3') = Distp3(u3,v3)/10;
d1(u4,v4,'w4') = Distp4(u4,v4)/10 ;
d1(u5,v5,'w5') = Distp5(u5,v5)/10 ;
d1(u6,v6,'w6') = Distp6(u6,v6) /10;
*-------------------------------------------------------------------
* VARIABLE DECLARATIONS
*-------------------------------------------------------------------
Binary Variables
 X(v,w)  'x(v,w)= 1 If SVC $v$ is open in pool $w$, 0 otherwise'
 Y(u,v,w) y(u,v,w) =1, if customer u's demand is allocated to
 *SVC $v$ in pool $w$, 0 otherwise '
*-------------------------------------------------------------------
*DECLARING THE CONDITION FOR Y(u,v,w) and a(u,v,w)
*-------------------------------------------------------------------
loop((u,v,w),
if( [(d(u,v,w) > 0 and d(u,v,w)<= dmax)] ,
    a(u,v,w)= 1;
*    Y.l(u,v,w)= 1;
    k(u,w) = 1 ;
else
    a(u,v,w)= 0;
*    Y.l(u,v,w)= 0;
    k(u,w) = 0 ;
 );

);

loop((u1,v1,w1),
if ([(d(u1,v1,w1)=0) and (ord(u1)= ord(v1))],
 a(u1,v1,w1)= 1;
);
);

loop((u2,v2,w2),
if ([(d(u2,v2,w2)=0) and (ord(u2)= ord(v2))],
```

```
 a(u2,v2,w2)= 1;
);
);

loop((u3,v3,w3),
if ([(d(u3,v3,w3)=0) and (ord(u3)= ord(v3))],
 a(u3,v3,w3)= 1;
);
);

loop((u4,v4,w4),
if ([(d(u4,v4,w4)=0) and (ord(u4)= ord(v4))],
 a(u4,v4,w4)= 1;
);
);

loop((u5,v5,w5),
if ([(d(u5,v5,w5)=0) and (ord(u5)= ord(v5))],
 a(u5,v5,w5)= 1;
);
);

loop((u6,v6,w6),
if ([(d(u6,v6,w6)=0) and (ord(u6)= ord(v6))],
 a(u6,v6,w6)= 1;
);
);

*Display the parameter 'a' that contains all distances between cities
display d, a , lambu;

parameters
 S0, Bb0, Lb(v,w);
 S0  = 1;
 Bb0 = rho**(S0+1)/(1-rho) ;
*Wt0 = Bb0/lamb0 ;
*alpha(v,w)= dist1('18',v)/2400
*Ib0 = S0 - (rho/(1-rho))*(1-rho**S0);
```

```
 Lb(v,w) = Bb0/lamb0 + dist1('18',v)/2400;
* Determining the value of F(w)
set s2 /0*175/;  !! A set to be used for the summation
* number of svc in pools used to estimate Sw(w) /w1 6, w2 6,
*w3 5, w4 7, w5 6, w6 7 /
parameter
 Sw(w) / w1 30, w2 30, w3 25, w4 35, w5 30, w6 35 /
 Fw(w), Svw(v,w),  Fvw(v,w);
Svw(v,w) = Sw(w)/mode(w);
Svw(v,w) = Sw(w)/mode(w);
*here Svw(v,w) =  5. In order to find optimal Svw(v,w) the
*problem is solved for different scenarios of Svw(v,w) from
*0 to 5 0r from 0 to 10 and the minimum is selected.
*Y.lo(u,v,w) = 1;
Variables  OBJ;
option sysout = on;
option domlim = 2;
*$ontext
*-------------------------------------------------------------------
* EQUATION DECLARATIONS
*------------------------------------------------------------------
Equations
Eq1,   Eq3(u,v,w),  Eq5(v,w), Eq21(u1,w1) , Eq22(u2,w2),Eq23(u3,w3),
Eq24(u4,w4),Eq25(u5,w5), Eq26(u6,w6) ;

* THE ObJECTIVE FUNCTION  TO BE MINIMIZED
Eq1 .. OBJ =e= sum((v,w), f(v,w)*X(v,w)  -p(v,w)*Svw(v,w)+(h0 +p(v,w))
*(sum(s2$(s2.val<=Svw(v,w)-1),(Svw(v,w)-s2.val)*exp[-sum(u,lambu(u)*
Y(u,v,w))*Lb(v,w)]*(([sum{u,lambu(u)*Y(u,v,w)}*Lb(v,w)]**s2.val))
/fact(s2.val)))+sum(u, [p(v,w)*(Lb(v,w)+  d1(u,v,w))]*lambu(u)
*Y(u,v,w)) ) + h0*[S0 - (rho/(1-rho))*(1-rho**S0)];
* SUBJECT TO
Eq21(u1,w1) .. sum[v1, Y(u1,v1,w1)] =e= 1;
Eq22(u2,w2) .. sum[v2, Y(u2,v2,w2)] =e= 1;
Eq23(u3,w3) .. sum[v3, Y(u3,v3,w3)] =e= 1;
Eq24(u4,w4) .. sum[v4, Y(u4,v4,w4)] =e= 1;
Eq25(u5,w5) .. sum[v5, Y(u5,v5,w5)] =e= 1;
Eq26(u6,w6) .. sum[v6, Y(u6,v6,w6)] =e= 1;
```

```
Eq3(u,v,w) .. Y(u,v,w) =l= a(u,v,w)*X(v,w) ;
Eq5(v,w) .. (Lb(v,w)- tau)*sum(u,lambu(u)*Y(u,v,w)) =l=
sum(s2$(s2.val<=Svw(v,w)-1),[1- (Svw(v,w)-s2.val)*exp[-sum{u,
lambu(u)*Y(u,v,w)}*Lb(v,w)]*(([sum{u,lambu(u)*Y(u,v,w)}*Lb(v,w)]
**s2.val))/fact(s2.val)]);

Model Optim /all/;
solve Optim minimizing OBJ using minlp ;
*$offtext
parameter cnt ;
cnt = sum((v,w)$(x.l(v,w)),1)
display X.l, Y.l, cnt;
```