1.

CHAPTER 1

INTRODUCTION

1.1. SUPPLY CHAIN CONCEPTS

1.1.1. Production Eras and Challenges

The challenges of production of goods and services after the Second World War have gone through three main chronological stages as outlined by Hopp and Wallace (2008). The first era focuses mainly on productivity, and this leads to the traditional focus on cost. Some visible developments in this era included fast paced development in scientific management, especially the reductionist techniques of work study, and more pervasive development and deployment of financial ratios for monitoring the health of firms. This was dated back to manufacturing itself, but received boost immediately after Second World War till the seventies. This trend was supported by the relatively sole strong position of the American economy at the time.

The productivity era was succeeded by the era of quality movement, which was dated back to the seventies and eighties, although the pioneering work appears to have been done as far back as 1931 by Shewhart. Some of the important tools of this holistic management era included the Total Quality Management (TQM) and Just in Time (JIT). These were later revived again in the Six Sigma and Lean movement. This movement was bolstered by the advent of competing nations like Japan and Germany among others that have started emerging from the rubbles of the war and are entering the same market that has been hitherto dominated by America.

The latest era appears to be that of integration, and this is assumed to have commenced in the nineties. This development was driven especially by the rapid development in the Information and Communication Technology (ICT) that makes the whole world to become more integrated than it has ever been. This globalisation trend has been further enhanced by the changing econo-political structure in most Asian, Latin American, Eastern Europe and African countries (from centrally planned to market driven philosophies), and the advancement of the World Wide Web that makes countries to locate their various offices where ever they feel is most appropriate for their businesses.

The Asian Tigers' miracle at the Han River and the emergence of China as strong manufacturing centres, with the later entrance of the Indo-Brazil and South African centres has created massive international competition.

With the possibility of savings by focusing only on the traditional methods of work reductionism thinning out, more focus shifts to the total manufacturing system (the network: from the supplier's supplier to the customer's customer), especially since most international legal, econo-political, fiscal and technical barriers are being constantly lowered. This era birthed the current production trend of Supply Chain Management (SCM). Some other related ideas in this era include Business Process Modelling (BPM) and Enterprise Resource Planning (ERP) amongst others. Supply Chain has been defined severally by a number of authors, but one definition that seems succinct but exhaustive in this thesis' context is presented next.

1.1.2. Supply Chain Definition and Concepts

A supply chain has been defined as a goal oriented network of processes and stock points used to deliver goods and services to customers (Hopp, 2008). This definition highlights the key features of any supply chain to be: the goal, the network, the stock points, the process stations, the products (goods and/or services) and the customers. This definition actually summarises all that is done in a supply chain (especially from the market perspective). This is further explored.

The basic goal of most organisations is profit. Two paths usually lead to increase in profit: cost reduction or growth in market size. But progress along one of these paths may actually degrade the other. So, organisations need to decide how much efforts are put into these two paths to realise the organisational goal of profit, both in the short and the long run. This makes the goal to be closely related to the strategy of the organisation, which is done at the highest planning level, and decides how much of what is traded off to achieve the other, and thereby , hopefully, placing the organisational plan on some sort of "efficient frontier".

Customers are important in the chain because they are the market. The second path to profitability implies ensuring that they are satisfied and delighted. But only if their needs (i.e. demand and timing) are known well in advance would managing the whole chain become easy and all unnecessary costs could be easily eliminated (or reasonably reduced). But, unfortunately, these customers are not so predictable, and hence comes in variability into the system. This is the first level of variability in the supply chain; which is related to the management of the external influences on the chain. This comes in the form of uncertain external demands and lead times.

Supply Chain Process Points, or the work stations, are the resources that actually get out goods and services ready for the customers one wishes to delight. These are the transformation centres that, in the word of Langley et al. (2009) add utilities of form to the input material by transforming its form (or may be servicing the customer). These process centres also contribute the second level of uncertainty, which in this case is internal to the system. This is in the form of uncertain process times of the process centres as products are transformed at these centres. This unavoidable variability in the system forces the strategic deployment of reserves in the supply chain. These reserves have been referred to by Webster (2008) as system slacks. These slacks are in the form of extra capacities or inventories. Therefore, the process points also serve as strategic capacity reserve points while stock points serve as strategic material reserve points. This leads to the discussion of stock points.

Stock points are positions in the supply chain network where inventories of materials or goods are found. These points exist due to two reasons: firstly, they may exist as a result of deliberate plan to keep some materials in some identified locations in the network, e.g. finished goods, some important raw materials, etc. The second reason is because some inventory build up in the system and are controlled by some natural laws like the Little's law. These form part of the work-in-process inventory and cannot be controlled directly but by regulation of flow through the system. Flows are now discussed next.

Flows are the actual products (or even customers) that are processed at the processing centres. They are generated basically by actual orders or demand forecasts. Above the decoupling point, they are driven by a push (production plan/forecast) while below the decoupling point, they are driven by a pull (customer orders). Another very close term is scheduling. Management of flows are very important in any supply chain that would be successful. Flows through the chain or the stations are usually stochastic, and this affects deployment and management of slacks of capacity and inventory. Decisions about full or under-utilisation of capacity affect the inventory cost and profitability of the system. Also, decisions about level of inventory necessary to support a given level of flow are crucial because this affects the level of customer service as well as operating cost of the whole network. These are all inter-related decisions that must be made in the production context. The decisions could sometimes be simplified (howbeit to some level) by choosing a suitable management philosophy (or a mix of such) to adopt. These philosophies are briefly discussed later.

1.1.3. The Goal of a Supply chain

One key issue about which most stakeholders in a supply chain have a common agreement is the provision of superior customer service. Doing this at a low cost is another important thing, and so, the interest in the landed cost of the product and not just the production cost.

Making goods available to customers when needed (referred to as the utility of time) could be achieved through two main means: superior transport service or keeping stock near customers. Two focus areas concerned about this are transport management and inventory management. It is therefore not surprising that transport and inventory costs have been identified as the two major costs of any supply chain. (Langley et al., 2009).

1.1.4. Importance of Inventory

Inventory occupies a strategic position in a production system. Apart from being a major means of fulfilling customer orders, it also has a major effect on the books of the company in that it affects both the balance sheet and income statement; hence its effective management is crucial. The main function, though, is like insurance in the production system, absorbing the variability shocks. Based on the function it performs, it has been classified as cycle stock, safety stock, contingency stock, process stock etc (Jacobs et al., 2009). It is generally true that the level of uncertainty of demand and lead time are the two main parameters that affect the modelling of its behaviour.

1.1.5. Some Production Management Philosophies

Production management philosophies are developed to guide management through effective decision making in the processes of production management that involves intricate and dependent trade-offs. The main difference between all these philosophies is the perception and treatment of slacks in the system. Both slacks cost the system, but one is usually more acceptable than the other depending on the philosophy. Three basic philosophies to be considered are Material Requirement Planning (MRP), Lean Manufacturing and the Theory of Constraints (TOC).

Lean is very critical of inventory, and in ideal Lean environment, the batch size is equal to the actual demand. It works by pure pull and rather tolerates extra capacity than extra inventory. Inventory there is hardly zero, however, but the Kanban controls both the scheduling and the effective quantity of inventory in the system. The MRP accepts more slacks of inventory and tends to utilise capacity more than Lean. Inventory is also used to support capacity utilisation. Theory of constraint, however, is built entirely around flow. Inventory is placed in strategic locations to support the critical resources, while the capacity slacks in the non-critical resources are also used to support flow through the entire system; especially through the critical resources.

1.2. SUPPLY CHAIN SYSTEMS AND MODELLING

Systems have many definitions depending both on the discipline and the issue of interest. In the current context, the system is basically some sort of processes of interest. Systems have some state variables of interest, in this case the level of inventory present in the system. Usually, these state variables can only be manipulated indirectly through the control of some other variables called the control variables. Systems have decision variables, in this case the order policy, order quantity, or the rate of flow through the system, all of which could be manipulated to affect the positions of the state variables, which in turn determine the overall system performance. These state variables together with the parameters, which in many cases are constants or variables with known patterns are what determine the values of the system performance indicators. Such indicators in this context include system cost, level of customer service, utilisation, etc. It is usually necessary to have models that represent these systems so that the behaviour of the systems could be understood through the behaviour of these models.

The contextual and semantic definition of model is quite diverse, but a succinct definition for the current context is that a model is a representation of a system that allows for investigation of the properties of the system and in some cases prediction of the future outcomes.

Models are important in systems analysis and engineering, and the complexity could be viewed along the two dimensional axes of time changes and level of certainty. This makes all systems to be reasonably captured in a four quadrant space of deterministic-static, deterministic-dynamic, stochastic-static and stochastic-dynamic regions. This makes the system whose variables are in the deterministic-static quadrant the most tractable in respect of their mathematical computation, while the stochastic-dynamic models are the least tractable problems. The quadrant to which a problem falls also usually determines the type of models that would be most appropriate for it. Usually, most typical supply chain models fall in the stochastic region and so may need some sophisticated level of mathematical manipulation.

Most models presented in this work are Markovian, so, the problems require the instruments of probability theory, and in some instance matrix mathematics, or some level of differential calculus.

Modelling is both an art and a science. It is an art because the dexterity often improves with usage. It is a science because most techniques have logical sequences and formal methods that are followed. A good modeller knows the level of complexity at which to pitch the modelling of a system. Sometimes, it suffices to use simple models and allow for the inclusion of the simplifying assumptions in the interpretation of the results. This saves a great deal of modelling and solution efforts while still effective at achieving the intention of the model. But in certain instances, there may be the need to develop some more complex models without which some important characteristics of the systems would be sacrificed. These facts have been well noted by Sterman (2000) and Zipkin (2000) and were taken note of in the development of models in this work. It, thus, became necessary to employ the probability tools while solving for the steady state probability distribution of the input parameters of the selected problems, and the use of simple differential calculus in determining the optimal flow parameters given that the system is operating at the steady state.

Supply chain modelling has utilised many analytical tools for the management of stock level and flow of products in the entire chain or at a station in the chain. These techniques include classical optimisation tools, mathematical programming, simulation modelling and probability models. Cases where one or more input into the system (usually the demand or/and lead time) are stochastic have always called for the use of probability techniques, either as simulation models, or in the estimation of the equilibrium properties of the system.

1.3. LITERATURE REVIEW

Various analytical tools have been used in the analysis of production systems to optimise the levels of stock (inventory) it holds. The type of tool depends on the assumptions made about the nature of product flow through the system. This ranges from the deterministic-static type to the dynamic-stochastic type discussed earlier. Such tools include classical optimisation tools, mathematical programming tools, probability models and simulation. Some popular works have been produced in each category.

1.3.1. The Harris Model

The use of deterministic optimisation techniques in the management of the appropriate stock levels to keep in a production environment is pervasive. The seminal model in this category is the Economic Order Quantity (EOQ) model, developed by Harris (1913) and popularised by Wilson (1934). This model is deterministic and static. It also has many other assumptions including zero (or deterministic) lead time, shortages and backlogging not allowed, unit purchase price independent of order quantity, infinite product life, instantaneous product availability (infinite capacity), perfect order quality, fixed set up cost, single item, and probably more. This model has been modified in diverse ways by relaxing one or more of its assumptions. And it is the relaxation of some such assumptions that made the use of classical optimisation techniques inadequate for analysis in certain instances.

There have been some major groups of extensions to this classic work. The Dynamic Economic Lot (DEL) Model by Wagner and Whitten (1958, 2004) removes the static demand assumption, but still assumes the future demand pattern is known with certainty. The Silver-Meal heuristics is another seminal work in this direction. Another interesting extension is in that of single item assumption. The Joint Replenishment Problem (JRP) has been studied by many authors. Goyal and Soni (1969) and Goyal (1974) are notable. Other contributors include Van Eijs (1993), Viswanathan (2002), Fung and Ma (2001), Chan, Cheung and Langevin (2002) and Federgruen and Zheng (1992).

Multi-echelon inventory is another area that has generated much interest, starting from Clark and Scarf (1960). Others include Graves (1985), Erkip, et al (1990) Chen (2000), Rau et al (2003) and Viswanathan and Piplani (2001).

1.3.2. Deteriorating inventory

An area that has enjoyed an extensive research is the deteriorating inventory studies. Starting from the seminal work by Ghare and Shrader (1963) which is a deterministic demand model, much work has followed since. Nahmias (1982) made a detailed survey of the work done on deteriorating inventory up until that time. He summarised the contribution of the various authors reviewed and classified the work into five main areas based on:

- Fixed life perishability,
    - deterministic demand and stochastic demand, single and multi products, exact and approximate solutions, single and multi echelon
- Random lifetime models
    - Periodic review and exponential decay models
- Queuing models with impatience
- Applications.

Raafat (1991) extended the survey to the contributions made after Nahmias. While most of the models reviewed by Nahmias are fixed lifetime models, Raafat extended the survey to cover a lot more random deterioration models. Raafat classified the literatures as single or multiple items, deterministic or probabilistic demand, static or varying demand, single or multiple period, purchase or production model, availability of quantity discount(s), allowance for shortages, constant or varying deterioration rate.

Since the two compendia are quite detailed, effort would be concentrated on reviewing some of the more recent works done after Raafat. Goh et al (1993) presented a model in which inventory deteriorates in two stages. The arrival is a Poisson process with rate λ and the demand rates are μ_1 for stage 1 (fresh) product and μ_2 for the product older than stage 1 but not yet obsolete. Various system parameters were considered in this model. The model was modified in Yadavalli, et al. (2004) with the inclusion of lead time with arbitrary distribution and solved for the various system parameters. Vaughan (1994) presented a customer realised product expiration, in which he treated the expiration date of the product as a decision variable, and the product life time is treated as a random variable.

Kalpakam and Sapna (1995) dealt with a base stock policy, where the lead time is stochastic and correlated with the possibility of lost demand. Products are taken out of the system due to failure or demand. The system parameters were determined. Hariga (1996) developed an EOQ model for deteriorating inventory with time varying demand and with shortages allowed and completely backlogged. The performance of the model with linear and exponential demand inputs was analysed. Yadavalli et al (2006) also presented a model for two component production-inventory assembly system in which products are assembled from two components. A component is produced with the lead time following an arbitrary distribution and the other component is purchased with an exponential lead time. System parameters were estimated.

Chakrabathy et al (1998) presented a model in which the deterioration of inventory follows a three parameter Weibull distribution. The demand is assumed to be time varying and shortages are allowed in the system. Lee and Wu (2002) is a model with Weibull distribution deterioration and power demand with complete backlogging of shortages, and this model was extended by Dye (2004) to a general type time-proportional backlogging rate model. The backlogging rate was defined as a function of the waiting time. Chiao et al (2008) presented a model with two storage facilities, partial backlogging and quantity discount. In this model, the excess product is kept in a rented warehouse due to capacity constraint in own warehouse.

Cases of joint demand have also been investigated by Yadavalli et al. (2004) where there is capacity constraint on stored items and each has different reorder points, but the reorder for one item triggers reorder of all other items. In another paper, Yadavalli et al. (2006) considered a case where two products have individual Poisson demand, and the demand for the first item also generates demand for one of the second. Systems parameters were evaluated. A case of substitutable products with joint demand and joint ordering policy was also considered in Yadavalli et al (2005b). A multi-item inventory with fuzzy deterministic demand has also been considered. (Yadavalli et al. 2005a)

Lee and Hsu (2009) is a model of a two-warehouse inventory management of a free form time dependent demand, where both the replenishment rate and planning horizon are finite. They used an approach which permits variation in production cycle time to determine the number of production cycles and time of replenishment during a finite planning horizon. Ferguson et al (2007) showed that EOQ model with nonlinear holding cost is an approximation of optimal order policy for perishable goods sold in small to medium size grocery stores where there are delivery surcharges due to infrequent ordering, and managers frequently utilize markdowns to stabilize demand as the product's expiration date gets nearer. They showed how the holding cost curve parameters can be estimated via a regression approach from the product's usual holding cost (storage plus capital costs), lifetime, and markdown policy.

Ho et al (2007) considered the effects of deteriorating inventory on lot-sizing in material requirements planning systems. They used simulation studies to evaluate the performance of five existing heuristics using three factors: rate of inventory deterioration, percentage of periods with zero demand, and setup cost. Hwang and Hahn (2000) investigated an optimal procurement policy for items with an inventory level-dependent demand rate and fixed lifetime, being a case for a fish cake retailer. Lin and Gong (2006) considers the impact of random machine breakdowns on the classical Economic Production Quantity (EPQ) model for a product, manufactured in batches, and subject to exponential decay and under a no-resumption ($\boldsymbol{NR}$) inventory control policy. The time-to-breakdown also follows an exponential distribution.

Chung and Wee (2007) developed an integrated deteriorating inventory policy for a single-buyer, single-supplier model with multiple $\boldsymbol{JIT}$ deliveries considering the transportation cost, inspection cost and the cost of less flexibility. Shah and Shukla

(2009a) presented an algorithm and models for a retailer's optimal procurement quantity and the number of transfers from the warehouse to the display area are determined when demand is decreasing due to recession and items in inventory are subject to deterioration at a constant rate. They also presented a deterministic inventory model in Shah and Shukla (2009b) where items are subject to constant deterioration and shortages are allowed. The unsatisfied demand is backlogged as a function of time.

Baten and Kamil (2010) presented a continuous review model for the control of production-inventory system subject to generalised Pareto distributed deterioration. They used the principle of control theory to determine what should be the optimal level of inventory in the system. Benhadid, et al (2008) also used control theory to show how to manage inventory in a production system with deteriorating items and dynamic costs.

Inventory models with Markov Arrival Processes ($\boldsymbol{MAP}$) and/or retrial queues have not been fully studied. The study of systems with $\boldsymbol{MAP}$ input systems have been focused mainly in telephone network systems. This has been highlighted in Gomez-Corral (2006) and Artalejo (1999). The only inventory related $\boldsymbol{MAP}$ input literatures documented is in Gomez-Corral (2006), and it was done by Krishnamoorthy et al. (2003, 2004) and even then, the inventory focus is also related to communication system as well. Some works have started being reported in this area. Yang and Templeton (1987) is another review. Lian, Liu and Zhao (2009) presented a continuous review model for a one item product where the demand has a distribution that is the Markov Arrival Process. The lifetime of the product is exponentially distributed with a constant failure rate λ. All arrival demand requests only one unit of item and all unmet demand is backordered.

Manuel et al. (2007) developed a continuous review perishable ($\boldsymbol{s}, \boldsymbol{S}$) model where there is an $\boldsymbol{MAP}$ arrival and $\boldsymbol{PH}$ service time. There is also a negative flow of unsatisfied customer, following the $\boldsymbol{RCE}$ policy for removal of customers. System parameters were determined. Yadavalli et al (2006) have also presented a model of service facilities where customers do not receive services immediately but have to wait till some services are performed on these products being waited for before the product is brought into stock. Two cases were considered: first where the product is brought in immediately after service; and the second case was where the product is brought in only at the next epoch. System parameters were determined. A model of perishable inventory in a random environment according to an alternating renewal process has also been studied in Yadavalli and Van Schoor (2004). The rate of perishing depends on the state of the system. Generally, it does not appear as if a lot has been done in deteriorating inventory systems with $\boldsymbol{MAP}$ arrival pattern and/or $\boldsymbol{PH}$ service pattern.

1.4. STOCHASTIC PROCESSES

Lindsey (2004) defined a stochastic process as some phenomenon that evolves over time (i.e. a process) and that involves a random component. It involves some response variable x_t that takes values varying randomly and in some way over time $t = 1 \dots T \; or \; 1 \dots$. and/or space $n = 1 \dots n \; or \; 1 \dots$. The variable may also be a scalar or vector. The observation of a state (or a change of state) is called an event. Usually, the probabilities of possible events would be conditional on the state of the process. The main properties, among other things, distinguishing a stochastic process are:

- The frequency or periodicity with which observations are made
- The set of all its observable values (state space)
- The sources and forms of randomness present, including the nature of the dependence among the values in a series of realisations
- The number of copies of the process available (only one or several)

1.4.1. Distribution and Transformation of the Random Variable

A random variable can be defined as a real-valued function defined over a sample space. The distribution of a random variable is the sample space of all its possible outcomes and the probability of each one occurring. The distribution function of a random variable plays an important role in the determination of the various parameters of the system in which it occurs.

Solving the state equations of a variable, especially since it is usually a joint distribution, could be quite challenging. It usually necessitates the need to transform the variable from one form to another in which it could be handled in a more straight forward manner. Bocharov et al. (2004) has used the term characteristic transform to describe all the transformations that are used in such manner. This term, he stated, comprises of the characteristic function, Laplace-Steiltjes transform and the moment generating function, depending on which ever is best applied.

1.4.2. Other Properties of the Stochastic Process

Some other issues that would be worth mentioning, apart from the randomness of the variable(s) and its distribution, are state dependence, serial dependence, stationarity, equilibrium, ergodicity, and regeneration point.

A stochastic process is said to be state dependent if the probability of being in a future state is dependent on the present state in which the state is found. This principle is exploited in Markov processes.

A stochastic process is said to have serial dependence if some parameters of the system depend not directly on the previous state of the system, but somehow on the previous state and the prediction at that time. It is a useful mechanism in time series analysis. Such dependencies could be on the location parameter, as in most such models, or on the spread parameter as in heteroscedastic models.

A stochastic process is said to be strictly stationary if sequences of consecutive responses of equal length in time have identical distributions. This means the values of the statistical parameters of the process are assumed constant with respect to time.

A process is said to be in equilibrium if the flow of a parameter of interest (including probability) into and out of a space (or point) balances out. The process may not be in equilibrium when it starts, but may enter a state of equilibrium over time, making it possible to observe its behaviour before entering equilibrium (i.e. while in transit – transient properties) and when it has entered equilibrium. In other words, if equilibrium has been reached, the probability that the process is in a given state, or the proportion of time spent in a given state, has converged to a constant that does not depend on the initial condition, and in essence the system become quite stationary.

Ergodicity is a concept quite related to equilibrium. Ergodic theorems provide identities between probability averages, such as an expected value, and the long run averages over a single realisation of a process. Thus, if the equilibrium probability of being in a given state equals the proportion of a long time period spent in that state, it is called an ergodic property of the process.

A regeneration point is a time instant at which the process returns to a specific state such that the future evolution of the process does not depend on how that state was reached. This means whenever a process arrives at the regeneration point, all of its previous history is forgotten. The renewal process, describing the time between recurrent events, is a well known case of such.

1.4.3. Types of stochastic processes and methods of observation

Basically, there are two main types of stochastic processes: survival processes and recurrent processes. The basic natures of each of these processes also affect the natures of its observations.

Survival processes are those that involve entering into a final state at which the process could be assumed to have terminated. Such processes are very useful in reliability studies in which the process of interest may not have the opportunity to regenerate itself. This limits the type of methods available for its study.

Recurrent processes are characterised by the possibility of the occurrence of more than one event (usually taken as two states in regeneration processes) over the time of study. One state is assumed to dominate while the other occurs occasionally. The latter that sparsely occurs is treated as a point event, and by focusing on its process of occurrence, the process is referred to as a stochastic point process. In contrast to a survival process, the point process only signals a transitory stage such that the event does not really signal a change of state. A binary indicator can, therefore, be used to signify a 1 if the point process occurs and a 0 otherwise. The process can, thus, be called a binary point process.

1.4.4. Method of Observation, Replications and Stopping Time

Two approaches could be used to observe accurate information from a stochastic process.

- One series for a long enough period (if it is reasonably stable)
- Several short replications of the process (if they are reasonably similar)

The nature of survival processes has confined their observation strictly to the second method since the process enters into an absorbing state. But for recurrent processes, one may use either of the two. Using the second method in a recurrent system raises the question of specifying an appropriate time origin. But in a stationary process, the principle of ergodicity makes it fairly simple to use the first method. The regeneration point process then acts as the appropriate time origin from which a datum could be taken for the initialisation of the observation process again.

Cinlar (1975) has defined a stopping time as any random time, T, having the property that for every $n \in N$ the occurrence or non-occurrence of an event $\{T \leq n\}$ can be determined by looking at the values of $x_o \ldots . x_n$.

1.4.5. Observation of Variables of interest

The variable of interest in a stochastic process could be one or more of the following:

- The inter occurrence time i.e. the duration between the occurrence of two consecutive events of interest, e.g. the time between two consecutive regeneration points
- The count of the number of occurrence of an event in a given interval e.g. the number of regenerations or renewals that have occurred between two periods of time
- The cumulative number of events of interest that occurred till date

The subject of renewal theory seeks to answer these questions. A summary of an overview of Renewal process, Markov theory and Queuing theory is included in Appendix 2.

1.5. POPULAR MANAGEMENT PHILOSOPHIES

Production managers have different perceptions about the importance and significance of the different system slacks. While some would not accept the presence of significant idle capacities, others are more critical of excess inventory. The decision about which one appears more critical is also dependent on the production philosophy. But the philosophies address not only issues of system slacks, but also issues of quality and job scheduling among others. This is because these are surrogate issues to the issues of slacks themselves.

Inventory is present in these systems, both as a stock build up, consequent to the job scheduling and flow management techniques as well as a result of deliberate actions of building up strategic reserves as an insurance against demand and lead time uncertainties. While there could be many other ideologies considered as management philosophies, the discussion here is limited to Lean Manufacturing, material Requirement Planning ($\boldsymbol{MRP}$) and the Theory of Constraints ($\boldsymbol{TOC}$). Just an overview of these would be provided also. Volmann et al. (2005), Jacobs et al. (2009), Goldratt and

Cox (2004) and Jonsson and Mattsson (2009) are good further readings for the interested reader for further treatment of the philosophies.

Lean manufacturing is a system that would prefer to pull entirely through the system. It apparently is more critical of excess inventory than spare capacity. In the ideal Lean environment, replacement of outputs or inputs should be lot for lot. This does not give consideration to issues of set up (both of purchase and production). To achieve this, effort goes into eliminating causes of bad quality as well as lead time variation in the system. Efforts are also put into managing demand so that the production rate is quite level. Kanban is used both to control the level of allowable inventory as well as scheduling tasks. Efforts for continuous reduction of set up times are also made consistently in Lean systems.

The Material Requirement Planning ($\boldsymbol{MRP}$), however, has a less critical view of inventory. Inventory is used to support utilisation of resources. Production is back-scheduled. Extra inventory is allowable as safety stock along various points in the network, and capacity utilisation is usually higher than that obtained in Lean.

Theory of Constraints ($\boldsymbol{TOC}$) also has a critical view of inventory in a manner probably similar to the Lean technique. It also would, however, not only allow for spare capacities in the various locations in the production network, but believes they are good. These spare capacities are used to break the production batches of such systems further down to the end that the average work-in-process inventory is further minimised. Strategic reserves are allowed in certain parts of the network where they are used to support the most critical station.

In a $\boldsymbol{TOC}$ environment, the critical station should be fully exploited, but only to the point where it does not also create an unnecessary inventory (finished good or work-in-process). Productivity is different from activation of resources. Productivity is about actual sales and not hours worked. Throughput is only about products that the market is ready to absorb and convert to money, and not just finished product. Finished product not going for sale is just another “undesirable” inventory. Scheduling is about creating an imaginary rope from the strategic buffer locations to the entry point to the flow line, and that suffices to control the flow through all processing stations in the entire line.

An important issue is the treatment of the statistical variations in the processing time and the complex stochastic and dynamic nature of demand that are basically not directly implied in all these models. Determinism is somehow implied to a large extent in the deployment of all these processes. This is the cause of system nervousness in such processes and their treatment has not been fully studied by researchers.

Of particular interest is the determination of the ideal buffer size to place ahead of the critical work station. This station could be a Bottleneck ($\boldsymbol{BN}$) or a Capacity Constrained Resource ($\boldsymbol{CCR}$) depending on if it has demand for production that is more than its capacity or close to its capacity respectively. While $\boldsymbol{TOC}$ seeks to eliminate unnecessary inventory in the system, it deliberately keeps time buffers ahead of the critical station to eliminate unplanned resource idleness and at junctions where other lines meet the critical line to eliminate waiting for parts or components along the critical line. The determination of this buffer size and its relationship to the flow rate in a $\boldsymbol{TOC}$ environment is an issue that still needs investigation, especially in the light of possible variation in resource processing time.

1.6. RESEARCH FOCUS AND CONTRIBUTION

1.6.1. Area of Interest

It has been stated that the aim of the supply chain management is a holistic approach for managing production throughout the entire production network, whereby some of its issues focus on the management of stations and some on the links. Issues of interest in station management relate to those of the traditional productivity and quality issues while issues of link management are those of logistics and information systems.

The focus of this work is on some of the station management principles. The main focus in stations is actually on the management of flows. Of particular interest is in the strategic management of inventories in the system as a result of the variability in the supply chain. Inventory has been mentioned earlier as strategic reserves of materials. They are said to occur both as deliberate strategic stocks and accumulation of flows in the production network.

Queuing principles are the basic tools used throughout this work. In some instances, it was used to determine the steady state parameters of some selected systems of interest. In other instances, the steady state parameters of some queuing processes were used to derive the control parameters (optimal feed rate) of some specific queuing processes considering a particular Operations Management principle.

1.6.2. Contributions to Knowledge

The purpose of this research in station flows in a supply chain is two pronged:

a) The first main contribution in this work is to the body of knowledge in the area of management of production system due to the nature of input system (i.e. pattern or arrival of demand from outside the production network). This involves the understanding of how the system behaves due to the nature of the demand and the characteristics of the processing centre. Zipkin (2000) has noted quite well that the only time in a supply chain when variability in input or processing time becomes important is during lead time, when there is a reasonable possibility of not meeting demand due to non availability of stock, and the attendant cost implication. So, the modelling interest is to understand the joint distribution of demand and lead time so that the steady state distribution of such system is determined, and from there, the system parameters can be calculated.

   This area is actually well researched, and there exists many probability models that have been developed as such. But the area is not yet full researched as there are still cases of some possible input types and demand characteristics not yet solved (e.g. the various MAP and PH distribution considerations being done in this thesis). The theoretical probability distribution of some such Markov processes were developed in this regards in chapters 2 to 4.

b) The second main contribution is in the area of management and accumulation of flows. The Theory of Constraints philosophy was particularly used as the reference philosophy. Contributions are made in the management of flow in a production environment that utilises this theory. This area appears to have an enormous potential for studies by applying the solutions of some of the steady state parameters of the various queuing processes already derived in regulating flows in such production environment. But the area does not appear well researched, and so, considered in this work.

1.7. CHAPTER OVERVIEW

The first chapter of this work contains the background to the study and a review of the relevant literature. The focus of the research is defined and the anticipated contributions to the field of learning were stated.

In chapter two, a multi-server service facility of a perishable inventory system with negative customer is presented. The item demanded is presented to the customer only after some service has been performed on the item. The inventory is depleted at the service rate rather than the demand rate. The arrival of customers follows a Markov Arrival Process (MAP) and the service time has an exponential distribution. The ordering policy is (s, S), and the lead time has exponential distribution. A customer whose service could not be provided immediately moves into an orbit of infinite size, from where requests are sent back to the system at random intervals characterised by exponential distribution. In addition, a second flow of negative customers following an MAP removes one of the customers from the orbit. The joint probability of the number of busy servers, the inventory level and the number of customers in the orbit is obtained at the steady state. Various stationary system performance measures were calculated, and the result illustrated numerically.

Chapter three is a study of a continuous review retrial inventory system with a finite source of customers and identical multiple servers in parallel. The customers arrive according to a quasi-random distribution. The customers demand unit items which are then delivered after some service has been performed on the items. The re-ordering policy is $(\boldsymbol{s}, \boldsymbol{S})$, and its distribution is assumed to be exponential. A customer with unfulfilled order joins an orbit from which only customers selected based on certain rules can reapply for service. The joint probability distribution of the number of customers in the orbit and the steady state number of busy servers and inventory level are obtained. Measures of system performance were derived.

Chapter four is a study of two-commodity perishable inventory with bulk demand for one commodity. It is a continuous review process in which three flows of customers could demand single item of the first, bulk item of the second or both single item of the first and bulk of the second. The arrival pattern is assumed to be $\boldsymbol{MAP}$. Order policy is to place order for both items when inventory levels are below the fixed levels for both commodities. The lead time is assumed to have a phase type distribution and the demands that occur during the stock out period are lost. The joint probability distribution for both commodities is determined and the various measures of system parameters and the total expected cost rate in the steady state are derived and numerical illustration was done.

Chapter five studies the management of flow in a production environment managed through the Theory of Constraints approach. The system is a continuous or discontinuous flow process with a Poisson input flow and an exponential service time. The system is assumed to have only a Capacity Constrained Resource and no Bottle neck. The option of using a regulated input flow to dynamically control the buffer placed ahead of the critical resource to cover for variations in processing time was shown to provide better management approach than a case where a predetermined buffer size is placed ahead of the resource. This model was further modified to incorporate payment of penalty charges for cases of lost throughput. A formula for determining the optimal flow rate to allow in the system to maximise the system profit was developed. The effect of shortages on the system parameters was illustrated graphically.

Chapter six is basically the concluding overview, the contextualisation of some possible applications of the models developed in the thesis, and the identification of some suggested areas for further future research.

2. *

CHAPTER 2

A MULTI-SERVER PERISHABLE INVENTORY SYSTEM WITH NEGATIVE CUSTOMER

* A modified version of this chapter has been submitted to Computers and Industrial Engineering Journal. The revision has been completed and re-submitted.

2.1. INTRODUCTION

Stochastic inventory models in which the demanded item is not immediately delivered to the customer are being considered by many authors. As the item in the stock may require some time for installation or preparation etc, the time taken to deliver to the customers is positive and usually random. As this causes formation of queues, the inventory manager needs to consider the queue length as well as the waiting time apart from the mean inventory level, holding time, etc to evaluate the system performance and hence to implement various control policies.

Berman et al (1993) considered an inventory management system at a service facility which uses one item of the inventory for each service provided. They assumed that both demand and service rates are deterministic and constant, and queues can form only during the stock outs. They determined optimal order quantity that minimises the total cost rate. Berman and Kim (1999) analysed a problem in a stochastic environment where customers arrive at a service facility according to a Poisson process. The service times are exponentially distributed with mean inter arrival time which is assumed to be larger than the mean service time. Each service requires one item from the inventory. Under both the discounted and average cost cases, the optimal policy of both finite and infinite time horizon problems is a threshold ordering policy.

A logically related model was studied by He et al. (1998), who analysed a Markovian Inventory-Production system, in which the demands are processed by a single machine in a batch size of one. Berman and Sapna (2000) studied an inventory control problem at a service facility which requires one item of the inventory. They assumed Poisson arrivals, arbitrarily distributed service times and zero lead times. They analysed the system with a finite waiting room. Under a specified cost structure the optimal ordering quantity that minimises the long run expected cost per unit time has been derived.

Sivakumar and Arivarignan (2006) considered an inventory system with service facility and negative customers. Schwarz et al (2006) have considered an inventory system with

Poisson demand, exponentially distributed service time and deterministic and randomised ordering policies. Manuel et al (2008) analysed an inventory system with service facility and finite waiting hall. They assumed the customers arrive according to a Markovian arrival process, the service times have phase-distribution, the lead time of the reorder and the life time of each item are exponential. When the waiting hall is full, an arriving customer joins the orbit of infinite size and after a random time, the customer tries his/her luck. Yadavalli et al (2008) considered an inventory system with service facility and infinite waiting hall. They assumed that demands occur according to a renewal process with instantaneous supply of reorders.

In all the above models, the authors assume that the service facility had a single server. But in many real life situations, the service facility has more than one server, and this is incorporated in this paper by assuming multiple servers. It was also assumed that any arriving customers who find all the servers are busy or all the items are in service enters into an orbit of infinite size to try their luck again sometime later.

Queues in which customers are allowed to conduct retrials have been widely used to model many problems in production/manufacturing engineering, communication engineering, etc. A complete description of situations where queues with retrial customers arise can be found in Falin and Templeton (1997). A classified biography is given in Artalejo (1999). For more details on multi-server retrial queues, see Anisimov and Artalejo (2001), Artalejo and Gomez-corral (2008), Artalejo et al (2001,2007), and Chakravarthy and Dudin (2002).

The rest of the paper is organised as follows. The next section gives a description of the mathematical model and the notations used. The steady state analysis of the model is presented in section 3. In section 4, various system performance measures in the steady state were derived. In the final section, the total expected cost rate in the steady state was derived and the results are illustrated using numerical examples.

2.2. MODEL DESCRIPTION

Consider the service facility which can stock a maximum of S units and $c(\geq 1)$ identical servers. The customers arrive according to a Markovian Arrival Process ($\boldsymbol{MAP}$) with representation (C_0, C_1) where C's are of order $m_1 \text{x}\, m_1$. The underlying Markov Chain $J_1(t)$ of the $\boldsymbol{MAP}$ has the generator $C(= C_0 + C_1\,)$ and a stationary distribution vector v_1 of length m_1. The stationary arrival rate is given by $\lambda_1 = v_1 C_1 \boldsymbol{e}$, where $\boldsymbol{e}$ is a column vector of appropriate dimension containing all ones. For more details on $\boldsymbol{MAP}$ and their properties, the reader may refer to Neuts (1995). If a new customer finds that anyone of the servers is idle, he/she immediately accedes to the service. The customer who finds either that all servers are busy or there is no service item (excluding those in service) in stock enters into an orbit of infinite size. These orbiting customers send requests at random time points for possible selection of their demands. The interval time between two successive request-time points is assumed to have exponential distribution with parameter θ. It is assumed that the access from the retrial group to the service facility is governed by the constant retrial policy described in Falin and Templeton (1997); i.e. the probability of repeated attempt during the interval $(t, t + \Delta t)$, is given by that $\theta\Delta t + o(\Delta t)$ as $\Delta t\ \rightarrow 0$. The service times have exponential distribution with rate μ both for primary customers and successful repeat customers. The items are perishable in nature and the life time of each item has a negative exponential distribution with parameter $\gamma(> 0)$. It is also assumed that the servicing item cannot perish. The operating policy is as follows: as soon as the inventory level drops to $s(> c)$, a replenishment order for $Q(= S - s > s)$ items is placed. The lead time is assumed to have exponential distribution with parameter $\beta(> 0)$.

In addition to the regular customers, a second flow of negative arrival following a $\boldsymbol{MAP}$ with representation (D_0, D_1) where D's are of order $m_2 \text{x} m_2$ is also considered. The underlying Markov Chain $J_2(t)$ of the $\boldsymbol{MAP}$ has the generator $D(= D_0 + D_1)$ and a stationary distribution vector v_{-1} of length m_2. The stationary arrival rate is given by $\lambda_{-1} = v_{-1} D_1 \boldsymbol{e}$. A negative customer has the effect of removing a customer from the orbit. The removal policy adopted is $\boldsymbol{RCE}$, (removal of a customer from the end of the queue).

Notations

$[A]_{i,j}$: The element/sub matrix at $(i,j)th$ position of A

$\mathbf{0}$: Zero matrix

$e_n(m)$: A column vector of dimension n with 1 in the m^{th} position

I: An identity matrix

I_k: An identity matrix of order k.

$A \otimes B$: Kronecker product of matrices A and B

$A \oplus B$: Kronecker sum of matrices A and B

$W = \{0,1,\dots,\}$

$$h(x) = \begin{cases} 1, & if \quad x \geq 0; \\ 0, & if \quad x < 0; \end{cases}$$

$$\delta_{(i,j)} = \begin{cases} 1, & if \quad i = j; \\ 0, & otherwise; \end{cases}$$

$\overline{\delta}_{(i,j)} = 1 - \delta_{(i,j)}$

$E_i = \{1,2,\dots,i\}$

$E_i^0 = \{0,1,\dots,i\}$

2.3. ANALYSIS

Let $X(t), L(t), Y(t), J_1(t)$ and $J_2(t)$, respectively, denote the number of customers in the orbit, the on-hand inventory level, the number of busy servers, the phase of the arrival of ordinary demand process and the phase of the arrival of the negative demand process at time t. From the assumptions made on the input and output processes, it can be shown that the stochastic process $\{X(t), L(t), Y(t), J_1(t), J_2(t); t \geq 0\}$ is a Markov process with state space given by

$$E = \{(i,k,m,u_1,u_2); i \in W, k \in E_{c-1}^0, m \in E_k^0, u_1 \in E_{m_1}^0, u_2 \in E_{m_2}^0\}$$
$$\cup \ \{(i,k,m,u_1,u_2); i \in W, k \in E_S \backslash E_{c-1}, m \in E_c^0, u_1 \in E_{m_1}^0, u_2 \in E_{m_2}^0\}.$$

Define the following ordered sets:

$$< i,k,m,u_1 > = ((i,k,m,u_1,1),(i,k,u_1,2),\dots(i,k,u_1,m_2)),$$

$$< i,k,m > = (< i,k,m,1 >,< i,k,m,2 >,\dots < i,k,m,m_1 >),$$

$$< i,k > = \begin{cases} (< i,k,0 >,< i,k,1 >,\dots < i,k,k >) & k \in E_{c-1}^0, \\ (< i,k,0 >,< i,k,1 >,\dots < i,k,c >) & k \in E_S \backslash E_c, \end{cases}$$

$$< i > = (< i,0 >,< i,1 >,\dots < i,S >).$$

Then the state space can be ordered as $(< 0 >,< 1 >,\dots)$.

The infinitesimal generator, P, of this process can be written in block partitioned form where the rows and columns correspond to $(< 0 >,< 1 >,\dots)$.

$$P = \begin{pmatrix} B_1 & A_0 & 0 & 0 & 0 & \dots \\ A_2 & A_1 & A_0 & 0 & 0 & \dots \\ 0 & A_2 & A_1 & A_0 & 0 & \dots \\ . & . & . & . & . & \dots \\ . & . & . & . & . & \dots \end{pmatrix} \tag{2.1}$$

where

$$A_0 = diag(H_0,H_1,\dots,H_{c-1},H_c,H_c,\dots H_c)$$

$$H_v = e_{v+1}(v+1)e_{v+1}^T(v+1) \otimes \left(C_1 \otimes I_{m_2}\right),\quad v \in E_c^0$$

$$A_2 = diag(F_0,F_1,\dots,F_{c-1},F_c,F_c,\dots F_c)$$

$$F_0 = I_{m1} \otimes D_1 \tag{2.2}$$

For $v \in E_c$

$$[F_v]_{k,l} = \begin{cases} I_{m1} \otimes D_1, & l = k, \quad k \in E_S^0 \\ \theta I_{m1} \otimes I_{m2}, & l = k+1, \quad k \in E_{v-1}^0 \\ 0, & otherwise \end{cases} \tag{2.3}$$

$$[A_1]_{k,l} = \begin{cases} M_k, & l = k, \quad k \in E_S^0 \\ N_k, & l = k-1, \quad k \in E_S \\ G_k, & l = k+Q, \quad k \in E_c^0 \\ G_c, & l = k+Q, \quad k \in E_S \backslash E_c \\ 0 & otherwise \end{cases} \tag{2.4}$$

$$G_k = J_k \otimes (\beta I_{m1} \otimes I_{m2}),\quad k \in E_c^0 \tag{2.5}$$

$$J_l = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ l \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & \cdots & l & l+1 & \cdots & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \end{array}, \quad l \in E_c^0 \tag{2.6}$$

$$N_1 = \begin{pmatrix} \gamma I_{m1} \otimes I_{m2} \\ \mu I_{m1} \otimes I_{m2} \end{pmatrix} \tag{2.7}$$

For $v \in E_c \backslash E_1$

$$[N_v]_{k,l} = \begin{cases} (v-k)\gamma I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_{v-1}^0 \\ k\mu I_{m1} \otimes I_{m2}, & l = k-1, \quad k \in E_v \\ \mathbf{0}, & otherwise \end{cases} \tag{2.8}$$

For $v \in E_S \backslash E_c$

$$[N_v]_{k,l} = \begin{cases} (v-k)\gamma I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_c^0 \\ k\mu I_{m1} \otimes I_{m2}, & l = k-1, \quad k \in E_c \\ \mathbf{0}, & otherwise \end{cases} \tag{2.9}$$

$$M_0 = C_0 \oplus D_0 - \beta I_{m1} \otimes I_{m2} \tag{2.10}$$

For $v \in E_{c-1}$,

$$[M_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{v-1}^0 \\ C_0 \oplus D_0 - (v\gamma + \beta + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \beta + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_{v-1} \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k = v \\ \mathbf{0} & otherwise \end{cases} \tag{2.11}$$

For $v \in E_s \backslash E_{c-1}$,

$$[M_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{c-1}^0 \\ C_0 \oplus D_0 - (v\gamma + \beta + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \beta + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_{c-1} \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k = c \\ \mathbf{0} & otherwise \end{cases} \tag{2.12}$$

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

For $v \in E_S \backslash E_s$,

$$[M_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{c-1}^0 \\ C_0 \oplus D_0 - (v\gamma + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_{c-1} \\ C_0 \oplus D_0 - ((v-k)\gamma + k\mu + \theta) I_{m1} \otimes I_{m2}, & l = k, \quad k = c \\ \mathbf{0} & otherwise \end{cases} \tag{2.13}$$

$$[B_1]_{k,l} = \begin{cases} \widetilde{M}_k, & l = k, \quad k \in E_S^0 \\ N_k, & l = k-1, \quad k \in E_S \\ G_k, & l = k+Q, \quad k \in E_c^0 \\ G_c, & l = k+Q, \quad k \in E_s \backslash E_c \\ 0 & otherwise \end{cases}$$

$$\widetilde{M}_0 = C_0 \oplus D - \beta I_{m1} \otimes I_{m2}$$

For $v \in E_{c-1}$,

$$[\widetilde{M}_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{v-1}^0 \\ C_0 \oplus D - (v\gamma + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D - ((v-k)\gamma + k\mu + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_v \\ \mathbf{0} & otherwise \end{cases} \tag{2.14}$$

For $v \in E_s \backslash E_{c-1}$

$$[\widetilde{M}_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{c-1}^0 \\ C_0 \oplus D - (v\gamma + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D - ((v-k)\gamma + k\mu + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_c \\ \mathbf{0} & otherwise \end{cases} \tag{2.15}$$

For $v \in E_S \backslash E_s$

$$[\widetilde{M}_v]_{k,l} = \begin{cases} C_1 \otimes I_{m2}, & l = k+1, \quad E_{c-1}^0 \\ C_0 \oplus D - v\gamma I_{m1} \otimes I_{m2}, & l = k, \quad k = 0 \\ C_0 \oplus D - ((v-k)\gamma + k\mu + \beta) I_{m1} \otimes I_{m2}, & l = k, \quad k \in E_c \\ \mathbf{0} & otherwise \end{cases} \tag{2.16}$$

It may be noted that A_0, A_1, A_2, B_1 are square matrices of order $\left(c, \frac{c+1}{2}\right) m_1 m_2 + (S-c)(c+1) m_1 m_2$, $F_i, H_i, i \in E_c^0$ are square matrices of order $(i+1) m_1 m_2$, $\widetilde{M}_i, M_i, i \in E_{c-1}^0$ are square matrices of order $(i+1) m_1 m_2$, $\widetilde{M}_i, M_i, i \in E_S \backslash E_{c-1}$ are square matrices of order $(c+1) m_1 m_2$, $N_i, i \in E_c^0$ are of order $(i+1) m_1 m_2 \text{ x } i m_1 m_2$, $N_i, i \in E_S \backslash E_c$ are square matrices of order $(c+1) m_1 m_2$, $G_i, i \in E_{c-1}^0$ are of order $(i+1) m_1 m_2 \text{ x } (c+1) m_1 m_2$, and G_c is a square matrix of order $(c+1) m_1 m_2$.

2.3.1. Stability Analysis

To discuss the stability condition of the process, consider $A = A_0 + A_1 + A_2$ which is given by

$$[A]_{k,l} = \begin{cases} \widehat{M}_k, & l = k, \quad k \in E_S^0 \\ N_k, & l = k-1, \quad k \in E_S \\ G_k, & l = k+Q, \quad k \in E_c^0 \\ G_c, & l = k+Q, \quad k \in E_S \backslash E_c \\ \mathbf{0} & otherwise \end{cases} \tag{2.17}$$

where

$$\widehat{M}_k = \begin{cases} M_k + F_k + H_k, & k \in E_{c-1}^0 \\ M_k + F_c + H_c, & k \in E_S \backslash E_{c-1} \end{cases} \tag{2.18}$$

Let Π denote the steady state probability vector of A, which satisfies

$$\Pi A = \mathbf{0}, \Pi \boldsymbol{e} = 1$$

The vector Π can be represented by

$$\Pi = (\pi^{(0)}, \pi^{(1)}, \cdots, \pi^{(S)})$$

where

$$\pi^{(i)} = \begin{cases} (\pi_{(i,0)}, \pi_{(i,1)}, \dots, \pi_{(i,i)}), & i \in E_{c-1}^0 \\ (\pi_{(i,0)}, \pi_{(i,1)}, \dots, \pi_{(i,c)}), & i \in E_S \backslash E_{c-1} \end{cases} \tag{2.19}$$

with

$$\pi_{(i,k)} = (\pi_{(i,k,1)}, \pi_{(i,k,2)}, \dots, \pi_{(i,k,m1)}), \quad i \in E_S^0, k \in E_c^0$$

and

$$\pi_{(i,k,l)} = (\pi_{(i,k,l,1)}, \pi_{(i,k,l,2)}, \dots, \pi_{(i,k,l,m2)}), \quad i \in E_S^0, k \in E_c^0, l \in E_{m1}$$

It can be easily shown that

$$\pi^{(i)} = \pi^{(Q)} \Omega_i, i \in E_S^0 \tag{2.20}$$

where

$$\Omega_i = \begin{cases} (-1)^{Q-i} N_Q \widehat{M}_{Q-1}^{-1} N_{Q-1} \dots N_{i+1} \widehat{M}_i^{-1} \\ \qquad\qquad i = 0,1,2,\dots,Q-1 \\ I, \qquad\qquad\qquad\qquad\qquad i = Q \\ (-1)^{S-i+1} \Omega_S \left[\sum_{j=0}^{S-c} \psi(s,j) G_c \eta(S-j,i) + \sum_{j=i-Q}^{c-1} \psi(s,j) G_c \eta(Q+j,i) \right], \\ \qquad\qquad i = Q+1, Q+2, \dots, Q+c-1 \\ (-1)^{S-i+1} \Omega_S \sum_{j=0}^{S-i} \psi(s,j) G_c \eta(S-j,i) \\ \qquad\qquad i = Q+c, Q+c+1, \dots, S \end{cases} \tag{2.21}$$

with

$$\psi(i,j) = \begin{cases} N_i \widehat{M}_{i-1}^{-1} N_{i-1} \dots \widehat{M}_{i-j}^{-1}, & j \geq 1 \\ I & j = 0 \end{cases}$$

$$\eta(i,j) = \widehat{M}_i N_i \widehat{M}_{i-1}^{-1} \dots \widehat{M}_j^{-1}. \tag{2.22}$$

and $\pi^{(Q)}$ can be obtained by solving

$$\pi^{(Q)} \left(\Omega_{Q+1} N_{Q+1} + \widehat{M}_Q + \Omega_0 G_0 \right) = \mathbf{0}.$$

and

$$\pi^{(Q)} \left(I + \sum_{\substack{k=0 \\ k \neq Q}}^{S} \Omega_k \right) \boldsymbol{e} = 1 \tag{2.23}$$

Now the following result obtains on the stability condition.

Lemma 1 The stability condition of the system under the study is given by

$$\begin{aligned} &\sum_{i=0}^{c-1} \pi_{(i,i)} (C_1 \otimes I_{m2}) \boldsymbol{e} + \sum_{i=c}^{S} \pi_{(i,c)} (C_1 \otimes I_{m2}) \boldsymbol{e} \\ &< \begin{pmatrix} \sum_{i=0}^{c-1} \pi_{(i,i)} (I_{m1} \otimes D_1) \boldsymbol{e} + \sum_{i=c}^{S} \pi_{(i,c)} (I_{m1} \otimes D_1) \boldsymbol{e} \\ + \sum_{i=1}^{c-1} \sum_{j=0}^{i-1} \pi_{(i,j)} (I_{m1} \otimes D_1 + \theta I_{m1} \otimes I_{m2}) \boldsymbol{e} \\ + \sum_{i=c}^{S} \sum_{j=0}^{c-1} \pi_{(i,j)} (I_{m1} \otimes D_1 + \theta I_{m1} \otimes I_{m2}) \boldsymbol{e} \end{pmatrix} \end{aligned} \tag{2.24}$$

Proof: From the well known result of Neuts (1994) on the positive recurrence of P, there exists

$$\Pi A_0 \boldsymbol{e} < \Pi A_2 \boldsymbol{e}$$

and by exploiting the structure of the matrices A_0 and A_2 and Π, the stated result follows.

2.3.2. Steady State Analysis

It can be seen from the structure of the rate matrix P and from the Lemma 1 that the Markov process $\{(X(t), L(t), Y(t), J_1(t), J_2(t))\ t \geq 0\}$ on E is regular. Hence, the limiting distribution is defined by

$$\emptyset^{(i,k,m,u_1,u_2)} = \lim_{t\to\infty} Pr[X(t) = i, L(t) = k, Y(t) = m, J_1(t) = u_1, J_2(t) = u_2 |\ X(0), L(0), Y(0), J_1(0), J_2(0)], \tag{2.25}$$

where $\emptyset^{(i,k,m,u_1,u_2)}$ is the steady-state probability for the state (i, k, m, u_1, u_2), exists and is independent of the initial state.

The probabilities $\emptyset^{(i,k,m,u_1,u_2)}$ can be grouped as follows:

$$\emptyset^{(i,k,l,u_1)} = (\emptyset^{(i,k,l,u_1,1)}, \emptyset^{(i,k,l,u_1,2)}, \dots, \emptyset^{(i,k,l,u_1,m_2)}),\ i \in W, k \in E_0^S, l \in E_c^0, u_1 \in E_{m1}$$

$$\emptyset^{(i,k,l)} = \left(\emptyset^{(i,k,l,1)}, \emptyset^{(i,k,l,2)}, \dots, \emptyset^{(i,k,l,m_1)}\right),\ i \in W, k \in E_0^S, l \in E_0^c$$

$$\emptyset^{(i,k)} = \begin{cases} \emptyset^{(i,k,0)}, \emptyset^{(i,k,1)}, \dots, \emptyset^{(i,k,k)}, & k \in E_{c-1}^0 \\ \emptyset^{(i,k,0)}, \emptyset^{(i,k,1)}, \dots, \emptyset^{(i,k,c)}, & k \in E_S \backslash E_{c-1} \end{cases}$$

and finally, write

$$\Phi^{(i)} = \left(\emptyset^{(i,0)}, \emptyset^{(i,1)}, \dots, \emptyset^{(i,S)}\right),\ i = 0,1,2, \dots \tag{2.26}$$

The limiting probability distribution $\Phi = (\Phi^{(1)}, \Phi^{(2)}, \dots)$ satisfies

$$\Phi P = 0,\ \Phi \boldsymbol{e} = 1. \tag{2.27}$$

Theorem 1: When the stability condition (2.24) holds good, the steady state probability vector, Φ, is given by

$$\Phi^{(j)} = \Phi^{(0)} R^{(j)},\ j = 0,1, \dots \tag{2.28}$$

where the matrix R satisfies the quadratic equation

$$R^2 A_2 + R A_1 + A_0 = \mathbf{0} \tag{2.29}$$

and the vector $\Phi^{(0)}$ is obtained by solving

$$\Phi^{(0)}(B_1 + R A_2) = \mathbf{0}. \tag{2.30}$$

subject to the normalising condition

$$\Phi^{(0)}(1-R)^{-1}\boldsymbol{e} = 1. \tag{2.31}$$

Proof: The theorem follows from the well known result of the matrix-geometric methods (Neuts, 1994).

2.3.2.1. Computation of the R matrix

In this subsection, an algorithmic procedure for computing the R matrix is presented, which is the main ingredient for discussing the qualitative behaviour of the system under study.

Due to the special structure of the coefficient matrices appearing in (2.29), the square matrix R of dimension $\left(\frac{c(c+1)}{2}\right)m_1m_2 + (S-c)m_1m_2$ can be computed as follows: Note that $A_0\boldsymbol{e}$ is of the form

$$A_0\boldsymbol{e} = \begin{array}{c} 0 \\ 1 \\ \vdots \\ c-1 \\ c \\ c+1 \\ \vdots \\ S \end{array}\begin{pmatrix} H_0\boldsymbol{e} \\ H_1\boldsymbol{e} \\ \vdots \\ H_{c-1}\boldsymbol{e} \\ H_c\boldsymbol{e} \\ H_c\boldsymbol{e} \\ \vdots \\ H_c\boldsymbol{e} \end{pmatrix}, \ H_i\boldsymbol{e} = \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \end{array}\begin{pmatrix} 0 \\ 0 \\ \vdots \\ (C_1 \otimes I_{m2})\boldsymbol{e} \end{pmatrix}, \ i = 0,1,2,\dots,c \tag{2.32}$$

Due to the special structure of A_0 matrix, the matrix R has only$(S+1)m_1m_2$ rows of nonzero entries as shown below

$$R = \begin{pmatrix} R_{(0,0)} & R_{(0,1)} & \cdots & R_{(0,S)} \\ R_{(1,0)} & R_{(1,1)} & \cdots & R_{(1,0)} \\ \vdots & \vdots & \ddots & \vdots \\ R_{(S,0)} & R_{(S,1)} & \cdots & R_{(S,S)} \end{pmatrix} \tag{2.33}$$

where

$$R_{(0,i)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & i \end{array} \\ 0 & \begin{pmatrix} R_{(0,i)}^{(0)} & R_{(0,i)}^{(1)} & \cdots & R_{(0,i)}^{(i)} \end{pmatrix} \end{array}, \quad i = 0,1,\dots,c-1$$

$$R_{(0,i)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & c \end{array} \\ 0 & \begin{pmatrix} R^{(0)}_{(0,i)} & R^{(1)}_{(0,i)} & \cdots & R^{(i)}_{(0,i)} \end{pmatrix} \end{array}, \quad i = c, c+1, \dots, S$$

$$R_{(i,i)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & i \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,i)} & R^{(1)}_{(i,i)} & \cdots & R^{(i)}_{(i,i)} \end{pmatrix} \end{array}, \quad i = 1, 2, \dots, c-1$$

$$R_{(i,i)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & c \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ c \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,i)} & R^{(1)}_{(i,i)} & \cdots & R^{(c)}_{(i,i)} \end{pmatrix} \end{array}, \quad i = c, c+1, \dots, S$$

$$R_{(i,j)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & j \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,j)} & R^{(1)}_{(i,j)} & \cdots & R^{(j)}_{(i,j)} \end{pmatrix} \end{array}, \quad \begin{array}{c} i = 1, 2, \dots, c-1 \\ j = i+1, i+2, \dots, c \end{array}$$

$$R_{(i,j)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & c \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,j)} & R^{(1)}_{(i,j)} & \cdots & R^{(c)}_{(i,j)} \end{pmatrix} \end{array}, \quad \begin{array}{c} i = 1, 2, \dots, c-1 \\ j = c+1, c+2, \dots, S \end{array}$$

$$R_{(i,j)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & j \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,j)} & R^{(1)}_{(i,j)} & \cdots & R^{(j)}_{(i,j)} \end{pmatrix} \end{array}, \quad \begin{array}{c} i = 1, 2, \dots, c-1 \\ j = 0, 1, \dots, i-1 \end{array}$$

$$R_{(i,j)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & j \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ c \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{(0)}_{(i,j)} & R^{(1)}_{(i,j)} & \cdots & R^{(j)}_{(i,j)} \end{pmatrix} \end{array}, \quad \begin{array}{c} i = c, c+1, \dots, S \\ j = 0, 1, \dots, c-1 \end{array}$$

$$R_{(i,j)} = \begin{array}{cc} & \begin{array}{cccc} 0 & 1 & \cdots & j \end{array} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ c \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \\ R^{0}_{(i,j)} & R^{1}_{(i,j)} & \cdots & R^{c}_{(i,j)} \end{pmatrix} \end{array}, \quad \begin{array}{c} i = c, c+1, , \dots, S \\ j = c, c+1, , \dots, S \\ i \neq j \end{array} \tag{2.34}$$

The matrix R^2 is also of the form R with only $(S+1)m_1m_2$ nonzero rows. This form is exploited in the computation of R using (2.29). The relevant equations are given in the appendix.

2.4. SYSTEM PERFORMANCE MEASURES

In this section, some stationary performance measures of the system are derived. Using these measures, the total expected cost per unit time can be constructed.

2.4.1. Mean Inventory Level

Let Γ_I denote the mean inventory level in the steady state. Since $\emptyset^{(i,k)}$ denotes the steady state probability vector for kth inventory level with each component specifying a particular combination of the number of customers in the orbit, the number of busy servers, the phase of the ordinary arrival process and the phase of the negative arrival process, the quantity $\emptyset^{(i,k)}\boldsymbol{e}$ gives the probability that the inventory level is k in the steady state. Hence, the mean inventory level is given by

$$\Gamma_I = \sum_{i=0}^{\infty} \sum_{k=1}^{S} k\emptyset^{(i,k)}\boldsymbol{e} \tag{2.35}$$

2.4.2. Expected Reorder Rate

Let Γ_R denote the expected reorder rate in the steady state. Note that a reorder is triggered when the inventory level drops from $s+1$ to s. The steady state probability vector $\emptyset^{(i,s+1,l)}$ gives the rate at which $s+1$ is visited. After the system reaches the inventory level $s+1$, either a service completion of any of the l servers if $l > 0$ or a failure of anyone of $s+1-l$ items trigger the reorder event. This leads to

$$\Gamma_R = \sum_{i=0}^{\infty} \sum_{l=1}^{c} l\mu\emptyset^{(i,s+1,l)}\boldsymbol{e} + \sum_{i=0}^{\infty} \sum_{l=1}^{c} (s+1-l)\gamma\emptyset^{(i,s+1,l)}\boldsymbol{e} \tag{2.36}$$

2.4.3. Mean Perishable Rate

Since $\emptyset^{(i,k,l)}$ is a vector of probabilities with i customers in the orbit, the inventory level is k and l busy servers, the mean perishable rate, Γ_P in the steady state is given by

$$\Gamma_P = \sum_{i=0}^{\infty}\sum_{k=1}^{c}\sum_{l=0}^{k-1}(k-l)\gamma\emptyset^{(i,k,l)}\boldsymbol{e} + \sum_{i=0}^{\infty}\sum_{k=c+1}^{S}\sum_{l=0}^{c}(k-l)\gamma\emptyset^{(i,k,l)}\boldsymbol{e} \qquad (2.37)$$

2.4.4. Mean number of customers in the Orbit

Let Γ_O denote the expected number of customers in the orbit. Since $\Phi^{(i)}$ is the steady state probability vector for i customers in the orbit with each component specifying a particular combination of the inventory level, number of busy servers, the phase of the ordinary customers arrival process and the phase of the negative customers arrival process, the quantity $\Phi^{(i)}$ gives the probability that the number of customers in the orbit is i in the steady state. Hence, the expected number of customers in the orbit is given by

$$\begin{aligned}\Gamma_O &= \sum_{i=1}^{\infty} i\,\Phi^{(i)}\boldsymbol{e}. \\ &= \Phi^{(0)}R(I-R)^{-2}\boldsymbol{e}. \end{aligned} \qquad (2.38)$$

2.4.5. Mean Rate of Arrival of Negative Customers

Let Γ_N denote the mean arrival rate of negative demand in the steady state. This is given by

$$\Gamma_N = \frac{1}{\lambda_{-1}}\sum_{i=1}^{\infty}\left[\emptyset^{(i,0,0)}(I_{m1}\otimes D_1)\boldsymbol{e} + \sum_{k=1}^{c-1}\sum_{l=0}^{k}\emptyset^{(i,k,l)}(I_{m1}\otimes D_1)\boldsymbol{e} + \sum_{k=c}^{S}\sum_{l=0}^{c}\emptyset^{(i,k,l)}(I_{m1}\otimes D_1)\boldsymbol{e}\right] \qquad (2.39)$$

2.4.6. The overall Rate of Retrials

If Γ_{OR} is the overall rate of retrials in the steady state, then overall rate of trials at which the orbiting customers request service is given by

$$\Gamma_{OR} = \theta \sum_{i=1}^{\infty} \Phi^{(i)} \boldsymbol{e}$$
$$= \theta \Phi^{(0)} R(1-R)^{-1} \boldsymbol{e} \qquad (2.40)$$

2.4.7. The Successful Rate of Retrials

Let Γ_{SR} denote the successful rate of retrials in the steady state. Note that the orbiting customer can enter the service if there is at least one free server and there is at least one item which is not in service. Hence, the successful rate of retrial, Γ_{SR}, is given by

$$\Gamma_{SR} = \theta\left[\sum_{i=1}^{\infty} \sum_{k=1}^{c-1} \sum_{l=0}^{k-1} \emptyset^{(i,k,l)} \boldsymbol{e} + \sum_{i=1}^{\infty} \sum_{k=c}^{S} \sum_{l=0}^{c-1} \emptyset^{(i,k,l)} \boldsymbol{e}\right] \qquad (2.41)$$

2.4.8. The Fraction of Successful Rate of Retrial

The fraction of successful rate of retrial is given by

$$\Gamma_{FSR} = \frac{\Gamma_{SR}}{\Gamma_{OR}} \qquad (2.42)$$

2.4.9. The Expected Number of Busy Servers

If Γ_{BS} denotes the mean number of busy servers in the steady state, it is given by

$$\Gamma_{BS} = \sum_{i=1}^{\infty} \sum_{k=1}^{c-1} \sum_{l=0}^{k} l\emptyset^{(i,k,l)} \boldsymbol{e} + \sum_{i=1}^{\infty} \sum_{k=c}^{S} \sum_{l=1}^{c} l\emptyset^{(i,k,l)} \boldsymbol{e} \qquad (2.43)$$

2.4.10. The Expected Number of Idle Servers

If Γ_{IS} denotes the expected number of idle servers in the steady state, then Γ_{IS} is given by

$$\Gamma_{IS} = \mathrm{c} - \Gamma_{BS} \tag{2.44}$$

2.4.11. The Blocking Probability

Let Γ_B denote the blocking probability in the steady state. This is given by

$$\Gamma_B = \sum_{\mathrm{i}=0}^{\infty} \sum_{\mathrm{k}=0}^{\mathrm{c}-1} \emptyset^{(i,k,k)} \boldsymbol{e} + \sum_{\mathrm{i}=0}^{\infty} \sum_{\mathrm{k}=\mathrm{c}}^{\mathrm{S}} \emptyset^{(i,k,c)} \boldsymbol{e} \tag{2.45}$$

2.5. COST ANALYSIS

The total expected cost per unit time (expected cost rate) in the steady state for this model is defined to be

$$TC(S,s,c) = c_h \Gamma_I + c_P \Gamma_P + c_S \Gamma_R + c_w \Gamma_O + c_{ne} \Gamma_N \tag{2.46}$$

where

c_S: Setup cost per order

c_h: Inventory carrying cost per unit item per unit time

c_p: Perishable cost per unit item per unit time

c_w: Backlogging cost per unit time

c_{ne}: Loss per unit time due to arrival of a negative customer

Substituting Γs the cost rate becomes

$TC(S,s,c) = c_h\{\sum_{i=0}^{\infty}\sum_{k=1}^{s} k\emptyset^{(i,k)}\boldsymbol{e}\} + c_p\{\sum_{i=0}^{\infty}\sum_{k=1}^{c}\sum_{l=0}^{k-1}(k-l)\gamma\emptyset^{(i,k,l)}\boldsymbol{e} +$

*i=0∞k=c+1Sl=0ck−lγ∅i,k,l**e**+csi=0∞l=1clμ∅(i,s+1,l)**e**+i=0∞l=0c(s+1−l)γ∅(i,*

*s+1,l)**e**+cwi=1∞iΦ(i)**e**+cne*1λ−1*i=1∞∅i,0,0Im1⊗D1**e**+k=1c−1l=0k−1∅i,k,lIm*

*1⊗D1**e**+k=cSl=0c∅i,k,lIm1⊗D1**e***

(2.47)

Since the computation of the ∅'s involve recursive equations, it is difficult to study the qualitative behaviour of the total expected cost rate analytically. However, the following numerical examples are presented to demonstrate the computability of the results derived in this work.

2.6. NUMERICAL ILLUSTRATIONS[†]

As the total expected cost rate is obtained in a complex form, one cannot study the qualitative behaviour of the total expected cost rate by the analytical methods. Hence, some 'simple' numerical search procedures have been used to find the "local" optimal values by considering a small set of integer values for the decision variables. With a large number of numerical examples, it was found out that the total cost rate per unit time in the long run is either a convex function or an increasing function of any one variable.

Consider the following $\boldsymbol{MAP}$'s for arrivals of regular demands as well as of negative demands. These processes can be normalised so as to have specific demand rate λ_1 (or λ_{-1}) when considered for arrivals of regular (negative) demands. Each of the $\boldsymbol{MAP}$ will be represented by (Z_0, Z_1), where Z_i's will represent C's for regular (positive) demands and D's for negative demands.

[†] Tables (2.2 to 2.19) referenced but not included in the body of this chapter could be found in Appendix 3

1 $\boldsymbol{Exponential\ (Exp)}$

$$Z_0 = (-1) \quad Z_1 = (1)$$

2 $\boldsymbol{Erlang\ (Erl)}$

$$Z_0 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad Z_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

3 $\boldsymbol{Hyper-exponential\ (HExp)}$

$$Z_0 = \begin{pmatrix} -10 & 0 \\ 0 & -1 \end{pmatrix} \quad Z_1 = \begin{pmatrix} 9 & 1 \\ 0.9 & 0.1 \end{pmatrix}$$

4 $\boldsymbol{MAP}$ with negative correlation ($\boldsymbol{MNC}$)

$$Z_0 = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -81 & 0 \\ 0 & 0 & -81 \end{pmatrix} \quad Z_1 = \begin{pmatrix} 0 & 0 & 0 \\ 25.25 & 0 & 55.75 \\ 55.75 & 0 & 25.25 \end{pmatrix}$$

5 $\boldsymbol{MAP}$ with positive correlation ($\boldsymbol{MPC}$)

$$Z_0 = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -81 & 0 \\ 0 & 0 & -81 \end{pmatrix} \quad Z_1 = \begin{pmatrix} 0 & 0 & 0 \\ 55.25 & 0 & 25.75 \\ 25.75 & 0 & 55.25 \end{pmatrix}$$

All the above $\boldsymbol{MAP}$s are qualitatively different in that they have different variance and correlation structures. The first three processes are special cases of renewal processes and the correlation between the arrival times is 0. The demand process labelled $\boldsymbol{MNC}$ has correlated arrivals with correlation coefficient -0.1254 and the demands corresponding to the process $\boldsymbol{MPC}$ has positive correlation coefficient of 0.1213. Since $\boldsymbol{Erl}$ has the least variance among the five arrival processes considered here, the ratios of the variances of the other four processes labelled $\boldsymbol{Exp}, \boldsymbol{HExp}$, $\boldsymbol{MNC}$ and $\boldsymbol{MPC}$ above, with respect to the $\boldsymbol{Erl}$ process are 3.0, 15.1163, 8.1795, 8.1795 respectively. The ratios are given rather than the actual values since the variance depends on the arrival rate which is varied in the discussion. The parameters and values have been chosen in such a way that the system is stable.

In the following discussions, the notations $\boldsymbol{MAP}+, \boldsymbol{Exp}+, \boldsymbol{Erl}+$, ... were used when the $\boldsymbol{MAP}s, \boldsymbol{EXP}, \boldsymbol{Erl}, \ldots$ were consider respectively for positive demands. When the process for negative demand were considered, the $+$ were replaced by $-$. For example, when a case with $\boldsymbol{HExp}$ were considered for positive demands and $\boldsymbol{MPC}$ for negative demands, this will be denoted by $(\boldsymbol{HEXP}+, \boldsymbol{MPC}-)$.

Example 2.1: In the first example, the optimum values, S^* and s^* that minimise the expected total cost rate were given for each of the five $\boldsymbol{MAP}$s for arrivals of regular demands considered against each of the five $\boldsymbol{MAP}$s for negative demands (see table 2.1). The associated expected total cost values are also given. The lower entry in each cell gives the optimal expected cost rate and the upper entries give the corresponding S^* and s^*. Fixing $\lambda_1 = 10, \lambda_{-1} = 4, c = 3, \beta = 3, \mu = 5, \gamma = 0.6, \theta = 5, c_h = 0.1, c_s = 10, c_p = 1, c_w = 9, c_{ne} = 10$, the following were observed:

1. For the case $(\boldsymbol{Erl}+, \boldsymbol{Erl}-)$, the optimal total cost rate and the optimal inventory level are smaller
2. For the case $(\boldsymbol{MPC}+, \boldsymbol{Hexp}-)$, the optimal cost rate is large
3. For the case $(\boldsymbol{HExp}+, \boldsymbol{HExp}-)$, the optimal inventory level is large
4. For the case $(\boldsymbol{Erl}+, \boldsymbol{Erl}-)$, the optimal inventory level is smaller

Example 2.2: The effect of correlation among positive demands and the correlation among negative demands on the total expected cost rate is studied in this example. Fixing $S = 25, s = 6, \lambda_1 = 6, \lambda_{-1} = 4, \beta = 3, \mu = 5, \gamma = 0.6, \theta = 5, c_h = 0.1, c_s = 10, c_p = 1,$ $c_w = 9, c_{ne} = 10$, the following were observed:

1. When the correlation coefficient of demands of the $\boldsymbol{MAP}+$ increases, the total expected cost rate increases. The same result is observed for $\boldsymbol{MAP}-$.
2. If the correlation among the positive demands increases, the total expected cost rates when computed for each of the $\boldsymbol{MAP}$s of negative demands increase. This

trend is observed for $c = 1,2,3$ and 4. But all the curves become almost equal when $c = 4$.

3. When the correlation among the negative demands increases, the total expected cost rate corresponding to $\boldsymbol{HExp}+$ approaches that of $\boldsymbol{MNC}+$. When the number of servers and the correlation in the $\boldsymbol{MAP}+$ increases, the difference between the total expected cost rate corresponding to $\boldsymbol{MPC}+$ and $\boldsymbol{MPC}-$ increases.
4. The total expected cost rates for $(\boldsymbol{MAP}+, \boldsymbol{Erl}-)$ for all $\boldsymbol{MAP}+$, have smaller value. The same is observed for $(\boldsymbol{Erl}+, \boldsymbol{MAP}-)$**.**
5. The total expected cost rates for $(\boldsymbol{MAP}+, \boldsymbol{HExp}-)$ and for $(\boldsymbol{MPC}+, \boldsymbol{MAP}-)$ have high values.

Table 2.1: MAP of arrivals

		MAP of negative arrivals									
		Exp-		Erl-		HExp-		MNC-		MPC-	
MAP of positive arrivals	Exp+	32.6872		31.1528		39.3456		35.5992		37.5572	
		34	8	33	7	37	10	35	9	36	9
	Erl+	25.9807		24.9220		30.2158		28.0187		29.0862	
		32	6	31	6	35	9	34	8	34	8
	HExp+	63.6298		60.7149		77.5237		69.0841		74.1758	
		41	12	40	12	43	13	42	13	42	13
	MNC+	52.2187		49.5678		65.0810		57.1639		61.5312	
		37	10	36	10	41	12	38	11	39	11
	MPC+	82.0489		78.8221		98.6941		88.0139		94.0573	
		41	12	40	12	42	13	41	13	42	13

Example 2.3: In this example, the effect of each of the following were illustrated: the positive demand rate λ_1, the negative demand rate λ_{-1}, the lead time β, the service rate μ, the retrial rate θ, the perishable rate γ, the number of servers, $(\boldsymbol{MAP}+, \boldsymbol{HExp}-)$, on the fraction of the successful rate of retrial, Γ_{FSR}. From tables 2.2-2.7, the following were observed:

1. As λ_1 increases, Γ_{FSR} increases, except for the $(\boldsymbol{MPC}+, \boldsymbol{Erl}-)$.
2. Except $c = 1$, the values of Γ_{FSR} decreases as λ_{-1} increases for the model $(\boldsymbol{RP}+, \boldsymbol{RP}-)$, where $\boldsymbol{RP}$ represents the renewal processes, $\boldsymbol{Exp}, \boldsymbol{Erl}$ and $\boldsymbol{HExp}$. (In each of these cases, there is no correlation among the arrivals of demands).

3. In the case of correlated demand processes, i.e. those cases of $(\boldsymbol{NRP}+, \boldsymbol{NRP}-)$, where $\boldsymbol{NRP} = \boldsymbol{MNC\ or\ MPC}$, Γ_{FSR} decreases with $\beta$and increases with θ, when $c \neq 4$.
4. But Γ_{FSR} increases with γ for all c values.
5. It was noted that for all values of c, Γ_{FSR} assumes low value when the input nature is $(\boldsymbol{Erl}+, \boldsymbol{Erl}-)$. It was also noted that this value approaches zero as c increases.

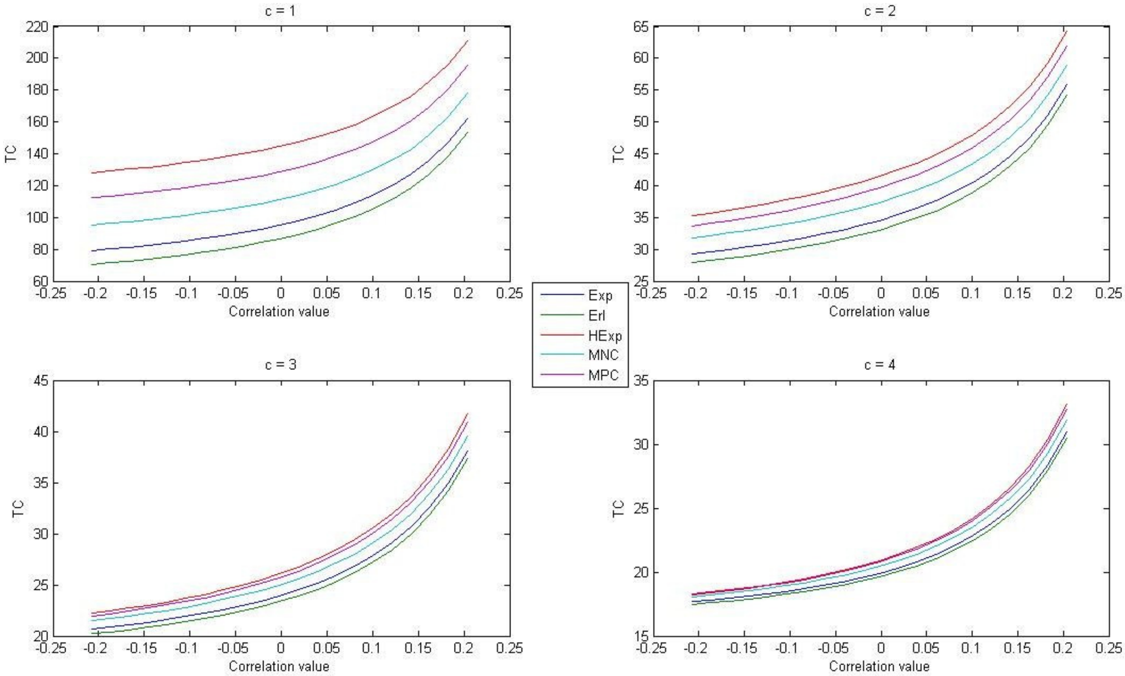

Figure 2.1: The effect of positive demand correlation on TC

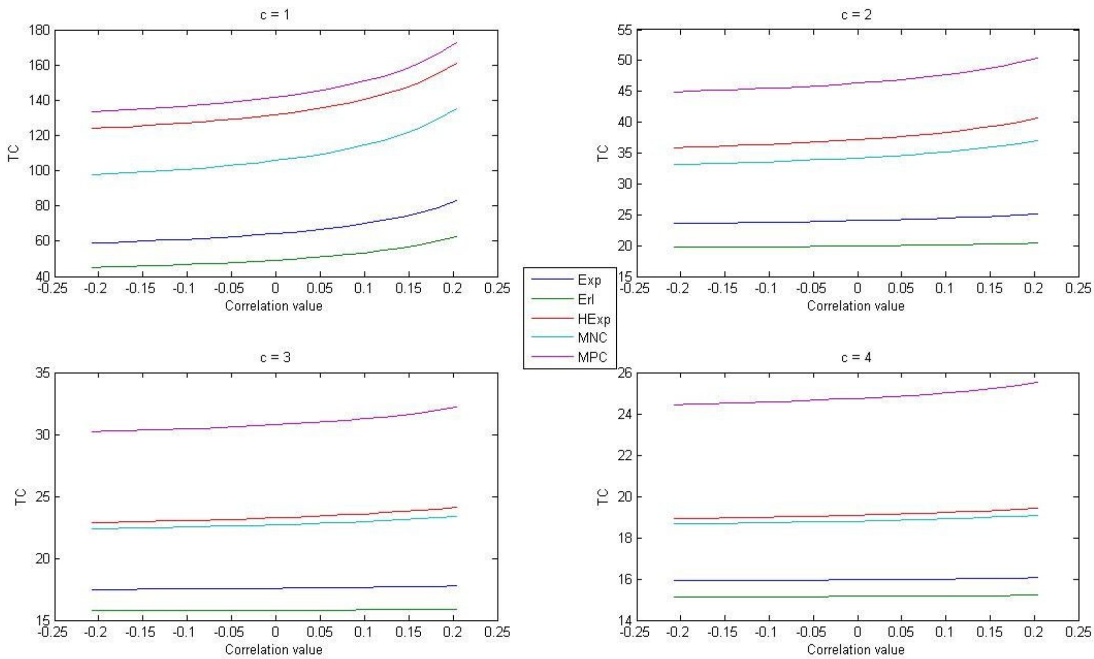

Figure 2.2: The influence of negative demand correlation on T C

Example 2.4: The influences of λ_1, λ_{-1}, β, μ, θ, γ, c and $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ on the blocking probability Γ_B is presented in this example. From tables 2.8 – 2.13, the following were observed:

1. Except for $c = 1$, as $\lambda_1$increases, Γ_B increases for each of the $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ process. For the single server case, as λ_1 increases, Γ_B decreases. The same behaviour is observed when θ increases.
2. Except for $c = 1$, Γ_B decreases when λ_{-1} increases.
3. Γ_B increases when the lead time rate β increases for each of the $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ process.
4. Whenever the number of servers is more than one, Γ_B increases with μ.
5. Γ_B increases with β for each of the $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ process.

Example 5: In this example, the effect of λ_1, λ_{-1}, β, μ, θ, γ, c, $\boldsymbol{MAP} +$ and $\boldsymbol{MAP} -$on the expected number of idle servers, Γ_{IS} were studied. From tables 2.14 - 2.19, the following were noted:

1. As is to be expected, as λ_1 increases, Γ_{IS} decreases except for single server case. This can be explained intuitively as follows. When the rate of positive customers increase, more number of servers would be engaged. This leads to decrease in the number of idle servers. For $c = 1$, Γ_{IS} increases with λ_1. This pattern is also observed for μ, θ.
2. Except for the single server case, Γ_{IS} increases as λ_2 increases. This is because as the negative customers frequently enter the orbit, they remove more customers from the orbit. Therefore, the number of retrying customers in the orbit decreases. Note that the servers will be occupied by both the positive demand and retrial customers. If the retrial customers' level decreases, then naturally, the customers from the orbit will also decrease. This forces the expected number of idle servers to increase.
3. As is to be expected, Γ_{IS} increases as β increases for each of the $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ process.
4. Except for $c = 4$, Γ_{IS} decreases as μ increases.
5. When γ increases, Γ_{IS} decreases for each of the $(\boldsymbol{MAP}+, \boldsymbol{MAP}-)$ process.

CONCLUSION

A continuous review perishable inventory system in a service facility with multi servers is studied in this work. The customers who could not get their demands attended to due to non-availability of items in stock or all the servers are busy join an orbit of infinite size. These customers attempt for service at random times. The customers are removed one by one by negative customers who could be touts of competing organisations. The novel attempt made in this work is to assume independent Markovian Arrival Processes $(\boldsymbol{MAP})$ for the positive demands and negative demands. By assuming $(\boldsymbol{MAP})$, one can also consider non renewal processes with correlated arrivals. Though, algorithmic solution is provided for this model, extended numerical examples were provided to discuss the behaviour of the expected total cost rate and the system performance measures due to changes or variations in the parameters.