Two Essays in Commonality

by

Hasan Arslan

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004
[June 2005]

Author ....
Sloan School of Management
September 14, 2004

Certified by ....
Thomas Roemer
Assistant Professor of Management
Thesis Supervisor

Certified by ....
Stephen Graves
Abraham J. Siegel Professor of Management Science & Engineering Systems
Thesis Supervisor

Certified by ....
Donald Rosenfield
Senior Lecturer
Thesis Supervisor

Accepted by ....
Birger Wernerfelt
Professor of Management Science, Chair PhD Program

Two Essays in Commonality

by

Hasan Arslan

Submitted to the Sloan School of Management
on September 14, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In the first part of this thesis we consider a manufacturer that introduces successive generations of products, possibly on multiple similar production lines. We consider two types of production phase for this firm; pilot production phase and full production phase. Pilot production phase for a newly developed product is an experimental production period in which a limited amount of production capacity of the production line is allocated to produce the new product so that production process of the new product is improved through cumulative production experience. At the conclusion of the pilot production phase, a production methodology is formalized and the obtained knowledge is utilized for full production.

The objective of the first part of this thesis is two-fold: First, to determine how to operate the pilot production phase of a newly developed product; Second, to characterize how production processes of a new product can be improved during its full production. To achieve the first objective, we develop two separate models. First, we consider a single production line alone. We analyze how to split production between the new and existing product to maximize profits, considering that the new product typically faces low yields initially. We show that it is never optimal to dedicate only limited capacity to the new product, — i.e., pilot production is not optimal. We then extend our view to multiple similar production lines when acquired knowledge can be passed on across production lines. We determine conditions under which launching pilot production becomes preferable. Finally, we consider the case of full production where a lead production line scans information on process improvements on a continuous basis with the remaining similar production lines during full production.

In the second part of this thesis we consider a single-product inventory system that serves multiple demand classes, which differ in their backlog costs or service level requirements. We develop a model for cost evaluation and optimization, under the assumptions of Poisson demand, deterministic replenishment lead-time, and a continuous–review (Q, R) policy with rationing. We show the value from a rationing policy and how to incorporate into a multi-echelon setting the single-item model with multiple demand classes.

Thesis Supervisor: Thomas Roemer
Title: Assistant Professor of Management

Thesis Supervisor: Stephen Graves
Title: Abraham J. Siegel Professor of Management Science & Engineering Systems

Thesis Supervisor: Donald Rosenfield
Title: Senior Lecturer

Acknowledgments

First, I thank the God for providing me the opportunity of pursuing my academic training at MIT among many good friends and colleagues.

I would like to express my warmest thanks to my academic and research supervisor Professor Thomas Roemer for his superb mentorship, professionalism, and constant care during my academic training at MIT. I have learned Operations Management through carefully listening and observing him. His passionate enthusiasm for research and teaching combined with profound intellectual integrity have become a great source of support and inspiration. I feel immensely privileged to have been his student. I am going to strive to teach my students what I learned from him.

I am greatly indebted to my other thesis committee members, Professor Steve Graves, and Professor Don Rosenfield, for their invaluable contribution and effort. Their great feedback was most helpful in improving this thesis as well as preparing my job talk. Professor Steve Graves has become a great role model for my academic life. He has always been a source of inspiration for my research and academic training. His impressive intellectual integrity, professional guidance, and generous research feedback greatly helped me with my intellectual growth. I feel extremely lucky to know such a unique personality.

I am greatly indebted to my great friend Anna Piccolo for her constant support, care, and friendship. I am also indebted to my other friends Sharon Cayley and Hillary Ross for their constant support and making our Ph.D program an enjoyable quality place.

Lastly, I want to thank all my friends whose constant care and support make my journey enjoyable.

Contents

List of Figures

Chapter 1

Introduction

1.1 Main Scope

This thesis deals with two important topics in Operations Management that have significant practical values. The first part of the thesis studies how pilot production strategies should be developed for production firms and how the existence of process commonality among different production facilities affect these strategies. The second part of the thesis, on the other hand, studies the inventory problem where multiple demand classes are served from the common inventory location. We analyze each subject separately. We present each subject by using the following methodology; we first construct the simplest setting and complete its analysis, then we extend our analysis to more general setting. For the clarity of our presentation, all supporting arguments and proofs will be presented in the appendices. We believe that our models and key findings both provide practical interest in each subject and establish new lines of further research in Operations Management field.

1.2 Process Commonality & Pilot Production

One common major challenge in many industries with shrinking product life cycles and selling prices is to begin high volume production when the underlying production process and technology are still ill-understood. Many firms are compelled to introduce new products constantly due to fierce competition in the market. At the same time, however, they suffer from low production yield levels associated with the lack of knowledge and experience in new production technologies. Hence, a through understanding of the dynamics of production yields is paramount for successful introduction of new products.

Production yield is defined as the ratio of usable parts to the total production quantity. When the production process is poorly understood, much of what has been produced would result in being defective, which causes having low production yield levels. As production continues, more problems related to the production process are identified and solved that provide better understanding about the production process. Through increasing production experience, employees gain better understanding of the production process, bottlenecks in production process are identified and circumvented, machine down-times are minimized, which all together result in better quality outputs. Hence, more production experience follows higher production yield levels, i.e, — learning-by-doing. These production yield improvements through production experience play an important role in profitability of a firm. For instance, 1% increase in production yield for a semiconductor fab producing 20, 000 wafers per month with an average selling price of \$2, 000/unit wafer can result in additional \$4.8M revenue in a year, which is almost entirely profit since only post production (i.e. storage and distribution) costs need to be covered by the \$4.8M.

As we have mentioned earlier, one major problem associated with the launching of a new product is the lack of production experience, which leads to low production yield levels. Pilot production phase for a newly developed product is an experimental production period in which a limited amount of production capacity of the production line is allocated to produce the new product so that the production process of the new product is improved through cumulative production experience. During pilot production phase manufacturers learn about the new recipes and tools required for the new product through small scale production. At the conclusion of the pilot production phase, the production methodology for the new product is better understood, formalized, and the obtained knowledge is exploited for full production. One major question that manufacturers, therefore, face is how to best learn about the production processes and corresponding yields of new product generations. This further begs the following two questions: is it better to improve the production process quickly by allocating a whole production capacity to experimental production of a new product? or is it better to improve the production process slowly, and still be able to reap profits from the old product that has already high and stable production yields? Answers to these questions are key to developing pilot production strategies, hence, there is a strong need to develop robust models to explore how fast to learn during pilot production phase.

In addition to process improvement through cumulative production experience, knowledge sharing is another opportunity to improve production process when there are multiple production lines with similar production processes. The production process of a new product is initially improved through cumulative production experience on a pilot production line. Then, acquired production experience and knowledge can be shared with the other production lines before they start launching the full production of the new product. Hence, having production lines with similar productions processes is helpful in carrying out knowledge transfer across production lines. For example, one of the largest automobile supplier, Visteon Corp., have recently started standardizing equipment and processes in its different manufacturing facilities to enable better knowledge sharing and to prevent duplicate improvement efforts among its plants. Intel's *copy exactly* strategy is another good example for the application of production lines with similar production processes. Under the copy exactly strategy, the whole manufacturing process (i.e., equipment set, process flow, suppliers, clean room, and training methods) devised in a development plant is duplicated across the other manufacturing plants for full production. As indicated in Clark (2002), by standardizing equipment and processes among different fabs through copy exactly, knowledge is transferred much faster and more efficiently among fabs. In addition, products and workers can be easily shifted among fabs to smooth operations, and engineers in different fabs with identical production lines can simultaneously deal with arising problems. When acquired knowledge can be passed on across production lines with similar production processes as pointed out by the aforementioned industry examples, it is essential to take this ability into account when developing effective pilot production strategies for newly developed products.

Production process during full production period can further be monitored to obtain additional yield improvements. When there are multiple production lines with similar production processes, a lead production line with better production experience would share information on process improvements on a continuous basis with the other production lines. Significant yield discrepancies among production lines during full production represent opportunities to share knowledge about the production process. For example, Intel Corp. exchanges employees across production lines with similar production processes and holds frequent meetings among their engineers to help close the gap in production yield levels. For example, Intel Corp. highly utilizes exchanging its employees among its production lines to make the detection of flaws in production processes easier. To illustrate, Bruce Sohn, the co-manager of Intel's Rio Rancho plant, reports that "identical tools in two factories of Intel kept producing different defect rates. By swapping the workers who maintained the tool, Intel learned that one group was cleaning the tool by wiping a towel in a circular motion; the other wiped back and forth. One motion was against the grain of the metal, spreading dirt particles rather than moving them." As supported by the previous example, exchanging labor force and carrying out knowledge sharing meetings among production lines with similar production processes are effective methods of further process improvements, however, continuous process improvement through knowledge and experience sharing process bears certain costs. For example, deployment of managers and engineers for meetings and exchanging employees among production lines cost both labor time and money. Hence, there is a strong need to develop a model that *i*) weights benefits from process improvements through knowledge sharing during full production and the resulting incurred costs; and *ii*) informs managers about when it would be best to carry out knowledge sharing among production lines.

To provide insights into the dynamics of production yields during ramp-up, this study considers a manufacturer that introduces successive generations of products, possibly on multiple production lines with similar production processes. We consider two types of production phase for this firm; pilot production phase and full production phase. Our objective in this paper is two-fold. First, we would like to determine how to operate the pilot production phase of a newly developed product. Second, we would like to characterize how the production process of a new product can be further improved during its full production phase. To determine how to operate the pilot production phase, we develop two separate models. First, we consider a single production line alone in chapter 2. We analyze how to split production between the new and old product to maximize total profits, considering that the new product typically faces low yields initially. We show that it is never optimal to dedicate only limited capacity to the new product, — i.e., experimental production is not optimal. We then extend our view to multiple similar production lines in chapter 2 when acquired knowledge can be passed on across production lines. We determine conditions under which launching pilot production becomes preferable. Next, in chapter 3, we consider the case of full production where a lead production line scans information on process improvements on a continuous basis with the remaining production lines. We develop a helpful yield control model that considers the yield gap between a lead production line and the other production lines during full production phase. Through this yield control model, we analyze the trade-off between the benefits from knowledge transfer from a lead production line during full production phase and any associated cost incurred through sharing knowledge. We show that this trade-off reveals a simple analysis that help managers in characterizing when knowledge sharing from a lead production line is desirable for a given the cost of knowledge sharing event. Finally, we summarize our main findings and point out some possible limitations and pitfalls to the approaches presented in this paper.

1.3 Service Commonality & Multiple Demand Classes

The demand for a product can often be categorized into classes of different priority. Rationing inventory among these different customer classes is an important tool for managing inventory when requirements for service vary greatly among customer classes. Issuing inventory to some customers while refusing or delaying the service for others is a commonly used practice in many industries. The following examples of inventory systems where different demand classes with different levels of service priority arise provide a motivation:

- A spare parts provider company is responsible to handle the inventory of serviceable parts of aircrafts. The company would face different types of demand for spare parts. Keeping an aircraft grounded can be very expensive. Moreover, the company has contractual agreements with some major airlines. Each of these airlines is assigned a specific service-level such that certain percent of its orders for spare parts must be filled within a pre-specified time window. Furthermore, there are some customers with whom no contractual agreement exist. The company, therefore, is in need of a operational support system to manage its spare-part inventory to minimize its operational costs while satisfying the promised service levels to its customers.

- A warehouse in a two-echelon inventory system for a single product has several retailers located in different regions with different market sizes and different profit margins. Each of these retailers offers three types of service delivery lead-time: 1 day delivery, 3 days delivery, and 1 week delivery. Therefore, each of these retailers faces 3 different types of demands for the single product, depending on different delivery lead-times chosen by customers upon their arrival. A manager faces the challenge of deciding how much inventory to place at warehouse and at each retailer to minimize the long term operational cost of this two-echelon inventory system.

Analyzing inventory models with different customer classes based on different service priorities has received some attention in the extant inventory management literature. However, existing models suffer from a lack of generality to cover most real problems or are very complicate to use. Hence, there is a strong need of having transparent inventory management models with multiple demand classes that are easy to understand and to be easily implemented in practice. Therefore, our objective in this study is to develop an approach that covers a wide variety of existing problems while being fairly transparent and easy to implement.

To achieve so, we consider a single-product inventory system that serves multiple demand classes, which differ in their backlog costs or service level requirements. We develop a model for cost evaluation and optimization, under the assumptions of Poisson demand, deterministic replenishment lead-time, and a continuous–review (Q, R) policy with rationing. We analyze three different problems: the *service level problem*, the *cost minimization problem*, and the *service time problem.* The main objective in each of these three problems is to find the optimal (Q, R) policy with rationing. The service level problem aims to achieve a pre-specified service level for each demand class by incurring the minimum long-run average inventory holding costs. The cost minimization problem, on the other hand, simply aims to minimize the sum of long-run average inventory holding and backlogging costs. In each of these aforementioned problems demand classes differ from each other in terms of their different service-level requirements and associated backlogging costs. The service time problem, however, considers a more elaborate approach to differentiate demand classes, where there is both an associated specific demand fulfillment lead-time and service level requirement to each demand class. It aims to satisfy the specified delivery time with the pre-specified service level for each demand class by incurring the minimum long-run inventory holding costs.

We develop transparent models to analyze each of the above mentioned three problems and provide effective numerical solution algorithms for each of them. Furthermore, we show the value of rationing policy and how to incorporate the single-item model with multiple demand classes into a multi-echelon setting.

Chapter 2

Pilot Production & Process Improvement

2.1 Introduction & Literature Review

The work in this chapter draws on both, the process improvement literature and the production yield management literature. The former branches into an area discussing pilot production and one dedicated to full production. For example, Bohn and Terwiesch (1999), Terwiesch and Bohn (2001), and Terwiesch and Xu (2004) discuss process improvement during pilot production phase. The main theme of these earlier works is that the production process can be improved by learning through deliberate experiments or process changes during pilot production phase. Experiments and process changes result in capacity reduction in the short run that causes a trade-off between how much to learn through experiments and process changes and how much to sacrifice from production to carry out experimentations and process changes. Our pilot production model in section 3 of this paper is related to these earlier studies since it focuses on interactions among capacity utilization and process improvement during pilot production. However, it differs from Bohn and Terwiesch (1999) and Terwiesch and Bohn (2001) in two aspects. First, we model the process improvement through cumulative production experience, i.e., learning-by-doing rather than carrying out controlled experiments. Therefore, production process in our model is improved by allocating more production capacity for pilot production rather more capacity for experiments. Hence, to this end, our model shifts the emphasis onto: trading production capacity among a new product undergoing pilot production and an old product that has been already in its full production phase. Second, our model follows a continuous time approach rather than a discrete time approach as used by the aforementioned works.

Since we model process improvement through learning with cumulative production experience, it is worth mentioning to cite what work has been done in the literature along this dimension. We begin with synthesizing production learning models in the literature. The traditional learning models in a typical manufacturing environment can be divided into two categories; autonomous learning and induced learning. Autonomous learning is the first generation learning model. It assumes that learning is merely a by-product of increased cumulative production volumes, — learning by doing. Induced learning model, on the other hand, is a deliberate improvement process, — i.e., knowledge-transfer. The first autonomous learning model was developed by Wright in 1936, who concluded that manufacturing labor costs in the airframe industry decrease as cumulative production increases. Arrow (1962) later considered a cumulative production experience model in which labor costs of production decrease as cumulative investments in production process increase. Baloff (1971), Joskow and Rozanski (1979), and Lieberman (1984) provided empirical support from a variety of different industries, — musical instruments, semiconductors, apparels, auto assembly, and nuclear power plants. Induced learning models, on the other hand, view learning about the production process as a deliberate improvement process rather than a solely autonomous process. For example, Mody (1989) considered cumulative production experience as a combination of continuous adjustment process and a team-work of skilled engineers rather than only a cost reduction process. And, he concluded that the level of learning is strongly affected by knowledge creation about production processes and how this knowledge is transferred within a production facility. Dorroh et al. (1994) stressed the importance of investing resources in attaining knowledge about production processes rather than depending simply on autonomous learning. Hatch and Mowery (1998) showed that learning in early stages of semiconductor manufacturing is a function of allocation of engineering resources rather than increasing production volumes. Adler (1990), who conducted a detailed field study of a global high-tech company, defined learning as a continuous improvement process through accumulation of knowledge about manufacturing process. Zangwill and Kantor (1998) later arrived at the same conclusion. Dutton and Thomas (1984) and Terwiesch and Bohn (2001) emphasized experimentation as a form of deliberate learning. They concluded that the learning rate should no longer be treated as a given constant, but rather as a dependent variable influenced by a firm's behavior.

Process improvement during full production also received important attention in the literature. Studies along this dimension can be classified further into two streams; improvement with process change and improvement through knowledge creation and sharing. Fine (1986), Li and Rajagopalan (1998), and Carrillo and Gaimon (2000), for example, provide a good literature review about how process change may benefit process improvement. The common theme in process change models is that process change provides process improvement through increasing effective production capacity in the long run. Characterizing how process change or investment in process change should be carried out given accumulated knowledge in the system is the typical objective sought in these models. In this paper we do not model process improvement through process change. Instead, we follow in the footsteps of Lapre and Wassenhove (2001), Argote and Ingram (2000), Szulanski (2000), Darr and Kurtzberg (2000), Adler (1990), Mody (1989), and Argote et al. (1990) and analyze knowledge creation and sharing among production facilities with *similar* production processes. Specifically, we study the impact of process commonality across production lines on knowledge creation and how that affects pilot production strategies. As Adler (1990) points out, increasing competition in industries with short-life cycles increases the importance of commonality of production lines at process design level. Commonality of production lines has two important advantages. First, process commonality provides identical products worldwide, — this is key to gradual discontinuance strategy of products with short-life cycles. Second, process commonality minimizes the risk of having operational problems propagated beyond managerial control due to divergence in process designs. Intel's copy exactly strategy, as explained in detail in McDonald (1998), further illustrates how common processes enable better knowledge creation and sharing across production lines. We model knowledge creation and sharing among production lines with similar production processes and show how process commonality affects pilot production strategies. In contrast to the extant empirical literature on knowledge creation and sharing, our model provides a descriptive analytical model that is supported by the existing empirical observations. Moreover, our model in section 5, to the best of our knowledge, provides the first analytical approach in characterizing when knowledge sharing should be desirable among production lines with common production processes. Finally, it provides an analytical benchmark for existing and future empirical models on when knowledge sharing should be carried out among production lines with process commonalities.

Production yields greatly impact the economic performance of production processes. Hence, there is a rich literature drawing attention on production yield management. As indicated by empirical findings of McIvor et al. (1997), production yield management is crucial for competitive industries with short-life cycles. Hence, pilot production becomes primarily a production yield improvement period in which firms strive to have as much process improvement as possible. Bohn and Terwiesch (1999) detail the economic impact of production yield management. They conclude that production yields are especially important in periods of constrained production capacity. Literature on production yield improvement can be divided into two main categories. First, the yield improvement is based on technical tools and methods. Detect classifications, statistical process control studies, in-line product inspections are some examples of methods cited in the literature along this direction. The interested reader is referred to Bohn and Terwiesch (1999) for additional references. Second, yield improvement is based on process and quality control. Some suggested approaches for better process and quality control in the literature are: inspection policies with better and quick information sharing among during the production process Tang (1991); providing job security to obtain employee commitment to process improvement Repenning (2000); and duplication of process equipment between development and manufacturing facilities in introducing new technologies Hatch and Mowery (1998). We model the production yield improvement through accumulated cumulative experience about the production process, i.e., the more production occurs, the higher production yields are. The most relevant yield model to ours in literature is developed in Terwiesch and Bohn (2001) where production yield is modelled to depend on production time and processing capability. Production yields improve over time in both models as production continues. However, our yield model does not consider changes in processing capability, but considers changes in production rate.

2.2 Pilot Production Model

The objective of this section is to derive optimal pilot production strategies for new product launches. Due to lack of production experience, the manufacturer initially faces low yields in producing new products. Hence, the manufacturer aims at gaining production experience through launching an experimental production on a single production line that currently produces an old product that has mature yield levels. Although the old product is still profitable, it will eventually become obsolete due to competition from new products in the market. The principal question for the manufacturer thus is whether he should allocate some limited capacity to experimental production to gain experience in producing the new product while still reaping profits from the mature product, or continue full production of the mature product and launch the full production of the new product when the former becomes obsolete.

To answer this principal question we develop an optimization problem that maximizes profits, — single line capacity allocation problem (SLCAP). In this problem we analyze two successive generations of a product on a single production line; an old product and a new product. We assume the following environment: an old product that currently has high and stable yields in production; and a new product that is ready to be launched into high-volume production. The main parameters that impact this analysis are time and price. Time considerations are: the time that the production life cycle of the old product comes to end is T_0; the earliest time of launching the production of the new product is 0; dedicated time to the new product before the launching of its successor is T. Price considerations are: the selling price of the old product before the new product is introduced is p_0^h; the selling price of the old product after the new product is introduced is p_0^l; and the selling price of the new product is p_1. We consider that the following generalizations are true for a new product relative to the old one. A new product is launched into production before its predecessor becomes obsolete, — at time $0 \leq T_0$. A new product performs better than its predecessor. Hence, a new product causes the price of the old one to fall, — $p_0^l < p_0^h$. The selling price of the new product does not decline much in its early life cycle, — the selling price of the new product over $[0, T]$ is constant at value p_1. Manufacturing costs in many industries are often constant, hence, it is possible to consider that unit production cost of both the predecessor product and its successor have the same value, — c \$/product. We assume that demand is sufficient enough to entail full production capacity, which is deterministic and denoted by K product/unit-time.

We model the production process in continuous time using known production rates that correspond to the amount of product produced per unit-time. We assume that the old product has matured yield levels by time $t = 0$, — i.e., 100% yield levels. As to the new product, the firm has low yields during its initial production phase, because of lack of experience. As the firm progresses through the production of the new product, its yield levels and production experience will increase. The model that is entailed assumes the following: the yield level of the new product at time t is derived from a yield function $y(Q(t)) \in [0, 1]$, — i.e., a function of the cumulative production quantity of the new product over $[0, t]$, $Q(t)$; $y(Q(t))$ is increasing and concave in $Q(t)$ with initial yield level $y(0)$; and the new product is produced at rate $x(t) \in [0, 1]$ product(s)/unit-time over $[0, T]$, as shown below in figure 1: We denote the capacity allocated to the new product at time $t \in [0, T]$ by $x(t) \in [0, K]$. Given a capacity allocation policy of $\mathbf{x} = \{x(t) \in [0, K],\ t \in [0, T]\}$ for the new product, the total profit obtained from both the old and new generations of the product in period $[0, T]$ can be stated as follows, — where it is assumed that any leftover product has

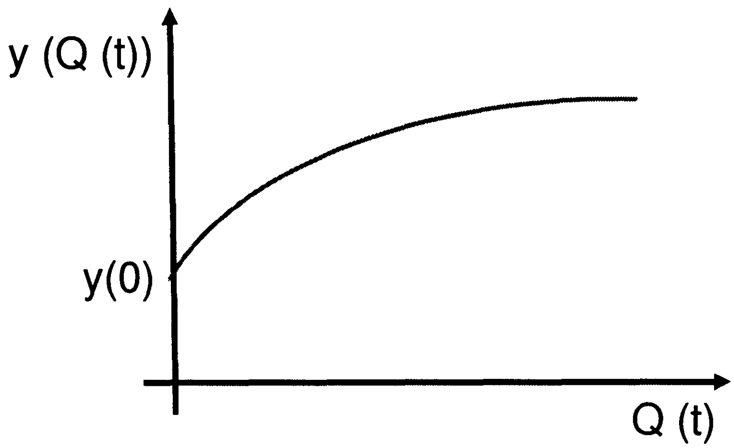


Figure 2-1: Production Yield Function, $y(Q(t))$

zero salvage value:

$$\Pi(\mathbf{x}) = \int_0^{T_0} [K - x(t)][p_0^l - c]dt + \int_0^{T_0} x(t)[y(Q(t))p_1 - c]dt + \int_{T_0}^{T} K[y(Q(t))p_1 - c]dt$$

This profit expression has three distinct parts: part I, — denotes profit from sales of the old product over $[0, T_0]$, where the old product is produced at rate $K - x(t)$ products/unit-time with yield level of 1; part II, — denotes profit from sales of the new product over $[0, T_0]$, where the new product is produced at rate $x(t)$ products/unit-time with yield function $y(Q(\cdot))$; and part III, — denotes profit from sales of the new product over the remaining time period $[T_0, T]$ in which the new product is produced at rate $x(t) = K$ product/unit-time with yield function $y(Q(\cdot))$. The maximization of total profit from both generations of the product, under limited production capacity entails the following inherent problem, — ***single line capacity allocation problem*** (hereinafter, "SLCAP"):

$$(SLCAP) \quad \text{Max}\Pi(\mathbf{x})$$
$$\text{s.t} \ \ 0 \leq x(t) \leq K \ \ \forall t \in [0, T]$$

A recap of all the parameters that impact the analysis in this section is as follows:

T_0	$=$ production end-time of the old product, where $0 \leq T_0$,
T	$=$ dedicated time for the new product, where $T_0 < T$,
$x(t)$	$=$ production rate of the new product at time t, $x(t) \in [0, K]\ \forall t \in [0, T]$,
$Q(t)$	$=$ cumulative production quantity of the new product during $[0, t]$,
$y(Q(t))$	$=$ production yield level of the new product at time t, $y(Q(t) \in [0, 1]$,
p_0^h	$=$ unit selling price of the old product before the launch of the new product,
p_0^l	$=$ unit selling price of the old product after the launch of the new product,
p_1	$=$ unit selling price of the new product during $[0, T]$,
c	$=$ unit production cost per product in period $[0, T]$,
$\Pi(\mathbf{x})$	$=$ total profit obtained in period $[0, T]$ given $\mathbf{x} = \{x(t) \in [0, K],\ t \in [0, T]\}$,

We would like to find out how fast we should learn about the production process of the new product during its pilot production. We could learn quickly and improve yields rapidly by allocating more production capacity for the new product. However, by allocating more production capacity for the pilot production an opportunity cost is incurred due to the reduction in the profitable output of the old product that has already mature yield levels. *SLCAP* stated above reflects this trade-off and its solution provides us with the optimal production capacity to be allocated for pilot production. We next characterize the solution for *SLCAP* through the following propositions:

Proposition 1 *The production of an old product stops permanently whenever a new product is introduced.*

Proposition 2 *Optimal capacity allocation policy for problem SLCAP is of bang-bang type; it is optimal to allocate the whole production capacity for a new product immediately if* $\frac{p_0^h}{p_1} \leq \frac{1}{T_0} \cdot \int_0^{T_0} y(Q(t))dt$, *otherwise it is optimal not to allocate any production capacity for a new product until the old product becomes obsolete.*

In summary, propositions 1 & 2 imply that it is never optimal to dedicate only limited capacity to the new product, — i.e., experimental production is not optimal, furthermore, it is optimal to launch full production of a new product either immediately or only when the mature product that is currently in production becomes obsolete. These results are robust to any production yield function that is concave-increasing in cumulative production quantity. Hence, these results are quite general and independent of any details on how production process is improved as long as production yield levels are increasing at a decreasing rate as more production occurs. Moreover, if production yield levels have random additive components due to process excursions or tool failures, then the results still hold true as long as the expected production yield levels are concave-increasing in cumulative production quantity. Also, by following similar line of proof optimal policy can be shown to be still bang-bang type even if selling prices and production cost were time dependent. Thus we do not loose generality by assuming constant selling prices and production cost in our analysis. We also need to emphasize that we ignored the issue of strategic competition in our analysis. Manufacturers may act strategically on how they allocate their total production capacity depending on their competitors' actions in the market. For this case, a different model is needed to determine the best capacity allocation between consecutive generations of products. In addition, we ignored any possible uncertainty in production capacity. Again, a different model is needed depending on how risk sensitive the manufacturer is. However, we conjecture that bang-bang type of policy would still hold true for any risk-neutral manufacturer. Moreover, we assumed that production yield level on each production line is independent of those on other production lines. We relax this assumption in the next section where we analyze how pilot production lines are managed when there is a dependency among production yield levels through knowledge transfer among production lines.

2.3 Pilot Production with Process Commonalities

In this section we extend our discussion to the case of multiple production lines with similar production processes. With multiple production lines with similar production processes experience gained on one production line could be transferred to other production lines as well. Thus, it is not clear any more if the simple result for the previous section will hold here, too. In particular, a manufacturer may now sacrifice some of the profit for the old product to gain experience on producing a new product early on. In other words, a manufacturer may launch the new product on some of its production lines to gain production experience while continuing the production of the old product on the remaining production lines, i.e., — pilot production process. Production experience gathered on a pilot production line, then, can be transferred to other production lines that have similar production processes as in pilot production line. The higher the commonality in production processes between pilot production line and any other production line implies the higher the amount of production experience transferred from pilot production line, hence the higher the benefit from pilot production. Thus, the principle question here for a manufacturer is how to manage pilot production for a new product when the manufacturer has multiple production lines with similar production processes.

We develop two different models to analyze this principal question. The first model analyzes how pilot production is managed when production yield levels are deterministic. This model, moreover, provides insights on the trade-off between the pace of pilot production and the cost of pilot production. The second model, on the other hand, analyzes how pilot production is managed when production yield levels have uncertainties. In the following, we describe and analyze each of these two models separately.

2.3.1 Pilot Production with Deterministic Yield Levels

We consider production yield levels that increase at a decreasing rate as the cumulative amount of production increases and do not involve any uncertainty. This implies that the rate of increase in yield levels is known to certainty and it is not affected by any unexpected excursions in production process and possible machine failures. We consider a manufacturer that has total of $n \in Z^+$ production lines. We argue that it is enough to assign at most one production line for pilot production when yield levels are deterministic because the same amount of production experience is accumulated no matter how many pilot production lines are used. Thus, we consider that at most one of the production lines is assigned as a pilot production line, which starts producing

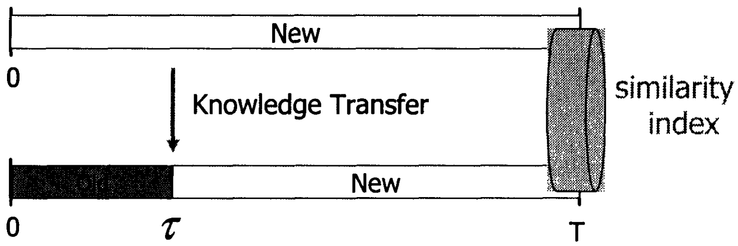


Figure 2-2: Knowledge Transfer Process with Similarity Index $\gamma \in [0, 1]$

the new product at time $t = 0$ while the other production lines launch the production of the new product by time $\tau \in [0, T_0]$. The production process of the new product on pilot production line is improved through cumulative production experience over $[0, \tau]$. At the end of pilot production period, knowledge about the production process is formed and transferred to all other production lines. All the parameters defined in section 3 hold true here. We model the similarity between the pilot production line and any of the other production lines through a similarity index $\gamma \in [0, 1]$ as shown in figure 2, where the high level of γ implies a high level of similarity between production processes. Furthermore, we consider that more knowledge can be transferred from pilot production line as the similarity index increases. Hence, we assume the following scheme for the similarity index: $\gamma = 1$, — implies that both production lines have identical production processes, hence, acquired knowledge on pilot production line is completely transferred; $\gamma = 0$, — implies that production processes have no commonality, as a result, no knowledge is transferred; and $\gamma \in (0, 1)$, — implies that production processes have some commonality, therefore, some portion of the acquired knowledge on pilot production line is transferred.

We also assume that the more knowledge transferred implies the higher production yield levels in production line acquiring knowledge. Hence, the production line acquiring knowledge from pilot production line begins to launch the new product with higher yields as the similarity index γ increases. In addition we assume that the more knowledge transferred implies the less effort in improving production yield levels in production line acquiring knowledge. Hence, production yield levels on production line acquiring knowledge improve at a slower rate as more knowledge is transferred,

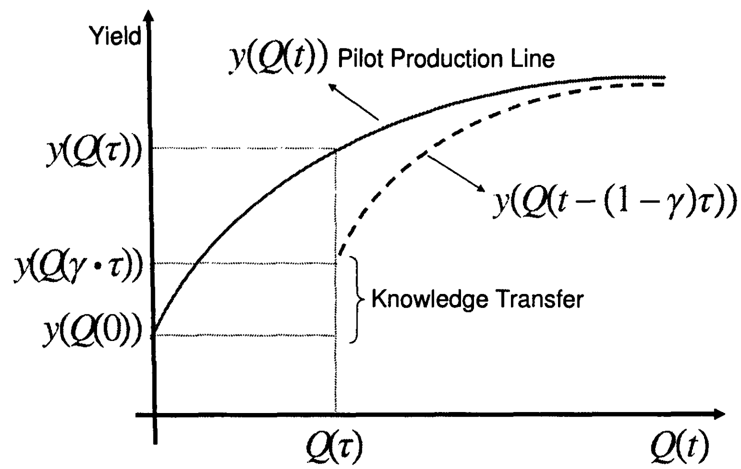


Figure 2-3: Production Yield Levels after Knowledge Transfer

in other words, when there is higher similarity between production lines. This boils down to the following production yield equations for the production line acquiring knowledge, and for the pilot production line transferring knowledge, respectively: $y(Q(t-(1-\gamma)\cdot\tau))$, $t \in [\tau, T]$; and $y(Q(t))$, $t \in [0, T]$. According to this model, production yield function on pilot production line is not affected by knowledge transfer, i.e., $y(Q(t))$, $t \in [0, T]$. However, production yield levels on any production line transferring knowledge from pilot production line is obtained by shifting the production yield function on pilot production line over time by $(1-\gamma)\cdot\tau$ units, i.e., $y(Q(t-(1-\gamma)\cdot\tau))$, $t \in [\tau, T]$. This representation guarantees the following for the production line acquiring knowledge: First, as the similarity index γ increases, its production yield function gets closer in shape to the production yield function on pilot production line; Second, as the similarity index γ decreases, its starting production yield levels decreases, but its production yield levels improve at a faster rate. These aforementioned features are depicted below in figure 3. We note that our choice of model for knowledge transfer from pilot production line is not necessarily unique. There might be many other ways to model it. Nevertheless, our knowledge transfer model captures the two common critical factors: the more similarity among production processes enables the more knowledge transfer from pilot production line; and the more knowledge acquisition from pilot production line implies the slower rate of improvement in production yields.

In the following we first analyze whether pilot production would become desirable in this setting. To simplify our presentation we assume that each production line has a production rate of $\frac{K}{n}$ units/unit-time. We denote the total profit from all n production lines by $\Pi(n, \tau(n))$:

$$\int_0^T \frac{K}{n}[p_1 y(\frac{K}{n}t)-c]dt+(n-1)[\int_0^{\tau(n)} \frac{K}{n}\left(p_0^l - c\right) dt+\int_{\tau(n)}^T \frac{K}{n}[y(\frac{K}{n}(t-(1-\gamma)\tau(n)))p_1-c]dt]$$

There are two distinct parts to this profit expression. They are: part I, — denotes the profit from the pilot production line over $[0, T]$; and part II, — denotes the profit from the remaining $(n-1)$ production lines over $[0, T]$.

According to our model, pilot production is pursued whenever optimal production launch time $\tau^*(n) > 0$. Otherwise, all production lines launch the full production of new product simultaneously at time $t = 0$ and no pilot production occurs. The following two propositions characterize when optimal production launch time $\tau^*(n)$ is non-zero for given $n \in Z^+$ production lines.

Proposition 3 *For any $n \geq 2$, $n \in Z^+$ and $\gamma \geq 0.5$ optimal production launch time of a new product $\tau^*(n)$ is non-zero iff $\frac{p_0^l}{p_1} > (1-\gamma) \cdot y(\frac{K}{n} \cdot T) + \gamma \cdot y(0)$.*

Above proposition implies that it becomes desirable to use pilot production when there are multiple similar production lines with deterministic yield levels if *i*) the old product is still profitable compared to the new product; *ii*) the similarity between the pilot production line and other production lines is high enough; *iii*) there are more production lines; and *iv*) the amount of dedicated time for the new product is high enough. More interestingly these results show that the simple result of section 3, which indicates that pilot production is not optimal, does not hold true in general when a manufacturer has multiple similar production lines and the ability to transfer production experience across them. This result holds true for any production yield function that is concave-increasing in cumulative production quantity. We need to point out, however, that the result strongly depends on how we model knowledge transfer process. Furthermore, total profit function $\Pi(n, \tau(n))$ can be shown to be

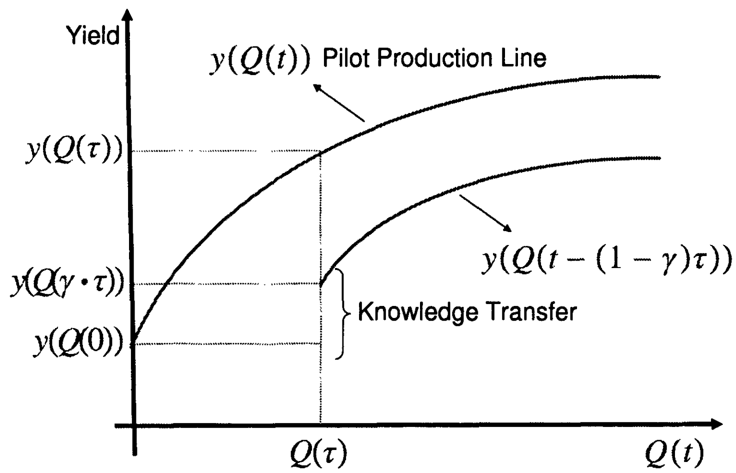


Figure 2-4: Production Yield Levels under Alternative Knowledge Transfer Model

strictly concave in duration of pilot production $\tau(n)$ for different knowledge transfer models as well. For example, we would alternatively model production yield function on production line acquiring knowledge by the following yield function, $y(Q(t)) + (1-\gamma) \cdot y(Q(\tau(n)))$, $t \in [\tau(n), T]$. This alternative yield model guarantees that production yield levels of the production line acquiring knowledge increases as more knowledge is transferred from pilot production line. This model, however, constraints production line transferring knowledge to have the same process improvement rate as on pilot production line as depicted in figure 4 below. Thus it becomes an appropriate model when production yield levels in both the pilot production line and the production line obtaining knowledge increase at the same rate regardless of when and how much knowledge is transferred. Under this alternative knowledge transfer model, $\Pi(n, \tau(n))$ still becomes concave in $\tau(n)$.

We next consider the question of how much production capacity should be allocated for pilot production. Sparing more production capacity for pilot production increases the amount of production experience gathered through pilot production, however, pilot production bears an opportunity cost due to producing the new product instead of the old product that has higher production yield levels. Hence, there exists a trade-off between the amount of experience obtained from pilot production and opportunity cost incurred from pilot production. Thus we analyze this trade-off to determine how much capacity should be allocated for pilot production. We allocate the total production rate K units/unit-time in the following scheme among $n \in Z^+$ production lines: allocate $\frac{K}{n}$ units/unit-time of production rate for pilot production line; and allocate the remaining capacity $\frac{K \cdot (n-1)}{n}$ units/unit-time equally among the remaining $(n-1)$ production lines. Equal capacity allocation on the remaining $(n-1)$ production lines simplifies our presentation. However, it does not limit our model because the essential trade-off does exist no matter how capacity is allocated among the remaining production lines.

According to the above capacity allocation scheme, increasing the number of production lines n refers to decreasing the amount of capacity allocated for pilot production; similarly, decreasing the number of production lines n refers to increasing the amount of capacity allocated for pilot production. For a given $n \in Z^+$, the total maximum profit from n production lines simply becomes $\Pi(n, \tau^*(n))$, where $\tau^*(n)$ stands for optimal production launch time as stated before. We would like to characterize how $\Pi(n, \tau^*(n))$ changes as the number of production lines $n \in Z^+$, in other words, the capacity allocated for pilot production, changes. Namely, we would like to determine whether it is better to allocate lower production capacity or higher production capacity to pilot production to maximize profits. We begin with the following proposition that checks whether using a pilot production line with smaller capacity would result in higher profits for the most simple case:

Proposition 4 *For* $\gamma \geq 0.5$, $\Pi(n=2, \tau^*(2)) > \Pi(n=1)$ *iff the price ratio* $\frac{p_0^l}{p_1} > \frac{2 \cdot \int_0^T y(K \cdot t)dt - \int_0^T y(\frac{K}{2} \cdot t)dt - \int_{T_0}^T y(\frac{K}{2} \cdot (t-(1-\gamma) \cdot T_0))dt}{T_0}$ *and optimal knowledge transfer time* $\tau^*(2) = T_0$.

The above proposition simply shows that allocating half of the production capacity for pilot production would bring higher profits than allocating the whole production capacity for pilot production as long as the price ratio between the old product and the new product exceeds certain threshold value. In other words, it shows that total profit may further increase as the production capacity allocated for pilot production is decreased if the old product is still profitable compared to the new product. This proposition triggers us to ponder whether it is better off to dedicate even smaller production capacity for pilot production when the old product is still profitable. To answer this question we next characterize how total maximum profit function $\Pi(n, \tau^*(n))$ varies with the allocated production capacity for pilot production.

Proposition 5 *If $\gamma \geq 0.5$ and $\Pi(n = 2, \tau^*(2)) > \Pi(n = 1)$ hold true, then for all $n \geq 2$, $n \in Z^+$ total profit $\Pi(n, \tau^*(n))$ can be divided into two separate parts; profit obtained from old product $\Pi_{Old}(n, \tau^*(n))$ and profit obtained from new product $\Pi_{New}(n, \tau^*(n))$ such that*

i) $\Pi_{Old}(n, \tau^*(n))$ is increasing and discrete-concave in n,

ii) $\Pi_{New}(n, \tau^*(n))$ is decreasing and discrete-convex in n.

The above proposition tells us that as we allocate less and less capacity for pilot production, profit from the old product increases at a decreasing rate while profit from the new product decreases at a decreasing rate under two sufficient conditions; first, similarity index γ is greater than or equal to 0.5; second, total profit with half of the total production capacity allocated for pilot production is greater than that with whole production capacity allocated for pilot production.

The first condition seems to limit the result in this proposition to pilot production lines with high similarity index, however, it is indeed necessary to have high similarity among production lines to benefit more from pilot production through knowledge transfer, thus analyzing pilot production lines with high similarity index is appropriate in our setting. Moreover, many manufacturers aim at having production lines with similar production processes to enable better knowledge and technology transfer among production lines. And, this supports our argument that pilot production gains more importance when there is high commonality and standardization among production lines. Intel Corp.'s copy exactly strategy is one extreme example in which the target is to have similarity index $\gamma = 1$ among its semiconductor fabs.

The second condition, on the other hand, simply restates the result in previous proposition. It is necessary to have $\Pi(n = 2, \tau^*(2)) > \Pi(n = 1)$, otherwise reduction in the production capacity for pilot production does not result in higher total profits. Thus, this condition does not limit us in analyzing how total maximum profit function

$\Pi(n, \tau^*(n))$ varies with the allocated production capacity for pilot production.

We showed in the above proposition that allocating less production capacity for pilot production results in deriving higher profits from the old product, but lower profits from the new product. And, the total profit is the sum of the profits from the old product and the new product. Hence, there exists a trade-off between the profits from the new product and the old product depending on how much production capacity is allocated for pilot production. The following proposition makes this trade-off more explicit.

Proposition 6 *If $\gamma \geq 0.5$ and $\Pi(n = 2, \tau^*(2)) > \Pi(n = 1)$ hold true, then*

i) total profit $\Pi(n, \tau^*(n))$ is unimodal in n, $n \in Z^+$,

ii) total profit $\Pi(n, \tau^*(n))$ is increasing for all $n < n^* \in Z^+$,

ii) total profit $\Pi(n, \tau^*(n))$ is decreasing for all $n > n^* \in Z^+$.

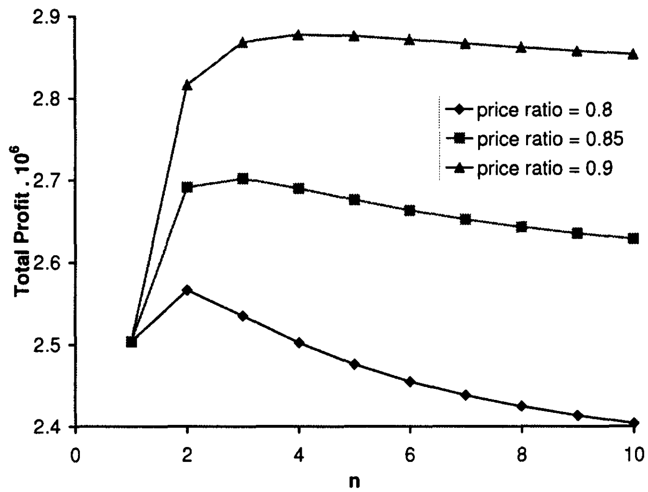
Above proposition, in summary, shows that there is a unique optimal production capacity to be allocated for pilot production. And, it can calculated through simple optimality search. This is helpful in determining how much production capacity managers should allocate for pilot production. Moreover, it clarifies the trade-off between lower opportunity cost of pilot production and slower learning through pilot production. It shows that this trade-off answers the question of how much production capacity should be allocated for pilot production. Namely, maximum profit obtained increases for a while as we reduce the amount of production capacity allocated for pilot production because the old product is still profitable. However, the effect of slower learning due to smaller production capacity outweighs the lower opportunity cost of pilot production as we further reduce the allocated capacity. Hence, there exists a unique optimal production capacity for pilot production when the old product is still profitable and there is a high similarity between production lines. We note that unimodality of total profit function in allocated production capacity for pilot production is robust to any production yield function that is concave-increasing in cumulative production quantity. And, it also holds true for the two knowledge transfer models that we described earlier. We next illustrate how total profit function behaves with production capacity of pilot production through a numerical example.

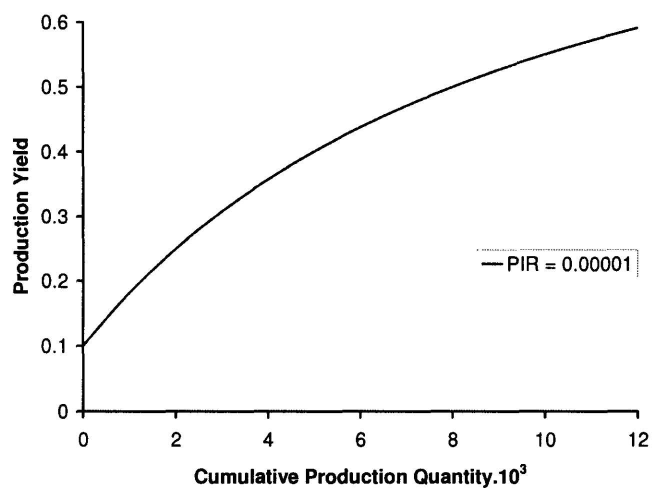
2.3.2 Example:

Suppose that we have the following production yield function that is concave and increasing in cumulative production quantity, $Q(t)$:

$$y(Q(t)) = y_0 + \frac{1 - y_0}{1 + \frac{1}{Q(t) \cdot PIR}}, Q(t) \in [0, \infty),\ \ y(Q(t)) \in [0, 1)$$

We represent the initial production yield level by y_0. It stands for the starting production yield level before any prior production experience except any possible knowledge transfer from other pilot production lines. We use an initial production yield of $y_0 = 0.1$, 10% production yield level, which is realistic for many production facilities including semiconductor fabs and chemical processing plants. Also, we represent how fast production experience improves production yield levels through variable PIR, process improvement rate. It refers to how quickly manufacturers ramp-up their production yield levels through production experience. We note that $\frac{dy(Q(t))}{d(PIR)} > 0$ and $\frac{d^2y(Q(t))}{d(PIR)^2} < 0$. This implies that production yield levels improve at a decreasing rate as process improvement rate increases. Moreover, process improvement rate is strongly affected by quality improvement efforts. Production lines or manufacturers would be at different stages of maturity in terms of quality management. And, this may greatly impact how they improve their production yield levels. Hence, we may even classify production lines or manufacturers into different categories depending on how sensitive their production yield levels to cumulative production experience. To illustrate this classification, we consider two extreme cases; $PIR = 0.00001$ and $PIR = 0.000004$. We choose the following parameters of interest: production time of interest for the new product $T = 12$ months; and production end time for the old product $T_0 = 5$ months; production rate $K = 10,000$ products/month; similarity index $\gamma = 0.8$; process improvement rate $PIR = 0.00001$; unit production cost $c = \$20$; and unit selling price for the new product $p_1 = \$100$. The following figures show how total profit function $\Pi(n, \tau^*(n))$ behaves with respect to change in $n \in Z^+$ for different unit selling prices, $p_0^l = \$80$, $\$85$ and $\$90$, where price ratio stands for $\frac{p_0^l}{p_1}$:





It is interesting to note that total profit function is not very sensitive to changes around the optimal n, i.e., $1\% \sim 5\%$. Some possible answers for this observation are: having deterministic production yield levels, having only a single pilot production line, and more importantly knowledge transfer happens at time T_0 for all cases in which $\Pi(2) > \Pi(1)$. We next consider the same example when process improvement rate $PIR = 0.000004$.

We observe from the above example that total profit becomes more sensitive as process improvement rate gets smaller. This is because opportunity cost of pilot production increases as production yield levels improve slowly due to smaller process improvement rates.

2.3.3 Pilot Production with Uncertain Yield Levels

Manufacturing processes on pilot production lines are often not mature, hence the rate of process improvement may vary across pilot production lines making the same product using the same technology. Sources of differences in process improvement rates can be classified in various ways. Some of them are operator adjustments, chemical contamination, ambient particles causing defects, process excursions, and tool failures. These sources might have both systematic and random effects on process improvement rates, which consequently result in random variations in production yield levels on pilot production lines. We observe that in many industries multiple production lines are often dedicated for pilot production to hedge against uncertainties in process improvement rates. Running multiple pilot production lines may provide the opportunity to tune up the manufacturing process through a comparative monitoring across pilot production lines so that yield levels on pilot production lines are balanced. The key question here that manufacturers face is how to manage pilot production lines that have uncertain process improvement rates, therefore, uncertain production yield levels.

The model here analyzes how pilot production is managed when production yield levels have uncertainties. We have already argued earlier that it is sufficient to have at most a single pilot production line when yield levels are known with certainty. This is because no additional production experience would be generated if we have multiple pilot production lines with deterministic yield levels. However, the same notion would not hold true when yield levels are not known with certainty in advance. Having multiple pilot production lines gives us the benefit of comparing yield levels among multiple lines and picking up the pilot production line with the highest yield level to transfer production experience to other production lines. However, this benefit has a trade-off because each additional pilot production line carries an opportunity cost. Here we develop a simple model to analyze this trade-off.

We consider a manufacturer that has total of $n \in Z^+$ production lines, each with production capacity of K units/unit-time. Out of these n production lines,

m of them are dedicated as pilot production lines and remaining $(n-m)$ production lines continue producing the old product. We assume that production yield on each pilot production line is random due to immaturity in production process, however, expected production yield on each pilot production line improves through cumulative production experience. We assume random production yield level of the form $Y(Q(t), \epsilon) = y(Q(t)) + \epsilon$, where ϵ represents the random change in production yield level due to variability in production process and $y(Q(t))$ is still defined as the same as earlier. $Y(Q(t), \epsilon)$ has a cumulative distribution function $F_Y(\cdot)$ with mean $\mu = y(Q(t)) + E[\epsilon]$ and variance $\sigma^2 = Var(\ \epsilon)$. We assume mean and variance values such that $Y(Q(t), \epsilon)$ is almost certainly non-negative. We note that we have modelled the randomness in production yield levels with an additive random term. Alternatively, we could model randomness in production yield levels with a multiplicative random term, i.e., $Y(Q(t), \epsilon) = y(Q(t)) \cdot (1 + \epsilon)$. Similar analysis can be carried out under this alternative model as well. We focus on the additive random term model purely for the clarity of presentation.

Pilot production lines start producing the new product at time $t = 0$ and expected production yield level for the new product on each pilot production line improves with cumulative production while the remaining production lines continue milking profits out from the old product. At the end of pilot production period $[0, \tau_m(n)], \tau_m(n) \in [0, T_0]$, actual realizations of production yield levels on pilot production lines are compared with each other and pilot production line with the highest yield realization is chosen to transfer knowledge to all other production lines. Knowledge transfer occurs at time $\tau_m(n)$ from the chosen pilot production line to all other production lines. Production of the old product is stopped and all production lines produce the new product over the remaining production period, $[\tau_m(n), T]$. We represent the production yield level in pilot production line $i \in \{1, 2, ..., m\}$ by the random variable $Y_i(Q(t), \epsilon_i) = y(Q(t)) + \epsilon_i$. Thus the highest random production yield level among all pilot production lines at time $\tau_m(n)$ will be $Y_h(Q(\tau_m(n)), \widehat{\epsilon}) = \max_{i \in \{1,2,...,m\}} \{Y_i(Q(\tau_m(n)), \epsilon_i)\}$, where $\widehat{\epsilon} = \{\epsilon_1, \epsilon_2, ..., \epsilon_m\}$ and we assume that yield levels $Y_i(Q(\tau_m(n)), \epsilon_i)$'s are mutually independent. Moreover, we assume that production yield levels for the new product are updated in the following scheme immediately after knowledge transfer occurs: production yield level on each pilot production line becomes $Y_h(Q(t), \widehat{\epsilon})$ for $t \in [\tau_m(n), T]$; whereas, production yield level on each of the remaining $(n-m)$ production lines becomes $Y_h(Q(t-(1-\gamma)\cdot\tau_m(n)), \widehat{\epsilon})$ for $t \in [\tau_m(n), T]$. According to this chosen scheme, knowledge gap between pilot production lines and the remaining production lines decreases as the similarity index γ increases; thus the production line acquiring knowledge from pilot production line begins to launch the new product with higher expected yield level. We note that this knowledge transfer scheme presumes that all pilot production lines have exactly the same production process, i.e., similarity index among pilot production lines is 1. This assumption is for the clarity of presenting knowledge transfer scheme. And, a similar analysis can be carried out if similarity index among pilot production lines is less than 1.

We represent the total expected profit from all n production lines by $E[\Pi(\tau_m(n), m)]$, which is the sum of expected profits from the old and new product over production period $[0, T]$ as shown below:

$$\begin{aligned} E[\Pi(\tau_m(n), m)] =& (n-m)\cdot[K\cdot(p_0^l - c)\cdot\tau_m(n) + \int_{\tau_m(n)}^{T} K\cdot[p_1\cdot E[Y_h(Q(t- \\ & (1-\gamma)\cdot\tau_m(n)), \widehat{\epsilon})] - c]dt] + m\cdot[\int_0^{\tau_m(n)} K\cdot[p_1\cdot E[Y(Q(t), \epsilon)] \\ & - c]dt + \int_{\tau_m(n)}^{T} K\cdot[p_1\cdot E[Y_h(Q(t), \widehat{\epsilon})] - c]dt] \end{aligned}$$

The above expression for total expected profit has four distinct parts. The first part is the expected profit over $[0, \tau_m(n)]$ from the old product on $(n-m)$ production lines. The second part is the expected profit over $[\tau_m(n), T]$ from the new product on $(n-m)$ production lines. The third part is the expected profit over $[0, \tau_m(n)]$ from the new product on m pilot production lines. And, the last part is the expected profit over $[\tau_m(n), T]$ from the new product on m production lines that are dedicated as pilot production lines over $[0, \tau_m(n)]$. In the following we characterize how total expected profit changes with respect to number of pilot production lines m.

Proposition 7 *If $E[Y_h(Q(t),\widehat{\epsilon})]$ is increasing and discrete-concave in $m \in Z^+$ and similarity index between pilot production lines and the remaining production lines $\gamma \geq 0.5$, then the total expected profit $E[\Pi(\tau_m(n),m)]$ is discrete-concave for all $m < n \in Z^+$.*

This proposition simply emphasizes the trade-off between the benefit from multiple pilot production lines and opportunity cost associated with pilot production lines. Simply put, it shows that it is better-off to have one more pilot production line if the gain in expected profit by dedicating one more line as a pilot production line is larger than the opportunity cost incurred in doing so. Furthermore, this result indicates that it would be better off to use multiple production lines when there are uncertainties in production yield levels. We illustrate the model here with the following numerical example:

2.3.4 Example:

We consider random production yield level of the form $Y(Q(t),\epsilon) = y(Q(t))+\epsilon$, where the random change in production yield level on each production line due to variability in production process, ϵ, is Uniformly distributed over $[-u,u], u \in (0,1]$. We assume identical distributions on random change for all production lines for clarity of presenting the example. Production yield levels for the new product in pilot production lines improve through cumulative production experience and have the following form over $t \in [0,\tau_m(n))$: $Y_i(Q(t),\epsilon_i) = y(Q(t)) + \epsilon_i$, where $\epsilon_i \sim$Uniform$[-u,u]$, $i \in \{1,2,...,m\}$. To derive an explicit expression for total expected profit, we first evaluate $E[Y_h(Q(\tau_m(n)),\widehat{\epsilon})] = E[\max_{i\in\{1,2,...,m\}}\{y(Q(t)+\epsilon_i\}]$. Suppose that $Y_h(Q(\tau_m(n)),\widehat{\epsilon})$ has a cumulative probability distribution $F_h(\cdot)$. We use the following relation between m random variables and the maximum among them in deriving $F_h(\cdot)$:

$$P(Y_h(Q(t),\widehat{\epsilon}) \leq y_h) = P(\bigcap_{i\in\{1,2,...,m\}} y(Q(t)) + \epsilon_i \leq y_h) = \prod_{i=1}^{m} \frac{z - [y(Q(t)) + u]}{2 \cdot u}.$$

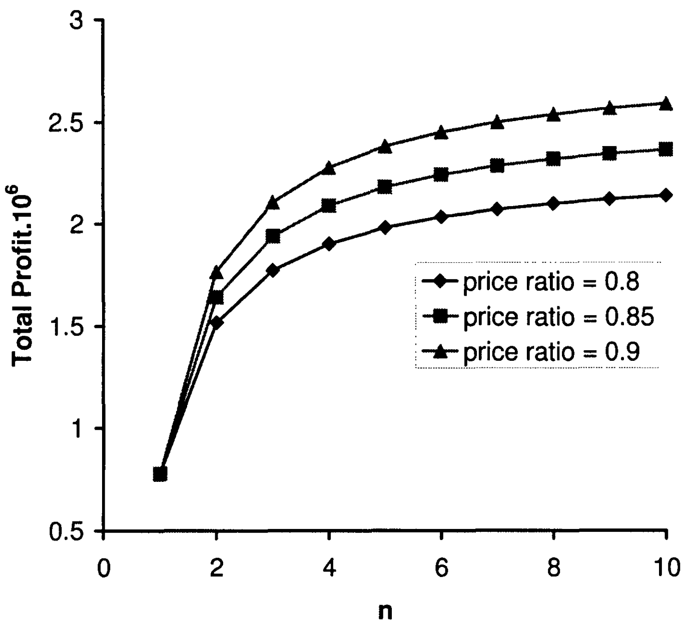
Taking the expectation of $Y_h(Q(t),\hat{\epsilon})$ yields that $E[Y_h(Q(t),\hat{\epsilon})] = \int_{-\infty}^{\infty} z \cdot dF_h(z)dz = \int_{y(Q(t))-u}^{y(Q(t))+u} z \cdot [\frac{m\cdot(z-(y(Q(t))-u))^{m-1}}{(2\cdot u)^m}]dz = y(Q(t)) + \frac{(m-1)\cdot u}{m+1}$. Similarly, we can derive $E[Y_h(Q(t-(1-\gamma)\cdot\tau_m(n)),\hat{\epsilon})] = y(Q(t-(1-\gamma)\cdot\tau_m(n))) + \frac{(m-1)\cdot u}{m+1}$. Total expected profit then will have the following expression:

$$
\begin{aligned}
E[\Pi(\tau_m(n),m)] =&(n-m)\cdot[K\cdot(p_0^l-c)\cdot\tau_m(n) + \int_{\tau_m(n)}^{T} K\cdot[p_1\cdot y(Q(t-(1-\gamma)\cdot \\
&\tau_m(n))) - c]dt] + m\cdot\int_0^T K\cdot[p_1\cdot y(Q(t)) - c]dt + \\
&\frac{m-1}{m+1}\cdot u\cdot n\cdot(T-\tau_m(n))\cdot K\cdot p_1
\end{aligned}
$$

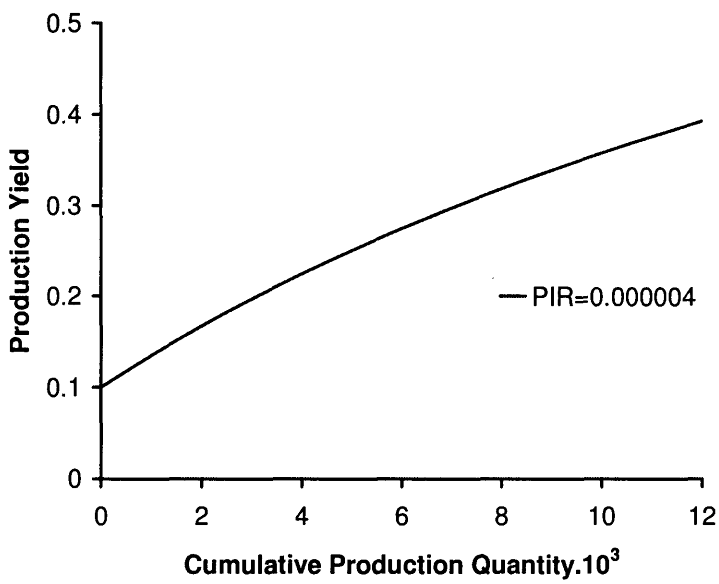
We can easily show that $E[Y_h(Q(t),\hat{\epsilon})] = y(Q(t)) + \frac{(m-1)\cdot u}{m+1}$ is increasing and discrete-concave in $m \in Z^+$. Therefore, using the previous proposition this implies that $E[\Pi(\tau_m(n),m)]$ is discrete-concave in $m \in Z^+$. Thus we can uniquely determine how many of production lines assign as pilot production lines. We use the following function for $y(Q(t)) \in [0,1)$, $y(Q(t)) = y_0 + \frac{1-y_0}{1+\frac{1}{Q(t)\cdot PIR}}$, the same as in previous example. And, we choose the following parameters of interest, which are also the same as in previous example: production time of interest for the new product $T = 12$ months; production end time for the old product $T_0 = 5$ months; production rate $K = 10,000$ products/month; similarity index $\gamma = 0.8$; total number of production lines $n = 20$; Uniform distribution over $[-u = -0.1, u = 0.1]$; initial yield level $y_0 = 0.1$; process improvement rate $PIR = 0.00001$; unit production cost $c = \$20$; and unit selling price for the new product $p_1 = \$100$. Figure 5 below shows how total expected profit function $E[\Pi(\tau_m(n),m)]$ behaves with respect to change in $m \in Z^+$ for different unit selling prices, $p_0^l = \$60$, \$70 and \$80, where price ratio stands for $\frac{p_0^l}{p_1}$:

Figure 5 depicts that decreasing price ratio $\frac{p_0^l}{p_1}$ favors increasing the number of pilot production lines as expected. It becomes favorable to initiate the production of the new product in more pilot production lines as the profit margin of the new product increases compared to that of the old product.

In this section we extended our discussion in section 3 to the case of multiple similar production lines. We modelled that with multiple production lines with similar production processes experience gained on one production line could be transferred to other production lines as well. And, we showed that the simple result for the previous section would not hold here if there is enough similarity of production processes among production lines to enable knowledge transfer among them. In particular, we showed that a manufacturer may now sacrifice some of the profit for the old product to gain experience on producing a new product early on. In other words, a manufacturer may now launch the new product on some of its production lines to gain production experience while continuing the production of the old product on the remaining production lines, i.e., – pilot production process. And, production experience gathered on a pilot production line, then, can be transferred to other production lines that have similar production processes as in pilot production line. We indicated that the principle question for a manufacturer with multiple production lines with similar production processes is how to manage pilot production for a new product. We analyzed this principle question in two different environments; pilot production with deterministic production yield levels and pilot production lines with uncertain yield levels. For each scenario, we derived conditions on when pilot production becomes desirable and furthermore characterized how total profits change with the allocated capacity for pilot production, which are essential in designing pilot production strategies upon existence of multiple production lines. In the next section, however, we focus on how production processes of a new product can be improved during its full production. Although production process of a new product is improved through pilot production, the quality of the production process is still being monitored during full production for further improvements. And, in the next section, we analyze trade-offs involved in this continuous improvement process.



3
2.5
2
1.5
1
0.5
Total Profit.10^6
0
2
4
6
8
10
n
price ratio = 0.8
price ratio = 0.85
price ratio = 0.9



0.5
0.4
0.3
0.2
0.1
0
Production Yield
0
2
4
6
8
10
12
Cumulative Production Quantity.10^3
PIR=0.000004

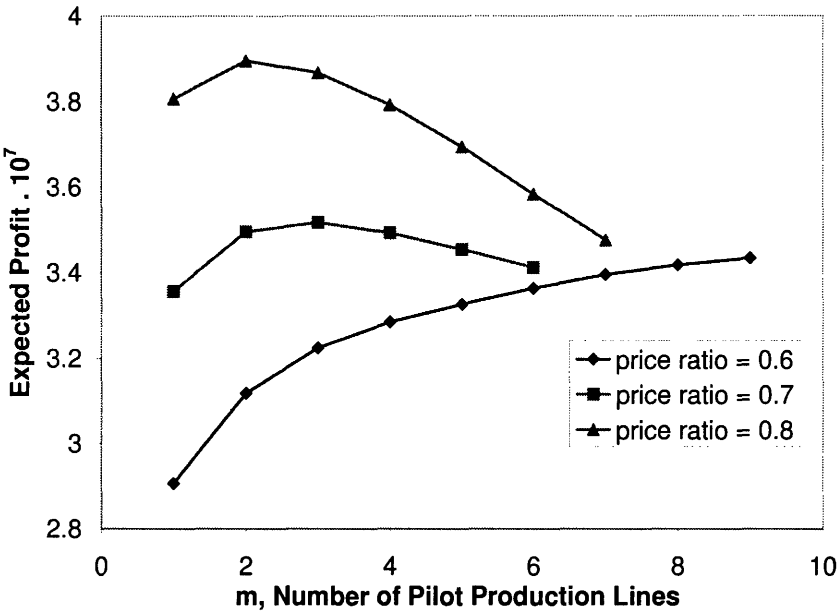


Figure 2-5: Expected Profit vs Number of Production Lines

Chapter 3

Process Improvement during Full Production

3.1 Introduction

Although production process of a new product is improved through pilot production when there are multiple similar production lines, the quality of the production process is still being monitored during full production for further improvements. The objective of this chapter is to characterize how production processes of a new product can be improved during its full production. We consider the case of full production where a lead production line scans information on process improvements on a continuous basis with the remaining production lines. However, this knowledge sharing process involves costs that need to be traded-off with benefits from process improvements. The principle question for a manufacturer is thus when to transfer knowledge from a lead production line about process improvements for a new product.

To answer this question we form a yield control mechanism for a production line so that its production process can benefit from that of a lead production line. This yield control mechanism is based on yield differences which stand out as an opportunity to improve production yields through knowledge transfer. We characterize when knowledge transfer from a lead production line should be desirable.

3.2 General Framework

Generally, production lines during full production fall into the following two categories. First, lead production lines, — generally associated with high yield levels, and denoted here as Production-line 1. Second, dominated production lines, — generally associated with low yield levels, and denoted here as Production-line 2. The yield level at time t in Production-line j is given by $y_j(Q(t))$, $j \in \{1,2\}$, $t \in [0,T]$. First set of assumptions are: production lines 1 and 2 will launch full production of a new product at time $t = 0$; production lines 1 and 2 will have the same initial yield levels; and the yield levels of both production lines will increase through cumulative production experience over time.

Production-line 2 is subordinate to Production-line 1, hence, its learning curve will accelerate more slowly and its yield levels will also be lower. The mathematical expression is given by $y_1(Q(t)) \geqslant y_2(Q(t))$ for all $t \in [0,T]$. Hence, the reason for production line 1 to transfer knowledge to production line 2. The second set of assumptions, which are functions of yield levels are: when knowledge transfer happens at time $\tau \in [0,T]$, the yield function of Production-line 1 is not affected, however, the yield function of Production-line 2 is increased by $y_1(Q(\tau)) - y_2(Q(\tau))$, - i.e., the new yield function in Production-line 2 becomes $y_2(Q(t)) + y_1(Q(\tau)) - y_2(Q(\tau))$, $t \in (\tau,T]$; T is finite - $T < \infty$; $y_1(Q(t)) - y_2(Q(t))$ is a continuous real-valued concave non-decreasing function, - yield difference between the two production lines, increases at a decreasing rate when there is no knowledge transfer between them; there are $N \geq 0$, $N \in \mathbf{Z}$ knowledge-transfer events during time period $[0,T]$; for each knowledge-transfer event there is an associated fixed cost F; the vector of times at which knowledge-transfer events occur are $\tau = (\tau_1, \tau_2, ..., \tau_N)$, such that $0 \leq \tau_i < \tau_{i+1} \leq T$, for $i \in \{|1, N-1|\}$; yield control policy (N,τ) is as such that knowledge-transfer events occur at times τ_i, $i \in \{|1,N|\}$ and the result of knowledge transfer is an increase in $y_2(Q(\tau_i))$ by $y_1(Q(\tau_i)) - y_2(Q(\tau_i))$ at time τ_i. To this extent, the yield control problem (YCP) is given by:

$$(YCP) \quad \text{Max} \quad -N \cdot F + \int_0^T [p_1 \cdot y_1(Q(t)) - \mathbf{c}]dt + \int_0^T [p_1 \cdot y_2(Q(t), N, \tau) - \mathbf{c}]dt$$

$$s.t \qquad N \geq 0, N \in \mathbf{Z} \text{ and } 0 \leq \tau_i < \tau_{i+1} \leq T,\ i \in \{|1, N-1|\}.$$

where $y_2(Q(t), N, \tau)$ denotes that yield level in Production-line 2 depends on yield control policy (N, τ) and p_1 and c refer to unit selling price and unit cost, respectively.

Proposition 8 *Given $\Delta Y(t) = y_1(Q(t)) - y_2(Q(t), N, \tau) \geqslant 0, t \in [0, T]$ under control policy (N, τ) and given $h = p_1$. Then, (YCP) is equivalent to the following minimization problem:*

$$(YCP/E) \quad Min \ \ N \cdot F + \int_0^T h \cdot \Delta Y(t)dt$$

$$s.t \qquad N \geq 0, N \in \mathbf{Z} \text{ and } 0 \leq \tau_i < \tau_{i+1} \leq T,\ i \in \{|1, N-1|\}.$$

According to the above proposition, a positive yield difference between the two production lines indicates that there is an opportunity cost for not producing at the same yield levels. A constant $h = p_1 > 0$ represents the incurred opportunity cost of having one unit of yield difference per unit time. Under linear cost structure, the incurred opportunity cost is at the rate $H(\Delta y) = h \cdot \Delta y$ when the yield difference at time t is Δy.

Given that, there are incurred opportunity costs for having non-zero yield discrepancy between the production lines, and a fixed cost F for transferring knowledge, the objective of the problem (YCP/E) is to find a knowledge transfer policy (N, τ) that minimizes the sum of opportunity costs that is due to yield differences between the two production lines, and the cost of transferring knowledge over a finite horizon $[0, T]$ starting with a zero initial yield difference level, – i.e., $\Delta Y(0) = 0$.

Proposition 9 *Under control policy (N, τ), where $N \geq 0, N \in \mathbf{Z}$ and $\tau = (\tau_1, \tau_2, ..., \tau_N)$ such that $0 \leq \tau_i < \tau_{i+1} \leq T$, for $i \in \{|1, N-1|\}$, if $\Delta Y(t) = y_1(Q(t)) - y_2(Q(t), N, \tau)$ is real-valued continuous function on $[0, T]$ such that $\frac{d\Delta Y(t)}{dt} \geqslant 0$ and $\frac{d^2\Delta Y(t)}{dt^2} < 0$ for*

$t \in [0, T]$, then the objective function in (YCP/E) can be replaced by the following expression, $C(N, \tau)$:

$$C(N,\tau) = N \cdot F + \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot [\Delta Y(t) - \Delta Y(\tau_i)]dt.$$

Assuming that, there are N meetings between the production lines over the production period $[0, T]$, solution to (YCP/E) provides optimal times to carry out N knowledge-transfer events. Since the objective function in (YCP/E) is a continuous real-valued function on a set $G = \{0 \leq \tau_i < \tau_{i+1} \leq T,\ i \in [|1, N-1|],$ $N \geq 0, N \in \mathbf{Z}, T \in [0, \infty)\}$ that is compact and non-empty, by Weierstrass' theorem G contains a vector $\{\tau_1, \tau_2, ..., \tau_N\}$ that minimizes the objective function in (YCP/E). The first order optimality conditions for (YCP/E) provides the following set of difference equations with boundary conditions $\tau_0 = 0$ and $\tau_{N+1} = T$:

$$[y_1(\tau_i) - y_2(\tau_i)] - [y_1(\tau_{i-1}) - y_2(\tau_{i-1})] = \frac{d}{d\tau_i}(y_1(\tau_i) - y_2(\tau_i)) \cdot (\tau_{i+1} - \tau_i),\ i \in [|1, N|$$

The above set of difference equations can be reduced to a single equation with a single variable τ_1, – time of the first knowledge-transfer event, which can be easily solved to optimality. Once the optimal τ_1 is found, the rest of $N-1$ knowledge transfer times, τ_i $i \in [|2, N|]$, can be recursively calculated. The obtained solution(s) can be easily tested with the second order optimality conditions to determine the global minimum for (YCP/E). Once the optimal vector of meeting times τ^* is obtained, the objective function can be rewritten with the optimal knowledge transfer times. And, the resulting objective function can be optimized with the decision variable $N \geq 0, N \in \mathbf{Z}$ to determine the optimal number of knowledge-transfer events. The following proposition provides upper bound on the optimal number of knowledge-transfer events.

Proposition 10 *Optimal number of knowledge-transfer events $N \geq 0, N \in \mathbf{Z}$ in*

(YCP/E) *is bounded by the following expression:*

$$N \leqslant N_{\max} = \frac{\int_0^T h \cdot \Delta Y(t)dt}{F}$$

The next issue is to determine the conditions that guarantee a unique solution for (YCP/E). The following proposition provides the necessary conditions when yield difference functions $y_1(Q(t)) - y_2(Q(t))$ are piecewise-linear-concave.

Proposition 11 *Suppose that:*

i) $y(t) = y_1(Q(t)) - y_2(Q(t))$ *is a piecewise-linear-concave function, that is piecewise continuous and bounded over* $[0, T]$*; and*

ii) $\{f_1, f_2, ..., f_m\}$ *is the set of distinct components of* $y(t)$ *such that* $y(t) = \max\{f_1, f_2, ..., f_m\}$ *and* $0 \leq f'_{i+1} \leq f'_i, i \in [|1, N-1|]$.

If $\frac{f'_1}{f'_m} \leq 1/\cos(\frac{\pi}{N_{\max}+1})^2$, *then there exists a unique solution for* (YCP/E).

3.3 Example

The following example illustrates the production yield control model. Consider a linear yield difference function $y(t) = y_1(Q(t)) - y_2(Q(t))$, $t \in [0, T]$ and $N \geq 0, N \in \mathbf{Z}$ number of knowledge-transfer events between the two production lines. According to Proposition 8 and 9, the objective function in (YCP/E) becomes:

$$C(N, \tau) = N \cdot F + \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot [\Delta Y(t) - \Delta Y(\tau_i)]dt = N \cdot F + y'(t) \cdot h \cdot \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} [t - \tau_i]dt$$

Taking the partial derivatives with respect to N decision variables $\{\tau_1, \tau_2, ..., \tau_N\}$ and equating them to zero give the following first order optimality conditions:

$$\frac{\partial C(N, \tau)}{\partial \tau_i} = y'(t) \cdot h \cdot (2\tau_i - \tau_{i-1} - \tau_{i+1}) = 0, \; i \in \{|1, N|\}, \; \tau_0 = 0, \; \tau_{N+1} = T$$

Let $H = (h_{i,j})$ denote the corresponding Hessian matrix of the objective function $C(N, \tau)$. The non-zero elements of H are $h_{i,i} = 2 \cdot y'(\tau_i)$, $h_{i,i-1} = -y'(\tau_{i-1})$, and

$h_{i,i+1} = -y'(\tau_i)$, $i \in \{|1, N|\}$, $h_{1,0} = h_{N,N+1} = 0$. The linear function $y(t)$ satisfies the conditions stated in Proposition 7. Hence, the Hessian matrix $H = (h_{i,j})$ is positive definite, and the solution to first-order optimality conditions results in the global minimum for (YCP/E). The solution to the set of indifference equations that resulted from the first-order optimality conditions provides the following optimal knowledge transfer times $\tau_i^* = i \cdot \frac{T}{N}$, $i \in \{|1, N|\}$. This implies that each knowledge-transfer event is successively carried out $\frac{T}{N}$ units of time after the preceding knowledge-transfer event. The optimal number of knowledge-transfer events N^* can also be determined by optimizing $C(N, \tau^*)$ over $N \geq 0, N \in \mathbf{Z}$. $C(N, \tau^*) = N \cdot F + \frac{y'(t) \cdot h \cdot T^2}{2 \cdot N}$ can be shown to be strictly convex in N. Hence, the optimal number of knowledge-transfer events is the greatest non-negative integer N such that $N \leqslant T \cdot \sqrt{\frac{y'(t) \cdot h}{2 \cdot F}}$.

Chapter 4

Inventory Management with Multiple Demand Classes

4.1 Introduction & Literature Review

In this chapter we consider a single-product inventory system that serves multiple demand classes, which differ in their backlog costs or service level requirements. We develop a model for cost evaluation and optimization, under the assumptions of Poisson demand, deterministic replenishment lead-time, and a continuous–review (Q, R) policy with rationing. We show the value from a rationing policy. We begin our study with an extensive literature review to lay down what has been done in the inventory management literature with multiple demand classes.

The literature on inventory management with multiple demand classes can be divided into two categories depending on review type: periodic-review models and continuous-review models. The first studies have focused on periodic review models, where inventory levels are observed periodically at pre-specified time points. Further studies later on shifted the focus to continuous review models where the inventory levels are continuously observed over time. The initial stream of studies in inventory literature on multiple demand classes has begun with analyzing periodic-review models. Veinott (1965) carried out the first study to analyze a periodic-review inventory model with several demand classes for a single product. He analyzed a multi-period model where random demands from m different classes occur in each period. He assumed that demand classes are differentiated based on either their occurrence time during a period or relative importance to fill them. Moreover, he assumed that there is an inventory replenishment opportunity with zero replenishment lead-time at the beginning of each period. He divided each period into several subintervals. He applied the following sequence of events in each subinterval: fill any backlogged demand starting with backlogs from highest priority demand class from on hand inventory, observe the demand from all classes at the beginning of the subinterval, fill demands starting from highest priority demand class from on hand inventory, backlog any unfilled demand. He assigned a different backlogging cost for each demand class. He showed that it is optimal to replenish inventory with a base stock policy. He further pointed out that rather than using a policy of filling demands from a demand class as much as possible before filling demands from the next highest demand class, critical inventory levels can be used to ration the on hand inventory among demand classes. He depicted this policy in a special case with two demand classes: high priority and low priority demand class. Under the proposed "critical level policy," demands from either class are satisfied as they occur until the inventory level depletes to a specified level, – critical level –, then only high priority demands are filled from on hand inventory while low priority demands are backlogged. Veinott (1965) introduced this inventory rationing policy, however, he did not analyze it in his paper.

The proposed critical level policy has been analyzed subsequently by Topkis (1968). Similar to Veinott (1965), he analyzed a periodic review inventory model with multiple stochastic demand classes for a single product. He first considered a single period inventory model with several stochastic demand classes that are differentiated based on different backlogging costs. He divided the single period into subintervals. At the beginning of each subinterval, demand for that interval from all classes is observed and a decision is made on how much outstanding demand from each demand class is to satisfied. Any unsatisfied demand is either outsourced or backlogged to the next subinterval. He showed that the optimal inventory allocation policy within each subinterval is characterized by a $n \times k$ dimensional matrix

$\{z_t^j : t = 1, 2, ..., k; j = 1, 2, ..., n\}$ such that in subinterval t one satisfies as much class j demand as possible as long as the inventory on hand is not below z_t^j and there is no backlogged demand from any demand class that has higher priority than class j. The main limitation of his model is that there is only a single replenishment opportunity, which is at the beginning of the period. He next extends his model to one with multiple replenishment periods where there is an opportunity to replenish inventory at the beginning of each period. He assumes that any backlogged demand at the end of each period is cleared by an immediate order so that no backlogged demand is carried to the next period. The later assumption enables him to conclude that the optimal rationing policy in each period is as described in his previous single period model. Moreover, he shows that the optimal rationing policy in each period is myopic if unmet demand in each subinterval is fully backlogged to the next subinterval. And, a base stock policy is optimal for ordering in each period. This extended model is not general enough to allow demands to be backordered between periods. However, it is the first model in the literature that analyzed the critical level policy suggested in Veinott (1965) and it inspired further studies in inventory management with multiple demand classes.

Similar periodic review models to Topkis (1968) are studied by Kaplan (1969) and Evans (1968). Different from Topkis (1968) these two studies focused on only two demand classes. Evans (1968) essentially analyzed the same multi period model in Topkis (1968) when there are only two demand classes assuming any unfilled demand is lost. He showed the same result as in Topkis (1968) that the optimal ordering policy is a base stock policy and the optimal rationing policy is of critical level policy with a single critical inventory level so that only demands from high demand class are filled when the inventory level drops to the critical level. Kaplan (1969) also analyzed the same multi period model in Topkis (1968) when there are only two demand classes. However, contrary to Evans (1968), Kaplan (1969) confined his model to only a critical level policy and assumed that any unfilled demand is fully backlogged. He showed the existence of optimal critical levels and provided an algorithm to find them.

After Kaplan (1969)'s work, periodic-review models with multiple demand classes have not received much attention for a while. Recently, Katircioglu and Atkins (1996) analyzed a periodic-review inventory system with multiple stochastic demand classes and fixed positive replenishment lead-time. In contrast to the previous literature on periodic-review models with multiple demand classes, they allowed unfilled demand to be fully backlogged to the next period and considered positive replenishment lead-times. Moreover, they required an associated service level for each demand class, which has also not been analyzed in previous literature. They first showed that there is no one-to-one mapping between a model with service levels and a model with back-ordering costs when backordering costs are linear in the amount backlogged. Hence, they chose to use quadratic backordering costs as an alternative. They formulated two problems: one with backordering costs in the objective function, "the cost problem" and one with service level constraints, "the service problem." In the cost problem they allowed negative inventory allocations for demand classes in each period and showed that a myopic heuristic allocation policy that considers only the current period is optimal. In the service problem, they developed a heuristic in which they considered controlling expected backorders only for some future period. They used the myopic heuristic allocation policy obtained in the cost problem for the service problem. They tested the effectiveness of these heuristics using simulation and observed that heuristics provide service levels close to the target service levels. However, this model has a limitation that it allows negative inventory allocations that are hard to explain and implement.

Another current study on periodic-review models is carried out by Frank, Zhang, and Duenyas (1999). They analyzed a periodic-review inventory model with two demand classes: a deterministic demand class and a stochastic demand class. They assumed that the deterministic demand must be filled immediately in each period while unfilled stochastic demands during the period is lost. Demands are observed at the beginning of each period and a decision is made on how much to order and how much demand from the stochastic demand class to fill. They assume that there is a fixed ordering cost and the replenishment lead-time is zero. For a finite horizon discounted model, they showed that the optimal ordering and rationing policies are state dependent and do not have simple structures. Hence, they proposed a simple heuristic (s, k, S) policy such that k determines the number of periods for which existing inventory is used to fill deterministic demand before ordering. This study departs from previous literature in terms of how rationing is applied because the objective of rationing is chosen to avoid incurring high fixed ordering costs rather than saving inventory for high priority demand.

Continuous review models in the literature can be divided into two categories: models with (s, Q) ordering policies and models with base stock ordering policies. The first continuous-review model was studied by Nahmias and Demmy (1981). They considered a continuous-review inventory policy with two stochastic demand classes: a high priority and a low priority demand class. They assumed that inventory is replenished according to a $(\mathbf{s}, Q)$ policy and inventory is rationed according to critical level policy that has been suggested by Veinott (1965). They assumed a fixed critical level such that when the inventory level drops to this level, all low priority demands are backordered while high priority demands are continued to be filled. In a different approach from previous literature, they focused specifically on determining expressions for expected backorders under a fixed critical level policy rather than deriving optimal ordering and rationing policies. They first considered a single period model with the assumption that demand from both classes occur simultaneously at the end of the period. They derived expressions for expected backorders, which may require numerical methods to compute. They elaborated on how this single period model can be extended to a multi-period model with zero replenishment lead-time. Next, they considered $(\mathbf{s}, Q)$ ordering policy with independent Poisson demand processes for both demand classes and fixed positive replenishment lead-times. They simplified their analysis by assuming that at most one order is outstanding at any time. This assumption implies that when the reorder point is reached and a replenishment order is triggered, the inventory level and inventory position become identical. This allows them to calculate expected backorders for both demand classes. However, the expressions simply become approximations due to the assumption that no more than one order is allowed to be outstanding. Their model can be simply extended to compound

Poisson demand processes as done by Moon and Kang (1998), who derived approximate expressions for fill rates for both demand classes. Moreover, they extended the single period model of Nahmias and Demmy (1981) to multiple demand classes with multiple critical levels. They assumed that all demands occur simultaneously at the end of the period and obtained expected backorder expressions for each demand class. However, they did not consider extending the $(\mathbf{s}, Q)$ model of Nahmias and Demmy (1981) with more than two demand classes. Deshpande et al. (2003) analyzed the same $(\mathbf{s}, Q)$ model with two demand classes as in Nahmias and Demmy (1981). Similar to Nahmias and Demmy (1981), they assumed a stationary critical level to ration inventory. However, their objective was to optimize ordering and rationing policy parameters rather than developing service level expression for each demand class. Unlike Nahmias and Demmy (1981) they did not constrain themselves to at most a single outstanding order, therefore, their model is more realistic than the one in Nahmias and Demmy (1981). Allowing more than one order to be outstanding poses a problem of characterizing how backorders are filled by replenishment orders. To handle this problem, they developed an interesting "threshold clearing mechanism" to fill backorders. The main idea of threshold clearing mechanism is to assume that the on hand inventory level a replenishment lead-time back equals to $s + Q$ and to assume that the critical level policy is followed subsequently. In other words, exact analysis through keeping track of every replenishment order is difficult, hence, this mechanism assumes that virtual replenishment orders have occurred a lead-time back to raise the inventory level on hand to $s + Q$. Assuming this mechanism to fill backorders enables deriving expressions for expected number of backorders for both classes. Based on these expressions, they developed algorithms to calculate the optimal ordering and rationing parameters. To test the effectiveness of their model, they compared their results numerically with a better backlog clearing mechanism, "priority clearing mechanism," where high priority backorders are filled before low priority backorders. And, they recommended using a hybrid policy in which the optimal ordering and rationing parameters are determined by the model with "threshold clearing mechanism" and backordered demands are filled through "priority clearing mechanism." The main contribution of this work is to provide an interesting heuristic approach to handle the problem of filling backorders. Optimal ordering and rationing policies for an $(\mathbf{s}, Q)$ inventory system even with only two demand classes is still unknown.

Melchiors et al. (2000) also analyzed $(\mathbf{s}, Q)$ inventory model with two demand classes. Similar to the above studies, they used a stationary critical level policy to ration the inventory among the two demand classes. Unlike Nahmias and Demmy (1981) and Deshpande et al. (2003), they considered a lost sales environment so that demands from the low priority class are rejected whenever inventory level drops to the critical level. Similar to Nahmias and Demmy (1981), they assumed that at most one order is outstanding , demand process from each demand class is Poisson, and there is a fixed positive replenishment lead-time. In a lost sales environment, assuming $s < Q$ implies that at most one order is outstanding. Hence, the analysis in this work is exact whereas the analysis in Nahmias and Demmy (1981) results in an approximation because of the allowed backorders. They carried out a similar analysis as in Nahmias and Demmy (1981) and derived an exact expression for the expected cost. And, they presented an optimization procedure based on enumeration and bounding. They, however, did not generalize their model to more than two demand classes. Melchiors (2001), on the other hand, extended the model in Melchiors et al. (2000) to multiple Poisson demand classes with stochastic replenishment lead-times. He considered two types of critical level policy to ration the inventory among demand classes: a stationary critical policy where critical levels are constant, and a non-stationary critical level policy where critical levels are allowed to depend on the elapsed time since the outstanding order is triggered. We note that a time dependent critical level policy has been first used by Topkis (1968) in a periodic-review environment and has not been analyzed since then. Hence, Melchiors (2001) is the first study considering time remembering critical level policy in a continuous review environment. He formulated a Markov Decision model where the decisions are allowed to depend on the inventory level and the time elapsed since the replenishment order has been placed if the inventory level is below the reordering inventory level. He showed that the optimal policy is a non-stationary critical level policy when replenishment lead times are deterministic. Moreover, he showed that the critical levels decrease over the replenishment period. He carried out a numerical study to compare the simple stationary critical policy with the time dependent critical level policy. He observed that the difference in optimal cost between two policies is higher than 2 percent for only a few cases in all his examples. Hence, he suggested to use a stationary critical level policy because of its simplicity to apply. The main contribution of this work is that it proved that the optimal rationing policy is a non-stationary critical level policy in an $(\mathbf{s}, Q)$ inventory model with multiple Poisson demand classes, fixed replenishment lead-time, and lost-sales environment. Moreover, this work has an important contribution by considering a non-stationary critical level policy that provides an important benchmark to evaluate the stationary critical level policy employed by Nahmias and Demmy (1981), Melchiors et al. (2000), and Deshpande et al. (2003). There is no work in the literature, however, analyzing a non-stationary critical level policy for an $(\mathbf{s}, Q)$ inventory model allowing backorders, even with only two demand classes.

In contrast to the $(\mathbf{s}, Q)$ inventory models, continuous-review models with base stock ordering policies have not gotten much attention in literature. Dekker et al. (1998) provided the first study for an inventory model with two demand classes and one-for-one replenishment policy. They assumed that the demand process for both classes is Poisson and there is a positive fixed replenishment lead-time. They also allowed unfilled demands to be backordered. They furthermore assumed that a stationary critical level policy is used to ration inventory between the two demand classes. The model is similar to the one in Nahmias and Demmy (1981) except that a one-for-one replenishment policy is applied instead of a $(\mathbf{s}, Q)$ policy. Hence, it is possible that more than one order is outstanding, which is not the case in Nahmias and Demmy (1981). Allowing more than one order to be outstanding again creates difficulties here in filling backorders, which has been analyzed in Deshpande et al. (2003) for an $(\mathbf{s}, Q)$ inventory model with backorders. Namely, when there is a backorder for high priority demand, upon arrival of a replenishment order, it is optimal to first use the replenishment order to fill these backorders for high priority demand. Inventory level refers to on hand inventory minus all the existing backlogs while inventory position refers to inventory level plus all the existing orders that are on replenishment process. According to their model, when the inventory level is at least at the critical level, it is optimal to use the replenishment order to fill any backorders for low priority demand before replenishing the inventory. However, when the inventory level is below the critical level upon arrival of a replenishment order and there are backorders for low priority demand, we have to make a decision whether to use the incoming order to fill backorders for low priority demand or to replenish the inventory instead. To deal with this dilemma, they considered three methods to handle incoming replenishment orders. In the first method, replenishment order is first used to satisfy backorders for high-priority demand before filling backorders for low priority demand; and if there are no backorders, inventory is replenished. The second method is first to satisfy backorders for high priority demand, then replenish the inventory until a critical level is reached before filling backorders for low priority demand. These two methods do not require keeping track of what demand class triggered the arriving replenishment order. The third method differentiates the incoming order depending on what demand class issued it. If the incoming order is triggered by a high priority demand, then replenishing the inventory to critical level has priority over filling backorders for low priority demand. If the incoming order is triggered by a low priority demand, then filling backorders for low priority demand has priority over increasing inventory level to the critical level. The last method results in lower service for high priority demand class, but it decreases the length of a stock-out for low priority demand class. Stock-out probabilities are numerically evaluated for the three different methods to handle incoming backorders. They observed that the way incoming replenishment orders are handled have a significant influence on service levels. All three methods suggested result in approximate expressions for the service levels and Dekker et al. (1998) emphasizes the importance of how incoming replenishment orders are handled. Optimal method to handle incoming replenishment orders even for two demand classes is an open question in literature and no attempt in the literature has been made so far to determine it.

Dekker et al. (2000) extends the model in Dekker et al. (1998) to multiple demand classes with stochastic replenishment leadtimes, but switching to a lost sales environment rather than allowing backorders. They assumed one-for-one replenishment policy and stationary critical level policy to ration inventory among demand classes. Since they assumed a lost sales environment, handling incoming replenishment orders does not become a difficult task; each incoming replenishment order simply replenishes the inventory. They utilized the previously derived results in inventory literature about base-stock policies with lost sales environment and stochastic replenishment leadtimes to characterize state dependent demand arrival rates. By using these state dependent demand rates, they derived the expressions for the service level for each demand class and for the expected cost of the system. They also developed numerical solution methods to efficiently calculate the optimal base stock level and critical levels with or without service level constraints. We note that optimizing policy parameters subject to service levels has previously been carried out by Katircioglu and Atkins (1996) for a periodic-review model.

All studies mentioned above in the literature assumed an uncapacitated exogenous supply system and did not analyze how capacity affects inventory ordering and rationing decisions upon existence of multiple demand classes. Ha (1997a), on the other hand, considered a make-to-order production system with a single production facility and multiple demand classes for the end product. He assumed that any demand that is not satisfied immediately from on hand inventory is lost. Moreover, he assumed that production time of a single unit is exponentially distributed and demand process from each demand class is Poisson. He considered Markovian policies such that control actions are taken only considering the current state of the system. Moreover, assumptions on exponentially distributed production times and Poisson demand process makes the system memoryless, which enables him to conclude that a Markovian policy is optimal. Also, there are no backorders in this model, hence, it makes the replenishment process easier, i.e. all produced products increase the inventory level. He models a dynamic production control and demand acceptance model such that the production facility at any time chooses to produce or stop production and the production facility may either satisfy an arriving demand or reject it. He considered an optimal control policy that minimizes the expected discounted system cost over an infinite horizon. He showed that the optimal production policy is a stationary base stock production policy. Furthermore, he showed that a stationary critical level policy is optimal. The main contribution of this work is that it lays out sufficient conditions that lead critical level policy to be optimal in a capacitated production system. However, the main limitation of the model is that only exponentially distributed production times are considered, which may not be realistic for many applications. Ha (1997b) extended the study in Ha (1997a) further by allowing backorders to occur. However, he only considered two demand classes, which limits its application. Similar to Ha (1997a), he assumed that there is a single production facility that produces a single type of product with exponential production times. He also assumed that there are only two Poisson demand classes with different backordering costs. He developed a Markov Decision model to determine the optimal production policy and inventory rationing policy among all Markovian policies. Since the system is memoryless, a Markovian policy will indeed be the optimal policy. At any time, three possible actions can be taken in the system: do not produce, produce either to increase on hand inventory by one unit or to decrease one unit of backorder either from high priority or from low priority demand class. Similar to Ha (1997a), he considered an optimal control policy that minimizes the expected discounted system cost over an infinite horizon. He showed that there exists an optimal switching curve $S(y)$, where y stands for the number of backorders from low priority demand class, that determines both the optimal production and optimal rationing decisions. He showed that the optimal control policy has the following form: production for filling high priority backorders has the highest priority; when there are only low priority backorders, it is optimal to produce to increase on hand inventory level if the inventory level is below critical level $S(y)$ and to produce to fill low priority backorders otherwise; when there is no backordered demand, it is optimal to produce to increase inventory on hand if the inventory level is below $S(0)$ and to stop production otherwise; it is optimal to fill low priority demand from on hand inventory as far as the inventory level is above $S(y+1)$ and to backorder it otherwise. Moreover, he showed that a switching curve $S(y)$ is decreasing in y, the number of backorders for low priority class. This implies reserving less inventory for high priority demand class when the number of backorders from low priority class increases. Unlike the results in Ha (1997a) where stationary base stock and critical levels are optimal, both optimal critical levels to ration inventory and optimal base stock level depend on the size of the backlogged demand from low priority demand class. Limitations of this model are two-fold. First, the optimal policy holds true only for two demand classes and it is not certain it will hold true for more than two demand classes. Second, the form of the optimal policy is not known if production times are not Markovian. Finding either optimal or good heuristic production and rationing policies for make-to-stock production systems with more than two demand classes and backordering is still an open question and there has not been a study carried out in the literature to analyze it.

4.2 General Framework

Consider a single facility that carries inventory for a single type of product to serve $N \in Z^+$ different types of customer classes. A customer class is characterized by a group of demands that have either identical service level requirements or impose identical penalty costs if they are not filled immediately from inventory. Holding inventory has a financial burden for the facility. This financial impact increases further when demands are not known with certainty. Uncertain demands poses an important challenge for the facility on how much inventory to hold either to satisfy the pre-specified service level requirements or to incur minimum amount of penalty costs for not serving demands on time. Therefore, the facility seeks to find an inventory management policy that will enable it either to achieve pre-specified service level requirements with minimum possible inventory holding costs or to incur minimum amount of inventory holding costs and minimum amount of penalty costs for any orders that are delayed. We analyze how to manage the inventory at this single facility under the following assumptions:

- inventory is replenished according to a continuous-review (Q, R) replenishment policy;

- there is a fixed positive replenishment lead-time $L > 0$;

- the demand classes are ranked according to their priority to the facility: class-1 has the highest priority and class-N has the lowest priority, as shown in Figure 4-1;

- the demand from class-$i, D_i, i \in [|1, N|]$ follows a stationary Poisson process with rate $\lambda_i, i \in [|1, N|]$, independent of demand processes from other demand classes;

- any demand not satisfied immediately from on-hand inventory is fully backlogged until it is filled;

- inventory holding costs are incurred at rate $h > 0$ per unit per unit time;

- the associated backlogging cost for demand class-i is $b_i > 0, i \in [|1, N|]$, per unit-backordered per unit-time;

- inventory is rationed among the N demand classes by using a critical level policy, $\mathbf{c} = \{c_0, c_1, ..., c_{N-1} \mid c_i \in Z^+ \cup \{0\}$ and $c_{i-1} \leq c_i,\ i \in \{|1, N-1|\}\}$; and

- incoming replenishment orders are allocated on a first-come-first-served basis either to fill existing backorders or to increase the on hand inventory level

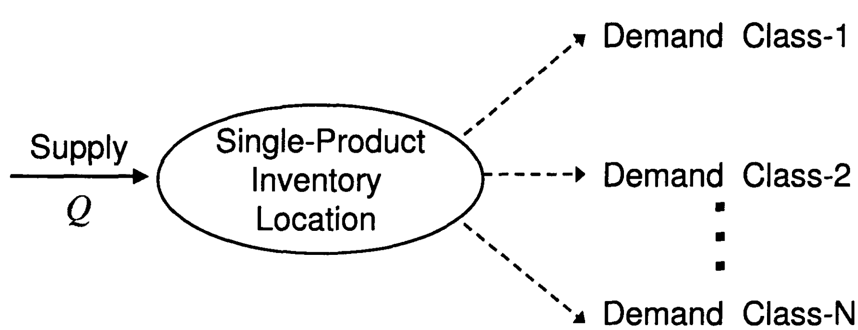


Figure 4-1: Single-Product Inventory Problem with Multiple Demand Classes

Under continuous-review (Q, R) replenishment policy, inventory position, which is the sum of the inventory level and any outstanding replenishment orders, is observed continuously. A replenishment order of size Q is placed with an outside supplier whenever the inventory position drops to level R. The outside supplier delivers replenishment requests after a fixed positive replenishment lead-time L. The inventory is rationed according to the critical level policy among the N demand classes. Under the critical level policy, for each demand class there is a critical stock level at and below which demand from that class is not satisfied from stock on hand. In this way inventory is reserved to serve demand from higher priority demand classes. The critical level for demand from class i is c_{i-1}, $i \in \{|1, N|\}$. If the physical inventory is at or below this level then the demand from class i is backordered. Whenever a replenishment request of size Q is received, the incoming replenishment quantity is allocated on a first-come-first-served basis between filling existing backorders and increasing the physical on-hand inventory.

The critical level c_1 refers to the reserved inventory for demand class 1 because any demand from other demand classes is backordered once the on-hand inventory level reaches this level. Similarly, $c_2 - c_1$ refers to the reserved inventory for both demand class 1 and demand class 2 because only demands from class 1 and class 2 are served once the on-hand inventory level reaches the critical level c_2; and exactly $c_2 - c_1$ units of inventory is used in common to serve demands from demand class 1 and demand class 2. In general, c_1 refers to the reserved inventory for demand class 1 and $c_i - c_{i-1}$ refers to the reserved inventory for demand classes 1 through i, $i \in [|1, N-1|]$. Let us denote the reserved on hand inventory levels by $s_1 = c_1$ and $s_i = c_i - c_{i-1}$, $i \in [|1, N-1|]$.

Our next step is to map our inventory problem into an equivalent inventory system that provides a more transparent analysis. This equivalent inventory system assumes the following:

- there are N different stockage shelves within the single facility;

- each shelf is designated to hold inventory for a different demand class;

- stockage shelves are numbered such that demands from demand class i is served from the inventory at shelf i and the inventory at shelf i is replenished from the inventory at shelf $i+1$, $i \in [|1, N|]$, where shelf $N+1$ refers to the outside supplier. The inventory at shelf i, therefore, faces two types of demand: regular demands from demand class i and inventory replenishment requests for shelf $i-1$;

- inventory at shelf N is replenished with a continuous-review (Q, R) replenishment policy from an outside supplier with $R = s_N$ while the inventory at shelf i is replenished from shelf $i+1$ with a one-for-one base stock policy with base stock level s_i, $i \in [|1, N-1|]$

- The inventory replenishment leadtime for shelf N is $L > 0$ time-units while the inventory of shelf i is immediately replenished from shelf $i+1$, $i \in [|1, N-1|]$;

- any unfilled demand or replenishment request at shelf i is backordered at shelf i until it is filled; and

- existing backorders at each shelf are filled on a FCFS (first-come-first-served) basis as soon as a replenishment order arrives there.

The above described inventory system can be considered a special serial inventory system with N installations, where there is a locational demand at each installation as shown in Figure 4-2. Each of N different demand classes is assigned to a differ-

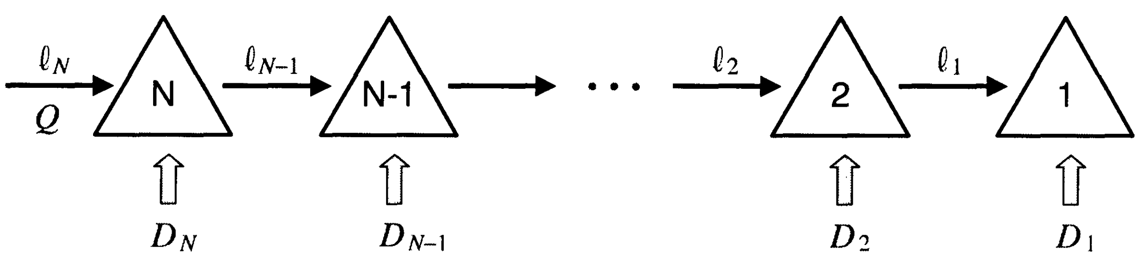


Figure 4-2: Serial Inventory System with Demand at Each Installation

ent installation. The most upstream installation corresponds to the lowest priority demand class, — i.e., demand class-N —, and the most downstream installation corresponds to the highest priority demand class,— i.e., demand class-1. The remaining demand classes are assigned to the remaining installations in the order of increasing priority towards the downstream installation. Each installation replenishes its inventory from its immediate upstream installation, where installation N replenishes from an outside supplier. The replenishment leadtime for installation i is given by $\ell_i = 0$ time-units, $i \in [|1, N-1|]$, which represents the required time to replenish inventory from the immediate upstream installation when there is enough inventory available there, while the replenishment leadtime at installation N is $\ell_N = L$ time-units. Any demand that is not filled immediately at each location by inventory on-hand is fully backordered until enough inventory becomes available.

The inventory at installation $i \in [|1, N-1|]$ is replenished with a one-for-one base stock policy. Whenever there is one unit of demand at installation $i \in [|1, N-1|]$, the inventory level at installation i is reduced by one unit and installation i places one unit of replenishment request to the immediate upstream installation $i+1$. When one unit of demand occurs at installation i, this demand is immediately satisfied if there is available inventory on-hand at installation i; otherwise, it is backordered at installation i until enough inventory becomes available. Also, upon arrival of a demand at installation $i \in [|1, N-1|]$, a replenishment request is placed to the most immediate upstream installation, installation $i+1$. The replenishment request for installation $i \in [|1, N-1|]$ is filled immediately if there is available inventory on-hand at installation $i+1$. Otherwise, it is delayed until sufficient inventory becomes available at installation $i+1$.

The inventory at installation N, on the other hand, is replenished from an outside supplier with a continuous-review (Q, R) replenishment policy. The inventory level at installation N is observed continuously and a replenishment order of size Q is placed to an outside supplier whenever the inventory position at installation N drops to an inventory ordering level R. Installation N obtains incoming replenishment orders of size Q that arrive after L time-units since they were placed.

The demand at installation $i \in [|1, N|]$ consists of the regular demands from demand class i and replenishment requests from installation $i-1$. Hence, backorders at installation i are of two types: backorders created by the regular demands from demand class i and backorders created by the replenishment requests from the immediate downstream installation. Upon arrival of a replenishment order to installation $i \in [|1, N|]$, any outstanding backorders at installation i are randomly filled. Clearing backorders randomly at each installation is clearly not optimal. However, it is also not that far from optimal. The reason is that the way the backorders are filled at each installation only matters when the incoming replenishment order Q is not enough to fill all outstanding backorders at installation N, and the probability of running into this case is very low. Using the inventory control mechanism described above, we can characterize the inventory dynamics in the serial inventory system, which enables us to derive explicit expressions for inventory levels and backorder levels at each installation. These expressions further help us characterize the required performance measures in managing the inventories at each installation. We use the following notation to analyze the inventory dynamics in the serial inventory system, where $i \in [|1, N|]$:

$$
\begin{aligned}
IL_i(t) &= \text{inventory level at time } t \text{ at installation } i, \\
IP_i(t) &= \text{inventory position (inventory level + inventory on order) at time} \\
&\quad\ t \text{ at installation } i, \\
B_i(t) &= \text{number of backorders at time } t \text{ at installation } i, \\
B_{i,i-1}(t) &= \text{number of backorders at time } t \text{ at installation } i \text{ that are due to} \\
&\quad\ \text{demand from installation } i-1, \\
B_{i,i}(t) &= \text{number of backorders at time } t \text{ at installation } i \text{ that are due to} \\
&\quad\ \text{demand at installation } i, \\
D_i(t, t+\ell_i) &= \text{demand during time interval } (t, t+\ell_i] \text{ from demand class } i,
\end{aligned}
$$

We can characterize how inventory level at each installation evolves over time using the following inventory dynamics equations for the serial inventory system, where $[\cdot]^+$ denotes the positive part of the inside expression, i.e, $[\cdot]^+ = \max([\cdot], 0)$, and

$i \in [|1, N|]$:

$$IL_i(t+\ell_i) = IP_i(t) - [D_i(t,t+\ell_i) + D_{i-1}(t,t+\ell_i) + \cdots + D_1(t,t+\ell_i)] - B_{i+1,i}(t) \quad (4.1)$$

$$B_i(t) = [-IL_i(t)]^+ \quad (4.2)$$

$$B_i(t) = B_{i,i}(t) + B_{i,i-1}(t), \quad (4.3)$$

$$B_{N+1,N}(t) = 0 \quad (4.4)$$

The result in (4.1) has been developed by Graves (1985). The intuitive explanation for this provided by Graves (1985) is as follows. At time t the outstanding orders for installation i are either in-transit to installation i or are backordered at the immediate upstream installation $i+1$. All items that were in-transit at time t will arrive at installation i by time $t+\ell_i$. But, none of the backorders at installation $i+1$ at time t can arrive at installation i by time $t+\ell_i$ because of the shipment time ℓ_i from installation $i+1$ to installation i. Furthermore, any demand during the time interval $(t, t+\ell_i]$ diminishes inventory level and any outstanding order created by the demand during the time interval $(t, t+\ell_i]$ can not be filled by time $t+\ell_i$. Therefore, at time $t+\ell_i$, the backorders at installation $i+1$ from time t, $B_{i+1,i}(t)$, and the demand during the time interval $(t, t+\ell_i]$, $\sum_{j=1}^{i} D_j(t, t+\ell_i)$, must be outstanding. Hence, the inventory level at time $t+\ell_i$ at installation i simply equals to the inventory position at time t at installation i minus the sum of $B_{i+1,i}(t)$ and $\sum_{j=1}^{i} D_j(t, t+\ell_i)$.

The equations (4.2), (4.3), and (4.4) on the other hand, tell us that backorder level at installation i is the negative part of the inventory level; backorders at installation i consists of backorders created by exogenous demand at installation i and backorders created by replenishment requests from the immediate downstream installation; and the outside supplier has enough capacity not to have any backorders, respectively.

The replenishment leadtime for the most upstream installation is $\ell_N = L > 0$, the same as the replenishment leadtime for the inventory facility, whereas the replenishment leadtime for all other installations are zero, $\ell_i = 0$, $i \in [|1, N-1|]$. Furthermore, inventory at any installation $i \in [1, N-1]$ is replenished according to a continuous review one-for-one replenishment policy with base-stock level s_i. Hence, inventory position at installation i simply equals the base-stock level s_i. Replacing these quantities into equations (4.1)- (4.4) provides the following set of equations in steady-state:

$$
\begin{aligned}
IL_N &= IP_N - \sum_{i=1}^{N} D_i^L, && (4.5)\\
IL_i &= s_i - B_{i+1,i}, \quad \text{for } i = 1, ..., N-1 && (4.6)\\
B_i &= [-IL_i]^+ && (4.7)\\
B_i &= B_{i,i} + B_{i,i-1} && (4.8)\\
B_{N+1,N} &= 0 && (4.9)\\
B_{1,0} &= 0 && (4.10)
\end{aligned}
$$

The steady state distribution of inventory level at each installation can be determined using the following sequential approach, starting from the most upstream installation N and moving downstream installations.

- Step 1: *Set $i = N$. Determine the steady-state distribution of IL_N, the inventory level at installation N.*

  For $i = N$, we use the equation (4.5). The inventory at installation N is replenished through a continuous review (Q, R) policy with $R = s_N$. Hence, the inventory position in steady-state at installation N, IP_N, has a continuous-time Markov chain and has a unique stationary limiting distribution, i.e. the inventory position at installation N in steady-state is Uniformly distributed over $[R+1, R+Q]$ (p.193 Zipkin 2003). Moreover, the distribution of $\sum_{i=1}^{N} D_i(t, t+L)$ is known. Hence, the steady-state distribution of IL_N can be easily derived through convolution.

- Step 2: *Derive the steady-state distribution of $B_i = [-IL_i]^+$, backorders at installation i.*

- Step 3: *Determine the steady-state distribution of $B_{i,i-1}$*

Backorders at each installation are filled randomly. Hence, the likelihood that a backordered demand at installation $i \in [|1, N|]$ is from demand class-i is proportional to the demand rate of demand class-i. Given a backorder level B_i at installation i, the conditional distribution of $B_{i,i-1}$ has a Binomial distribution. Hence, the steady-state distribution of $B_{i,i-1}$ can be derived by conditioning on B_i.

- Step 4: *Derive the steady-state distribution of* IL_{i-1}.

  We already derived the distribution of $B_{i,i-1}$ in the previous step. Hence, we can determine the distribution of IL_{i-1} from (4.6).

- Step 5: *Set* $i := i - 1$. *Move to the next downstream installation and go to Step* 2.

Continuing in this sequential fashion starting from the most upstream installation eventually provides us the steady- state distributions of inventory levels at all installations in the serial inventory system. Furthermore, knowing the steady-state distributions of inventory levels at all installations enables us to derive the required performance measures in managing the inventory at each installation, i.e. expected on-hand inventory at each installation, expected backorders in each installation, and desired service level at each installation.

We next analyze three different problems for cost evaluation and optimization using the aforementioned mapping to the serial inventory system. For the clarity of our presentation, our analysis for these three problems is based on $N = 3$ demand classes, nevertheless, our analysis holds true for any number of demand classes. We begin with describing the *Service Level Problem.*

4.3 Service Level Problem

The service level problem considers an associated service level for each demand class that measures what percentage of demand from each demand class is filled on time. Assigning a service level to a demand class rather than a penalty cost for not filling orders on time is of great interest for managers because specifying service levels is easier than determining how much it would cost if customer demands are not filled on time. The objective here is to find out how much inventory to hold for each demand class so that the total expected inventory holding costs are minimized while satisfying pre-specified service levels associated with demand classes.

We consider three demand classes, where class-1 refers to the highest priority demand class with the highest service level requirement and class-3 refers to the lowest priority demand class with the lowest service level requirement. We map the service level problem into an equivalent serial inventory system with three installations in total as described earlier and shown in Figure 4-3. Hence, total expected inventory

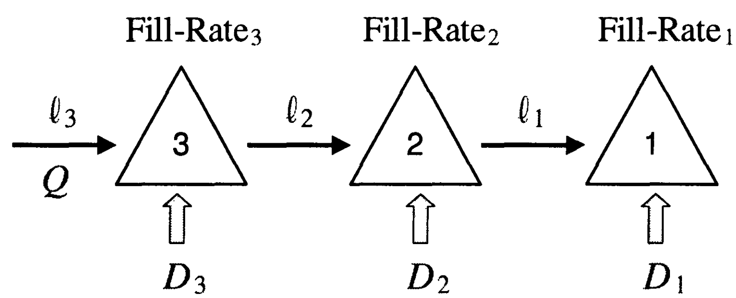


Figure 4-3: 3 Demand Classes with Fill-Rate Requirements

holding costs are simply (h, inventory holding cost rate)$\cdot \sum_{i=1}^{3}$(Expected on-hand inventory at installation i). We define the service level requirement for class-i, $i \in \{1, 2, 3\}$ as the long-run percentage of class-i demands that are filled immediately from on-hand inventory. Since arrivals of class-i demands are Poisson distributed, the service level requirement for class-i simply becomes equivalent to the long-run probability that on-hand inventory at installation $i \in \{1, 2, 3\}$ is greater than zero, i.e. $P(IL_i > 0)$.

Installation 3 manages its inventory with a continuous-review (Q, R) policy with $R = s_3$. Hence, for a given value of inventory ordering quantity Q, inventory level at installation 3 depends only on the inventory reorder level s_3. Therefore, we denote the expected on-hand inventory and service level at installation 3 simply by $E[IL_3(s_3)]^+$ and $P(IL_3(s_3) > 0)$, respectively. Inventory at installation 2 is replenished with a one-for-one replenishment process; and its inventory level also depends on how the inventory is replenished at the upstream installation 3. Hence, $E[IL_2(s_2, s_3)]^+$ and

$P(IL_2(s_2, s_3) > 0)$ denote the expected on-hand inventory level and service level requirement at installation 2, respectively. Similarly, inventory at installation 1 is replenished with a one-for-one replenishment process; and its inventory level also is affected by how inventory is replenished at the upstream installations. Therefore, $E[IL_1(s_1, s_2, s_3)]^+$ and $P(IL_2(s_1, s_2, s_3) > 0)$ denote the expected on-hand inventory level and service level requirement at installation 1, respectively. We also denote the pre-specified minimum service-level requirement for class-i by $\beta_i \in [0, 1]$, $i \in \{1, 2, 3\}$. The optimization problem of interest for $N = 3$ demand classes can be stated as follows:

$$
\begin{aligned}
\text{Min} \quad & h \cdot (E[IL_1(s_1, s_2, s_3)]^+ + E[IL_2(s_2, s_3)]^+ + E[IL_3(s_3)]^+) \\
& s.t \\
& \quad P(IL_1(s_1, s_2, s_3) > 0) \geq \beta_1 \\
& \quad P(IL_2(s_2, s_3) > 0) \geq \beta_2 \\
& \quad P(IL_3(s_3) > 0) \geq \beta_3 \\
& \quad s_1, s_2, \in Z^+ \text{ and } s_3 \in Z
\end{aligned}
$$

Inventory reorder level s_3 at installation 3 can take any integer value. However, stockage levels at installation 1 and installation 2 must be positive since both of these two installations replenish their inventories through a one-for-one inventory replenishment policy and there is a service level requirement for each installation. Therefore, s_3 can take any integer value while s_1 and s_2 must take positive integer values.

We note that $s_1 = 0$ implies that demand class 1 and demand class 2 combine together and treated as a single single demand class. Similarly, $s_2 = 0$ implies that demand class 2 and demand class 3 are merged and treated as a single demand class. Each of these two cases results in an optimization problem with only $N = 2$ installations rather than $N = 3$ installations. Furthermore, $s_1 = 0$ and $s_2 = 0$ together imply that all three demand classes are treated as a single demand class. This last case, on the other hand, results in an optimization problem with a single installation rather than $N = 3$ installations.

To solve the above optimization problem, we need to derive explicit expressions for the expected on-hand inventory and the service level at each installation. As we describe earlier, we begin with deriving the steady-state distribution of inventory level at installation 3, which enables us to calculate the expected on-hand inventory at installation 3. The steady-state distribution of inventory level IL_3 at installation 3 can be calculated by conditioning on IP_3, the inventory position at installation 3, which is Uniformly distributed over $[s_3 + 1, s_3 + Q]$, as follows:

$$P(IL_3 = i) = \frac{1}{Q} \sum_{j=s_3+1}^{s_3+Q} P(D_1^L + D_2^L + D_3^L = j - i) \tag{4.11}$$

Knowing the steady-state distribution of IL_3 enables us to derive both the expected on-hand inventory level $E[IL_3(s_3)]^+$ and service level requirement at installation 3 as shown below, where for the service level problem the service level measure at any installation is simply the probability that on-hand inventory level on that installation is positive,— i.e, fill-rate type service level:

$$E[IL_3(s_3)]^+ = \sum_{i=1}^{s_3+Q} i \cdot P(IL_3 = i) \tag{4.12}$$

$$P(IL_3(s_3) > 0) = \sum_{i=1}^{s_3+Q} P(IL_3 = i) \tag{4.13}$$

Next, we derive the distributions of B_3 and $B_{3,2}$ as shown below:

$$P(B_3 = i) = P(IL_3 = -i), \text{ for } i \in Z^+ \text{ and } P(B_3 = 0) = P(IL_3 \geq 0) \tag{4.14}$$

$$P(B_{3,2} = i) = \sum_{j=i}^{\infty} P(B_3 = j) \cdot \binom{j}{i} \cdot (\frac{\lambda_1 + \lambda_2}{\lambda})^i \cdot (\frac{\lambda_3}{\lambda})^{j-i} \tag{4.15}$$

Having derived the distribution of $B_{3,2}$ enables us to determine the distribution of IL_2, the inventory level at installation 2, which is shown below:

$$P(IL_2 = i) = P(B_{3,2} = s_2 - i), \ i \in Z, \ i \leq s_2 \tag{4.16}$$

Expected on-hand inventory at installation 2 is a function of both the reorder level s_3 and s_2. And, knowing the steady-state distribution of IL_2 enables us to derive both the expected on-hand inventory level $E[IL_2(s_3, s_2)]^+$ and service level requirement at installation 2 as shown below:

$$E[IL_2(s_3, s_2)]^+ = \sum_{i=1}^{\infty} i \cdot P(IL_2 = i) \tag{4.17}$$

$$P(IL_2(s_3, s_2) > 0) = \sum_{i=1}^{\infty} P(IL_2 = i) \tag{4.18}$$

Next, we derive the distributions of B_2 and $B_{2,1}$ as shown below:

$$P(B_2 = i) = P(IL_2 = -i), \text{ for } i \in Z^+ \text{ and } P(B_2 = 0) = P(IL_2 \geq 0) \tag{4.19}$$

$$P(B_{2,1} = i) = \sum_{j=i}^{\infty} P(B_2 = j) \cdot \binom{j}{i} \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^{j-i} \tag{4.20}$$

Having derived the distribution of $B_{2,1}$ we can determine the distribution of IL_1, the inventory level at installation 1, which is shown below:

$$P(IL_1 = i) = P(B_{2,1} = s_1 - i), \ i \in Z, \ i \leq s_1 \tag{4.21}$$

Expected on-hand inventory at installation 1 is a function of the reorder levels s_3, s_2, and s_1. Furthermore, from the steady-state distribution of IL_1, we derive both the expected on-hand inventory level $E[IL_1(s_3, s_2, s_1)]^+$ and the service level requirement at installation 1 as indicated below:

$$E[IL_1(s_3, s_2, s_1)]^+ = \sum_{i=1}^{\infty} i \cdot P(IL_1 = i) \tag{4.22}$$

$$P(IL_1(s_3, s_2, s_1) > 0) = \sum_{i=1}^{\infty} P(IL_1 = i) \tag{4.23}$$

We have so far derived all the expressions that we need to characterize the objective function and service level constraints in the service level problem, which becomes a non-linear integer programming problem. Next, we need to solve it for optimality to determine the minimum inventory requirements. We also develop an optimal solution algorithm for the service level problem in the Numerical Algorithms section.

4.4 Cost Minimization Problem

Different from the service level problem, the cost minimization problem simply aims to determine the inventory policy that minimizes the total inventory costs that consist of expected inventory holding costs, and expected backlogging costs for not filling demands on time. It is important to point out that the cost minimization problem has a considerable drawback in practice because it requires measuring the cost of not filling customers orders on time — backlogging costs — which is often very hard to measure accurately in practice. Nevertheless, it is a good alternative model to the service level problem when accurate estimates of backlogging costs exist. We denote the backlogging cost per unit backlogged per unit time associated with class-i demands by $b_{i,i}, i \in [|1, N|]$. Similarly, the optimization problem of interest for $N = 3$ demand classes can be stated as follows, where we note that inventory holding cost h is the same for all demand classes. However, the backlogging costs are different across different demand classes:

$$
\begin{aligned}
\text{Min} \quad & h \cdot (E[IL_1(s_3, s_2, s_1)]^+ + E[IL_2(s_3, s_2)]^+ + E[IL_3(s_3)]^+) + \\
& b_{1,1} \cdot E[B_{1,1}(s_3, s_2, s_1)] + b_{2,2} \cdot E[B_{2,2}(s_3, s_2)] + b_{3,3} \cdot E[B_{3,3}(s_3)] \\
s.t & \\
& s_1, s_2, \in Z^+ \cup \{0\} \text{ and } s_3 \in Z
\end{aligned}
$$

We have already derived the expected on-hand inventory level at each installation in (4.12), (4.17), and (4.22) for the service level problem. The same expressions hold true for the cost minimization problem as well. Different from the service level problem, we additionally need to derive expressions for expected backorders that belong to each installation, i.e, $E[B_{i,i}]$, $i \in \{1, 2, 3\}$. These expressions can be similarly derived using the distribution of inventory level at each installation. We utilize the fact that the conditional distribution of backorders $B_{i,i}|B_i = x$, $i \in \{1,2,3\}$, is Binomial with parameters $\frac{\lambda_i}{\sum_{j=1}^{i}\lambda_j}$ and $x \in Z^+ \cup \{0\}$. Hence, $E[B_{i,i}|B_i = x]$ simply equals $x \cdot \frac{\lambda_i}{\sum_{j=1}^{i}\lambda_j}$. Namely,

$$E[B_{3,3}(s_3)] = \sum_{i=1}^{\infty} E[B_{3,3}|B_3 = x] \cdot P(B_3 = x) = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \cdot E[B_3(s_3)] \tag{4.24}$$

$$E[B_{2,2}(s_3, s_2)] = \sum_{i=1}^{\infty} E[B_{2,2}|B_2 = x] \cdot P(B_2 = x) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \cdot E[B_2(s_3, s_2)] \tag{4.25}$$

$$E[B_{1,1}(s_3, s_2, s_1)] = \sum_{i=1}^{\infty} i \cdot P(B_1 = i) \tag{4.26}$$

The cost minimization problem also becomes a non-linear integer programming problem. And, effective solution methods can be easily developed.

4.5 Service Time Problem

Timely fulfillment of orders is an important customer service measure in manufacturing and distribution practice. Many companies have set targets for fulfilling customer orders within a certain time period. One commonly used service measure is the percentage of times that orders are completely fulfilled when they are due. For example, when a manufacturer promises to provide a 90% service level for orders within 3 weeks of the demand arrival, it simply means that the orders are fulfilled after 3 weeks 9 out of 10 times. If the manufacturer promises instant fulfillment, then this boils down to the service level measure we analyze earlier in the service level problem — fill-rate type service level. In the service level problem our objective has been to satisfy each customer order immediately while satisfying certain service level requirement for each demand class. In this section, however, we generalize our customer service measure by incorporating a service time for each demand class. This new service measure is suitable for production/distribution environments where at least some demand classes are willing to wait for a specific time interval for the delivery of their orders.

We next model and formulate this problem for any finite number of demand classes. Consider a single location inventory system with N different demand classes and its equivalent serial inventory system with N installations. As described earlier, whenever a class-i demand occurs at installation $i \in [|1, N|]$, a replenishment request is placed at the immediate upstream installation $i+1$, that fills the replenishment request immediately if it has available inventory. Installation N, on the other hand, places its replenishment request of size Q to an outside supplier whenever inventory position at installation N drops to level s_N. And, each replenishment to installation N arrives in L time-units. We assume that class-i demand has a service time of w_i time-units. In other words, a class-i demand arriving at time t has a due-date at time $t+w_i$. We also assume that demands are not filled earlier than their due-dates. This assumption may be violated in practice. However, it is realistic in many JIT (Just-in-Time) environments where customers expect deliveries exactly at the promised date.

Installation $i \in [|1, N-1|]$ replenishes its inventory with a one-for-one replenishment process. Each replenishment request from the downstream installation at installation i makes the installation i to order one unit immediately. However, each class-i demand makes the installation i to order one unit after w_i time units. At installation i, replenishment requests are filled immediately while the fulfillment of each class-i demand at installation i is postponed at least w_i time units.

Each class-i demand that arrived at installation i over $(t-w_i, t]$ is filled later than time t. Moreover, each demand at any downstream installation $j \in [|1, i-1|]$ creates a replenishment request at installation $i \in [|1, N-1|]$, which is filled immediately at installation i. These replenishment requests from downstream installations, therefore, decrease the inventory position at installation i. The following relation takes into account the aforementioned arguments and provides us with the inventory level at installation i at time $t + \ell_i$:

$$IL_i(t+\ell_i) = s_i - \sum_{j=1}^{i-1} D_j(t, t+\ell_i) - D_i(t-w_i, t-w_i+\ell_i) - B_{i+1,i}(t) \tag{4.27}$$

(4.27) simply tells us that anything that is on order to installation i and not backlogged at installation $i+1$ at time t arrives at installation i over $(t, t+\ell_i]$. And, local demands at installation 1 through $i-1$ over $(t, t+\ell_i]$ create replenishment requests at installation i under one-for-one replenishment policy that reduce the inventory level at installation i. We note that we make no assumption on the relation between the replenishment leadtime and service time window at installation i.

The above expression in (4.27) is for any installation $i \in [|1, N-1|]$. Similarly, we can develop an expression for the inventory level at installation N. Hence, $IL_N(t+\ell_N)$, inventory level at the last installation N at time $t + \ell_N$ has the following form:

$$IL_N(t+\ell_N) = IP_N(t) + D_N(t-w_N, t) - \sum_{j=1}^{N-1} D_j(t, t+\ell_N) - D_N(t-w_N, t-w_N+\ell_N) \tag{4.28}$$

Our model development for the service time problem so far is general for a serial inventory system with exogenous demand at each installation. As described earlier in Section 4, we can map the original service time problem with N demand classes into an equivalent serial inventory system when internal replenishment leadtimes in the serial inventory system for multiple demand classes are all zero, i.e, $\ell_i = 0$, $i \in [|1, N-1|]$ while replenishment leadtime at installation N is positive, $\ell_N = L$. Applying this mapping yields the following relation for the inventory level at each installation.

$$IL_i(t) = s_i - B_{i+1,i}(t), \quad i \in [|1, N-1|] \tag{4.29}$$

$$IL_N(t+L) = IP_N(t) - \sum_{j=1}^{N-1} D_j(t, t+L) - D_N(t, t-w_N+L) \tag{4.30}$$

To help us in characterizing the service-time-fill-rates in the equivalent serial inventory system, we define the amount of replenishment arriving to installation to i from installation $i+1$ during time period $(t, t+w_i]$ by $A_i(t, t+w_i), i \in [|1, N|]$:

$$A_i(t, t+w_i) \triangleq \text{ the amount of replenishment arriving to installation } i \text{ in } (t, t+w_i]$$
$$= \sum_{j=1}^{i-1} D_j(t, t+w_i) + D_i(t-w_i, t) \tag{4.31}$$

We also observe that a class-i demand arriving at time t will be satisfied within w_i units of time after its arrival if the inventory level at installation i at time t plus the total replenishment to installation i over $(t, t+w_i]$ is greater than zero.

$$P(\text{class-}i \text{ demand filled within } (t, t+w_i]\ ) = P(IL_i(t) + A_i(t, t+w_i) > 0)$$

Combining equations (4.29), (4.30), and (4.31) provides us with the following expression(s), where $i \in [|1, N-1|]$:

$$\begin{aligned} P(\text{class-}i \text{ demand filled within } (t, t+w_i]\ ) = {} & P(s_i - B_{i+1,i}(t) + \\ & \sum_{j=1}^{i-1} D_j(t, t+w_i) + D_i(t-w_i, t) > 0) \end{aligned} \tag{4.32}$$

$$\begin{aligned} P(\text{class-}N \text{ demand filled within } (t, t+w_N]\ ) = {} & P(IP_N(t-L) - \\ & \sum_{j=1}^{N-1} D_j(t-L, t) - D_N(t-L, t-w_N) + \\ & \sum_{j=1}^{N-1} D_j(t, t+w_N) + D_N(t-w_N, t) > 0) \end{aligned} \tag{4.33}$$

Equilibrium distribution of inventory level at each installation exists, hence, in steady state the above set of equations become as follows, where D_i^w denotes the equilibrium distribution of demand over w time units at installation $i \in [|1, N|]$:

$$P(\text{class-}i \text{ demand filled within } w_i \text{ time units }) = P(s_i - B_{i+1,i} + \sum_{j=1}^{i-1} D_j^{w_i} + D_i^{w_i} > 0)$$

$$P(\text{class-}N \text{ demand filled within } w_N \text{ time units}) = P(IP_N - \sum_{j=1}^{N-1} D_j^L - D_N^{L-w_N} + \sum_{j=1}^{N-1} D_j^{w_N} + D_N^{w_N} > 0) \qquad (4.34)$$

As we mentioned earlier, IP_N, the inventory position at installation N in (4.34), has a Uniform distribution over $[s_N + 1, s_N + Q]$. Using the above steady-state expressions in (4.29) - (4.34), we can derive all desired performance measures to construct our optimization problem. Namely, the service time problem simply aims to find out the optimal one-for-one base stock levels s_i, $i \in [|1, N-1|]$ and optimal reorder point s_N at installation N that minimize the average inventory holding costs subject to each service level requirement for each demand class, where the service level requirement for each demand class consists of filling its orders on the pre-specified due date with a pre-specified percentage — i.e, service time demand fulfillment rate:

$$\text{Min} \quad h \cdot \sum_{i=1}^{N} E[IL_i(s_N, s_{N-1}, \cdots, s_i)]^+$$

$$s.t$$

$$P(\text{class-}i \text{ demand filled within } w_i) \geq \beta_i, i \in [|1, N|]$$

$$s_i \in Z^+ \cup \{0\}, i \in [|1, N-1|] \text{ and } s_N \in Z$$

One very important attribute of the service time problem is that it can be used to serve as a system design tool. It enables us to construct cost curves that reflect the trade-off between holding inventory and achieved service time demand fulfilment rate for each demand class. Using these cost curves, we can synthesize the impact of providing shorter service time windows on total inventory costs.

Example

In this example, we consider that there are only three demand classes for the service time problem. Demand is Poisson distributed with rates $\lambda_1 = 8$ units/year, $\lambda_2 = 12$ units/year, and $\lambda_3 = 16$ units/year for demand class 1, demand class 2, and demand class 3, respectively. Inventory is replenished from an outside supplier with a replenishment lead-time of 3 months. Replenishment requests are done with a one-for-one replenishment process, $Q = 1$ units. Demand class 1 has a service time window $w_1 = 3$ days; demand class 2 has a service time window $w_2 = 7$ days; and demand class 3 has a service time window $w_3 = 15$ days. Also, minimum service level for demand class 1 is $\beta_1 = 0.98$; minimum service level for demand class 2 is $\beta_2 = 0.93$; and minimum service level for demand class 3 is $\beta_3 = 0.77$. Inventory dynamics at each installation in the equivalent serial inventory system in steady state become as follows:

$$\begin{aligned} IL_1 &= s_1 - B_{2,1} \\ IL_2 &= s_2 - B_{3,2} \end{aligned} \tag{4.35}$$

$$IL_3 = \begin{cases} IP_3 - \sum_{j=1}^{2} D_j^L - D_3^{L-w_3}, & \text{if } w_3 < L; \\ IP_3 - \sum_{j=1}^{2} D_j^L + D_3^{w_3-L}, & \text{if } w_3 \geq L. \end{cases} \tag{4.36}$$

Using the above expressions for distributions of inventory levels at installations, expected on hand inventory and service level at each installation can be easily derived. Minimizing the sum of the expected on hand inventories subject to the required minimum service level requirements result in the following optimal solution, $s_1 = 1$ unit, $s_2 = 3$ units, and $s_3 = 9$ units. If the service time for each demand class is zero $w_i = 0$, $i \in \{1, 2, 3\}$, then the problem boils down to the service level problem. Stocking $s_1 = 1$, $s_2 = 3$, and $s_3 = 9$ units in the service level problem results in the following service levels 0.98, 0.91, and 0.45 for demand class 1, class 2, and class 3, respectively. The difference in service levels between the service time problem and the service level problem clearly shows the benefit of having service time in reducing the amount of inventory to hold. Furthermore, if we treat all demand classes as if they were all class 1 with a service time $w_1 = 3$ days and a minimum service level $\beta_1 = 0.98$, then the optimal base-stock level for this single demand class problem will be $s = 16$ units.

4.6 Numerical Algorithms

In this section we develop numerical algorithm(s) for the service level problem that we analyzed earlier. We first consider the simplest scenario when there are only two demand classes $N = 2$ and order replenishment process is one-for-one with an order size of $Q = 1$ unit. Later, we extend our findings to solution algorithms for N demand classes. In this simple case with two demand classes, our problem maps to a serial inventory system with two installations that has the following forms of inventory equations in steady-state:

$$
\begin{aligned}
IL_2 &= s_2 - (D_1^L + D_2^L) \\
IL_1 &= s_1 - B_{2,1}
\end{aligned}
$$

We begin with developing a numerical solution method for the service level problem that can be stated as follows, where the objective is to minimize total expected inventory holding costs while satisfying fill-rate service level constraints for the two demand classes:

$$
\begin{aligned}
Min \quad & h \cdot (E[IL_2(s_2)]^+ + E[IL_1(s_2, s_1)]^+) \\
s.t \quad & \\
& P(IL_2(s_2) > 0) \geq \beta_2 \\
& P(IL_1(s_2, s_1) > 0) \geq \beta_1 \\
& s_2 \in Z, s_1 \in Z^+ \cup \{0\}
\end{aligned}
$$

We next analyze the objective function in the above service level problem in more detail. Since inventory holding cost rate h is constant, we can simply minimize the sum of expected on-hand inventory levels while satisfying service level constraints. For a given set of installation base stock levels, (s_2, s_1), at installations we denote total expected on-hand inventories with $TC(s_2, s_1) = E[IL_2]^+ + E[IL_1]^+$. Furthermore, the expected on-hand inventory at installation $i \in \{1, 2\}$ can be written as follows:

$$E[IL_i]^+ = E[IL_i + [IL_i]^-] = E[IL_i] + E[B_i] \tag{4.37}$$

$$E[IL_2]^+ = s_2 - E[(D_1^L + D_2^L)] + E[B_2] \tag{4.38}$$

$$E[IL_1]^+ = s_1 - E[B_{2,1}] + E[B_1] \tag{4.39}$$

Since $E[B_2] = E[B_{2,2}] + E[B_{2,1}]$ holds true, we can write $TC(s_2, s_1)$ in the following form:

$$TC(s_2, s_1) = s_2 + s_1 - E[(D_1^L + D_2^L)] + E[B_{2,2}(s_2)] + E[B_1(s_2, s_1)] \tag{4.40}$$

For a given backorder level at installation 2, backorders that are due to demand at installation 2 are Binomially distributed with rate $\frac{\lambda_2}{\lambda_1+\lambda_2}$. Hence, we can further write down the following explicit expression for expected backorders for demand class 2 that are due at installation 2:

$$E[B_{2,2}(s_2)] = \frac{\lambda_2}{\lambda_1 + \lambda_2} E[B_2] = \frac{\lambda_2}{\lambda_1 + \lambda_2} \sum_{i=1}^{\infty} i \cdot P(D_1^L + D_2^L = s_2 + i) \tag{4.41}$$

Similarly, expected backorders at installation 1 can be written in the following explicit form that depends on both stockage levels s_1 and s_2:

$$\begin{aligned} E[B_1(s_2, s_1)] &= \sum_{i=s_1}^{\infty} (i - s_1) \cdot P(B_{2,1} = i) \\ &= \sum_{i=s_1}^{\infty} (i - s_1) \cdot \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^{(j-i)} \cdot \\ &\quad P(D_1^L + D_2^L = s_2 + j) \end{aligned} \tag{4.42}$$

Let us now consider that one unit of stock is moved from installation 1 and placed into installation 2. This change in total allocation of stocks results in a new total expected on-hand inventory that is represented by the following expression:

$$TC(s_2+1, s_1-1) = s_2 + s_1 - E[D_1^L + D_2^L] + E[B_{2,2}(s_2+1)] + E[B_1(s_2+1, s_1-1)]$$

By moving one unit from installation 1 to installation 2, backorders at installation 2 that are destined to installation 2 changes in the following way:

$$\begin{aligned} E[B_{2,2}(s_2+1)] &= \frac{\lambda_2}{\lambda_1+\lambda_2} \sum_{i=1}^{\infty} iP(D_1^L + D_2^L = s_2 + 1 + i) \\ &= \frac{\lambda_2}{\lambda_1+\lambda_2} \sum_{i=1}^{\infty} (i-1)P(D_1^L + D_2^L = s_2 + i) \\ &= E[B_{2,2}(s_2)] - \frac{\lambda_2}{\lambda_1+\lambda_2} P(D_1^L + D_2^L \geq s_2 + 1) \end{aligned} \tag{4.43}$$

The above expression simply provides a relation between $E[B_{2,2}(s_2+1)]$ and $E[B_{2,2}(s_2)]$. Similarly, we can derive the following relation between $E[B_1(s_2+1, s_1-1)]$ and $E[B_1(s_2, s_1)]$, which informs us on how shifting one unit of stock to the upstream installation changes expected backorder level at installation 1:

$$\begin{aligned} E[B_1(s_2+1, s_1-1)] &= \sum_{i=s_1-1}^{\infty} (i-(s_1-1)) \cdot \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^i \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{(j-i)} \cdot \\ &\quad P(D_1^L + D_2^L = s_2 + 1 + j) \\ &= \sum_{i=s_1}^{\infty} (i-s_1) \cdot \sum_{j=i}^{\infty} \binom{j-1}{i-1} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{i-1} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{(j-i)} \cdot \\ &\quad P(D_1^L + D_2^L = s_2 + j) \\ &= E[B_1(s_2, s_1)] + \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot P(B_{2,1}(s_2+1) > s_1) \end{aligned}$$

Using the above derived relations between expected backorder levels at installation 1 and 2, we can derive the following relation between total expected on-hand inventory before one unit of stock is moved and after it is moved:

$$\begin{aligned}
TC(s_2+1,s_1-1) &= s_2+s_1-E[D_1^L+D_2^L]+E[B_{2,2}(s_2+1)]+E[B_1(s_2+1,s_1-1)] \\
&= s_2+s_1-E[D_1^L+D_2^L]+E[B_{2,2}(s_2)]-\frac{\lambda_2}{\lambda_1+\lambda_2}\cdot P(B_2(s_2)>0) \\
&\quad +E[B_1(s_2,s_1)]+\frac{\lambda_2}{\lambda_1+\lambda_2}\cdot P(B_1(s_2+1,s_1-1)>0) \\
&= TC(s_2,s_1)-\frac{\lambda_2}{\lambda_1+\lambda_2}[P(B_2(s_2)>0)-P(B_1(s_2+1,s_1-1)>0)]
\end{aligned}$$

Let us now compare $TC(s_2+1, s_1-1)$ with $TC(s_2, s_1)$. To achieve this comparison, we need to determine the relation between $P(B_2(s_2) > 0)$ and $P(B_1(s_2+1, s_1-1) > 0)$. And, the following proposition provides this information.

Proposition 12 *For $s_1 \in Z^+ \cup \{0\}, s_2 \in Z$, the following inequality holds true,*

$$P(B_2(s_2)>0) > P(B_1(s_2+1,s_1-1)>0)$$

We know that $TC(s_2+1, s_1-1) = TC(s_2, s_1) - \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot [P(B_2(s_2) > 0) - P(B_1(s_2+1, s_1-1) > 0)]$. Furthermore, $P(B_2(s_2) > 0) > P(B_1(s_2+1, s_1-1) > 0)$ holds true. Hence, moving one unit of stock from the downstream installation 1 to the upstream installation 2 results in a decrease in total expected on-hand inventory level. We next develop a sequential solution method that helps us in characterizing the optimal solution method for Service Level Problem with two demand classes and $Q = 1$. We call this auxiliary solution method as *Single-Pass-Algorithm.* Single pass algorithm finds the stockage level at each installation sequentially starting with the most upstream installation. Namely, when there are only two installations

$$\begin{aligned}
\widehat{s}_2 &= \min\{s_2 : P(IL_2(s_2)>0) \geq \beta_2\} \\
\widehat{s}_1 &= \min\{s_1|\widehat{s}_2 : P(IL_1(\widehat{s}_2,s_1)>0) \geq \beta_1\}
\end{aligned}$$

$\widehat{s}_2$ gives the minimum stock level at installation 2 such that the fill rate requirement at installation 2 is satisfied. Hence, $\widehat{s}_2$ is a lower bound on the optimal stock level at installation 2. Stock level at installation 1 does not affect the fill rate level at installation 2. However, the fill rate level at installation 1 is affected by the stock level at installation 2. Hence, Single-Pass-Algorithm first finds the minimum stock level $\widehat{s}_2$ at installation 2 that satisfies the fill rate requirement at installation 2, then it moves to installation 1. And, for this given stock level $\widehat{s}_2$ at installation 2, it finds the minimum stock level $\widehat{s}_1$ at installation 1 that satisfies the required fill rate at installation 1. After defining how Single-Pass-Algorithm works, we develop the following result that indicates how the solution for Single-Pass-Algorithm changes as we move units between the two installations.

Proposition 13 *If $\widehat{s}_1(s_2) \in Z^+$ denotes the optimal stockage level at installation* 1 *for a given $s_2 \in Z$ stockage level at installation* 2*, then the following holds true*

$$\widehat{s}_1(s_2 + 1) \in \{\widehat{s}_1(s_2) - 1, \widehat{s}_1(s_2)\}$$

The above result simply implies that if one unit of stock is added to installation 2, the solution at installation 1 for Single-Pass-Algorithm after this addition is either the same as before the addition or it is only one unit less than the solution before the addition. Once we solve the Single-Pass-Algorithm, we obtain the following set of solutions, $(\widehat{s}_2, \widehat{s}_1)$. The last result implies that adding one more unit into installation 2 results in a new feasible solution that is either $(\widehat{s}_2 + 1, \widehat{s}_1)$ or $(\widehat{s}_2 + 1, \widehat{s}_1 - 1)$. The total on-hand inventory level with solution set $(\widehat{s}_2 + 1, \widehat{s}_1)$ is surely larger than that with solution set $(\widehat{s}_2, \widehat{s}_1)$. However, the on-hand inventory level with solution set $(\widehat{s}_2 + 1, \widehat{s}_1 - 1)$ is smaller than that with $(\widehat{s}_2, \widehat{s}_1)$ because we already show earlier that shifting one unit of stock from a downstream installation to an upstream results in a lower total on-hand inventory level. This observation suggests that applying first Single-Pass-Algorithm and then shifting stocks from installation 1 to installation 2 as far as service level requirements are still satisfied results in an optimal solution set (s_2^*, s_1^*) for our problem with two installations. The formal description of this optimal solution algorithm is as follows:

Algorithm for Two Demand Classes:

1. carry out Single-Pass-Algorithm to determine $\widehat{s}_2$ and $\widehat{s}_1$,

2. set $s_2 = \widehat{s}_2$, $s_1(s_2) = \widehat{s}_1$, $s_2^* = \widehat{s}_2$, $s_1^* = \widehat{s}_1$, and *cost* $= TC(s_2, s_1)$,

3. DO WHILE $P(IL_1(s_2+1, s_1(s_2) - 1 > 0) \geq \beta_1$

   set $s_2 = s_2 + 1$, $s_1(s_2) = s_1(s_2 - 1) - 1$ and *cost* $= TC(s_2, s_1(s_2))$

   END DO

Next, we extend the above optimal solution algorithm for a problem with more than two demand classes. To achieve this, we first develop the extension to the result in Proposition 13 for any finite number of demand classes. For a problem with N demand classes, the Single-Pass-Algorithm is about solving the following set of equations sequentially:

$$
\begin{aligned}
\widehat{s}_N &= \min\{s_N : P(IL_N(s_N) > 0) \geq \beta_N\} \\
\widehat{s}_{N-1} &= \min\{s_{N-1}|\widehat{s}_N : P(IL_N(\widehat{s}_N, s_{N-1}) > 0) \geq \beta_{N-1}\} \\
&\;\;\vdots \\
\widehat{s}_1 &= \min\{s_1|\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_2) : P(IL_1(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_2, s_1) > 0) \geq \beta_1\}
\end{aligned}
$$

The following proposition characterizes how the solution for the Single-Pass-Algorithm changes as we change the stockage level at some installation. And, we utilize this result later on our solution algorithm.

Proposition 14 *If the solution for Single-Pass-Algorithm is increased by one unit at some installation* $i \in [|1, N|]$, *i.e,* $\widehat{s}_i \to \widehat{s}_i + 1$, *then the solution for Single-Pass-Algorithm for all downstream installations changes at most one installation.*

For the clarity of presentation we present a solution algorithm for a problem with $N = 3$ demand classes. However, a similar development applies for any finite number of demand classes. Consider now the equivalent serial inventory system with 3 installations. Then, the total on-hand inventory level for a given set of stockage level (s_3, s_2, s_1) becomes,

$$
TC(s_3, s_2, s_1) = \sum_{i=1}^{3} s_i - \sum_{i=1}^{3} E[D_i^L] + E[B_{3,3}(s_3)] + E[B_{2,2}(s_3, s_2)] + E[B_1(s_3, s_2, s_1)]
$$

Let us now consider how an increase in stockage level at any of the three installation affects total on-hand inventory level. There are three options to increase the stockage level; adding unit(s) to installation 3; installation 2; or installation 1. First, suppose that one unit of stock is added to installation 3 so that (s_3+1, s_2, s_1) becomes the new set of stockage levels. This additional inventory at installation 3 results in decrease in expected backorders at all three installations as indicated below:

$$\begin{aligned}
\Delta_{3,3}(s_3) &= E[B_{3,3}(s_3)] - E[B_{3,3}(s_3+1)] = \frac{\lambda_3}{\lambda} \cdot P(B_3(s_3) > 0) \\
\Delta_{2,2}(s_3) &= E[B_{2,2}(s_3, s_2)] - E[B_{2,2}(s_3+1, s_2)] \\
\Delta_1(s_3) &= E[B_1(s_3, s_2, s_1)] - E[B_1(s_3+1, s_2, s_1)]
\end{aligned}$$

We characterize the total change in expected backorders through adding one unit of stock in the following proposition.

Proposition 15 *For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation* 3 *is less than or equal to one unit.*

The above proposition implies that adding one unit of stock at installation 3 decreases the total expected backorders by one unit at most, therefore, increasing the stock level at installation 3 by one unit increases the total expected on-hand inventory level. This further implies that increasing the stockage level at installation 3 increases inventory holding costs. Let us now consider that the stockage level at installation 2 is increased by one unit so that (s_3, s_2+1, s_1) becomes the new set of stockage levels. Similarly, the increase in stockage level at installation 2 results in decrease in expected backorders both at installation 2 and installation 1 as shown below:

$$\begin{aligned}
\Delta_{2,2}(s_2) &= E[B_{2,2}(s_3, s_2)] - E[B_{2,2}(s_3, s_2+1)] \\
\Delta_1(s_2) &= E[B_1(s_3, s_2, s_1)] - E[B_1(s_3, s_2+1, s_1)]
\end{aligned}$$

We characterize the above mentioned decrease in expected backorders at installation 2 in the following proposition.

Proposition 16 *For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation* 2 *is less than or equal to one unit.*

Similarly, the above proposition implies that increasing the stockage level at installation 2 by one unit results in a decrease in expected backorders that is less than or equal to one unit. Therefore, increasing the stockage level at installation 2 by one unit increases the expected on-hand inventory level at installation 2, hence, the expected inventory holding costs. Lastly, we consider adding one unit of stock at installation 1 so that the new set of stockage levels becomes $(s_3, s_2, s_1 + 1)$. This additional one unit of stock at installation 1 does not affect the expected backorder levels at the upstream installations 2 and 3, hence, only expected backorder level at installation 1 decreases with the additional of one unit of stock at installation 1, i.e $\Delta_1(s_1) = E[B_1(s_3, s_2, s_1)] - E[B_1(s_3, s_2, s_1 + 1)]$. We characterize the decrease in the expected backorder level at installation 1 with the increase in stockage level at installation 1 in the following proposition.

Proposition 17 *For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation* 1 *is less than or equal to one unit.*

Similar to the previous two results, the above result simply implies that increasing the stockage level at installation 1 by one unit results in a decrease in expected backorders that is less than or equal to one unit. Increasing the stockage level at installation 1 by one unit, hence, increases the expected on-hand inventory level and the expected inventory holding costs.

In summary the above derived results together imply the following important finding: increasing stockage level at any of the three installations results in a decrease in expected backorders that is less than or equal to one unit. Hence, the total expected on-hand inventory level increases with the increase in stockage levels. Namely, it holds true that $TC(s_3, s_2, s_1)$ is an increasing function in stockage levels s_3, s_2, and s_1. Using the similar approach we can easily show that the same property holds true for the general case $TC(s_N, s_{N-1}, ..., s_1)$. Next, we develop our last auxiliary result that helps us derive an optimal solution algorithm when there are $N = 3$ demand classes.

In this last result we would like to derive how total expected on-hand inventory level $TC(s_3, s_2, s_1)$ changes as we move one unit of stock from a downstream installation and place it to an upstream installation. We characterize how this change in $TC(s_3, s_2, s_1)$ occurs in the following proposition:

Proposition 18 *For a given set of stockage levels $s_3 \in Z, s_2$ and $s_1 \in Z^+ \cup \{0\}$,*

$$i)\ TC(s_3, s_2 + 1, s_1 - 1) < TC(s_3, s_2, s_1)$$

$$ii)\ TC(s_3 + 1, s_2 - 1, s_1) < TC(s_3, s_2, s_1)$$

The above proposition simply tells us that total expected on-hand inventory level $TC(\cdot)$ decreases as we move stocks from downstream installations to upstream installations. The same result can be easily derived for any finite number of demand classes using the same procedure. Next, we gather all our findings together to construct an optimal solution algorithm. Before we construct our optimal solution algorithm we restate all our main results so far for the sake of clarity,

- Single-Pass-Algorithm yields a feasible solution for Service Level Problem,
- Single-Pass-Algorithm provides Service-Level-Problem with a feasible solution that has the smallest value of the sum of stockage levels, i.e min $\sum_{i=1}^{N} s_i$,
- If the solution for Single-Pass-Algorithm is increased by one unit at some installation $i \in [|1, N|]$, i.e, $\widehat{s}_i \rightarrow \widehat{s}_i + 1$, then the solution for Single-Pass-Algorithm for all downstream installations changes at most one installation by one unit decrease.
- For a given set of stockage levels $\{s_3, s_2, s_1\}$, increasing the stockage level at any installation results in an increase in the objective function value $TC(\cdot)$ of Service Level Problem,

- For a given set of stockage levels $\{s_3, s_2, s_1\}$, shifting the stocks from downstream installations to upstream installations yields lower total on-hand inventory level, hence, lower objective function value $TC(\cdot)$ for Service Level Problem,

Algorithm for Three Demand Classes: We now suggest a numerical solution algorithm for the Service Level Problem and discuss why it yields an optimal solution. Consider now the Service Level Problem with 3 demand classes. First, we suggest applying Single-Pass-Algorithm to obtain a good feasible solution for Service Level Problem. Let the solution set obtained by Single-Pass-Algorithm (SPA) be $\{\widehat{s}_3, \widehat{s}_2, \widehat{s}_1\}$ such that the following holds true:

$$
\begin{aligned}
\widehat{s}_3 &= \min\{s_3 : P(IL_3(s_3) > 0) \geq \beta_3\} \\
\widehat{s}_2 &= \min\{s_2 | \widehat{s}_3 : P(IL_2(\widehat{s}_3, s_2) > 0) \geq \beta_2\} \\
\widehat{s}_1 &= \min\{s_1 | \widehat{s}_3, \widehat{s}_2 : P(IL_1(\widehat{s}_3, \widehat{s}_2, s_1) > 0) \geq \beta_1\}
\end{aligned}
$$

Clearly, the sum of stockage levels $\widehat{s} = \widehat{s}_3 + \widehat{s}_2 + \widehat{s}_1$ obtained through Single-Pass-Algorithm provides a lower bound on the sum of stockage levels for an optimal solution for Service Level Problem. Hence, the optimal solution has the sum of stocks that is either $\widehat{s}$ or greater than $\widehat{s}$. Increasing stockage levels obtained in Single-Pass-Algorithm still yields a feasible set of stockage levels, however, we show that increasing stockage levels increases total expected on-hand inventory level, and therefore, the objective function value in Service Level Problem. Therefore, the sum of stockage levels in an optimal solution must be exactly equal to $\widehat{s} = \widehat{s}_3 + \widehat{s}_2 + \widehat{s}_1$. Also, we show that shifting stocks from downstream installations to upstream installations decreases the objective function value $TC(\cdot)$ of the Service Level Problem. Hence, moving stocks to upstream installations as long as feasibility is satisfied results in a better solution. This observation suggests that we should begin with moving stocks from installation 1 to installation 3 as long as service level requirements are still satisfied. When no more stock is allowed to be shifted from installation 1 to installation 3, we next move stocks to installation 2 instead of installation 3 as long as feasibility is preserved.

When no more stock is able to be moved to installation 2, we begin moving stocks from installation 2 to installation 3; and we continue this move from installation 2 to installation 3 as long as feasibility is preserved. This lateral shift of stocks does not change the optimal sum of stockage levels. Moreover, by shifting stocks to upstream installations while still satisfying feasibility further reduces the objective function value, which guarantees that the resulting set of stockage levels is indeed the optimal set of stockage levels. The formal description of this optimal solution algorithm is as follows:

Algorithm for Three Demand Classes:

1. carry out Single-Pass-Algorithm to determine $\widehat{s}_3$, $\widehat{s}_2$, and $\widehat{s}_1$,

2. set $s_3 = \widehat{s}_3$, $s_2 = \widehat{s}_2$, $s_1 = \widehat{s}_1$; and $cost = TC(s_3, s_2, s_1)$,

3. DO WHILE (service level at installation i) $\geq \beta_i, \forall i \in \{1, 2, 3\}$
   set $s_3 = s_3 + 1$, $s_1 = s_1 - 1$ and $cost = TC(s_3, s_2, s_1)$
   END DO

4. DO WHILE (service level at installation i) $\geq \beta_i, \forall i \in \{1, 2, 3\}$
   set $s_2 = s_2 + 1$, $s_1 = s_1 - 1$ and $cost = TC(s_3, s_2, s_1)$
   END DO

5. DO WHILE (service level at installation i) $\geq \beta_i, \forall i \in \{1, 2, 3\}$
   set $s_3 = s_3 + 1$, $s_2 = s_2 - 1$ and $cost = TC(s_3, s_2, s_1)$
   END DO

The above solution algorithm can be easily extended using the similar approach to account for any finite number of demand classes. To illustrate the above presented optimal solution algorithm we provide the following example:

Example

Consider that there are three demand classes with Poisson arrival rates $\lambda_1 = 8$ units/year, $\lambda_2 = 12$ units/year, and $\lambda_3 = 16$ units/year for demand class 1, class 2, and class 3, respectively. Constant order replenishment leadtime is $L = 3$ months. Inventory is replenished through a one-for-one replenishment process with order size $Q = 1$. Class 1 demands require a fill-rate of 99%, class 2 demands require a fill-rate of 94%, and class 3 demands require a fill-rate of 87%. Service Level Problem finds out the optimal set of stockage levels $\{s_3, s_2, s_1\}$ that minimizes the long-run average inventory holding costs while satisfying the required service level requirements. Applying Single-Pass-Algorithm yields the following stockage levels: $\widehat{s}_3 = 13$, $\widehat{s}_2 = 1$, and $\widehat{s}_1 = 2$. And, the total expected on-hand inventory level under these stockage levels becomes $TC(\widehat{s}_3, \widehat{s}_2, \widehat{s}_1) = 7.0905$. Shifting stocks from installation 1 to installation 3 violates feasibility. Hence, we move stocks from installation 1 to installation 2 instead. Due to feasibility requirement we are only able to move one unit from installation 1 to installation 2, which results in $s_3 = 13, s_2 = 2, s_1 = 1$. And, this move produces total expected on-hand inventory level $TC(s_3, s_2, s_1) = 7.0778$. Since moving any unit from installation 2 to installation 3 violates feasibility, $\{s_3 = 13, s_2 = 2, s_1 = 1\}$ becomes indeed the optimal solution set of stockage levels.

If fill-rate requirements become 99% for class 1, 93% for class 2, and 70% for class 3, then Single-Pass-Algorithm yields $\widehat{s}_3 = 11$, $\widehat{s}_2 = 2$, and $\widehat{s}_1 = 2$. And, this solution yields total expected on-hand inventory level $TC(\widehat{s}_3, \widehat{s}_2, \widehat{s}_1) = 6.2410$. It is feasible to move one unit from installation 1 to installation 3, which yields total expected on-hand inventory level $TC(\widehat{s}_3 + 1, \widehat{s}_2, \widehat{s}_1 - 1) = 6.1413$. Since no stock is feasible to move from installation 1 to installation 2, and from installation 2 to installation 3, optimal solution set of stockage levels becomes $\{s_3 = 12, s_2 = 2, s_1 = 1\}$.

In conclusion, we are able to develop a solution algorithm for Service Level Problem with one-for-one replenishment process such that it yields an optimal solution. Since the Service Time Problem is similar in structure to the Service Level Problem, the presented solution algorithm can be easily suited to solve the Service Time Problem as well. Furthermore, this solution algorithm can be modified to solve Service Level Problem with (R, Q) replenishment process. Suppose that order size Q is fixed. Then, it is clear that we need to set ordering level R high enough to meet the service requirement at the most upstream installation N. Moreover, increasing ordering level R improves service level not only at installation N, but also at all downstream installations. Hence, for given order size Q, our solution algorithm provides the optimal ordering level at installation N and optimal base-stock levels at all downstream installations. Furthermore, to determine the optimal ordering quantity Q, we can easily construct a search algorithm over the possible range of ordering size Q; and apply the solution algorithm for each of the possible ordering size to pick up the best Q.

Chapter 5

Summary & Future Research

We present in our paper several descriptive models with the objective of guiding a manufacturer about the optimal experimental and pilot production strategies for a new product that he plans to launch. We first analyze the question of how much capacity should be allocated for experimental production. To answer this question we develop an optimization problem that maximizes profits, — single line capacity allocation problem (SLCAP). Analysis of this problem provides us with the following result: it is never optimal to dedicate only limited capacity to the new product, — i.e., experimental production is not optimal. Further analysis also has yielded that it is optimal to launch full production of a new product either immediately or only when the mature product that is currently in production becomes obsolete. This result is quite interesting because experimental production is commonly used in many industries, which is in contrast to our derived result. Hence, our result lays out some sufficient conditions that suggest not to using experimental production. This is quite helpful for managers to determine what factors they need to focus on in developing experimental production strategies for their newly developed products. To the best of our knowledge, the previous work on experimental production in literature have not considered the issue of capacity allocation explicitly, thus our model extends the previous work in literature by explicitly modelling the capacity trade-off between two consecutive generations of products to develop experimental production strategies.

It is also worth mentioning that we model production yield levels as a function of cumulative production experience. Specifically, we treat production yield levels as an increasing-concave function of cumulative production quantity. Production experience models developed in the literature assume unit cost of production goes down with more production experience. We can easily develop a corresponding one-to-one mapping between unit production cost and production yield levels. This mapping is quite useful since we can characterize process improvement not only through reductions in unit production cost, but also through improvement in production yield levels. Hence, our production yield model generalizes the previously developed yield models in the literature.

We realize that additional questions arise from the insights obtained from the experimental production model. First, experimental production model treats each production line independent of other existing production lines. However, if there are multiple production lines with similar production processes, experience gained on one production line could be transferred to other production lines. Thus, it is not clear any more if the simple result for the experimental production model will hold true when there are multiple production lines with similar production processes. Second, experimental production model treats production yield levels as deterministic parameters. And, it is important to check what effect uncertain production yields, coupled with multiple production lines with similar production processes, would have on experimental and pilot production strategies.

These questions are answered by developing a knowledge transfer model among production lines. The two important characteristics of this model are: First, the higher the similarity in production processes between pilot production line and any other production line implies the higher the amount of production experience transferred from pilot production line; hence, the higher the benefit from pilot production. Second, the more production capacity for pilot production implies the faster improvement in production process, hence, the more production experience is gathered through pilot production.

We first show when a pilot production method becomes a desirable production strategy. We consider two different environments; production yields with no uncertainties and production yields with uncertainties. Next, we show how benefit from pilot production varies with the allocated production capacity for pilot production. Specifically, we derive conditions that guarantee that total profit function is discrete-unimodal with respect to allocated capacity for pilot production. These results provide managers guidelines on how to derive the most benefit from pilot production through optimal capacity allocation. Moreover, the model itself sheds light on trade-off between the pace of improvement through pilot production and the opportunity cost involved with pilot production Hence, our model make the trade-offs involved with pilot production more transparent for managers.

Next, we develop a model to characterize how production processes of a new product can be improved during its full production. To achieve this, we consider the case of full production where a lead production line scans information on process improvements on a continuous basis with the remaining production lines. However, this knowledge sharing process involves costs that need to be traded-off with benefits from process improvements. We determine when knowledge from a lead production line about process improvements for a new product should be transferred through a yield control mechanism. Furthermore, we provide an example to illustrate our suggested yield control mechanism. The developed yield control mechanism is based on yield differences which stand out as an opportunity to improve production yields through knowledge transfer. Observing yield differences among production lines is fairly easy task to achieve; therefore, the yield control mechanism would become a benchmark tool for managers on deciding when it would be best to transfer knowledge among production lines.

In conclusion, we would like to point out some possible limitations and pitfalls to the approaches presented here. First, we have implicitly assumed that there is no competition. Second, additive random production yield model described in this paper has not been modelled in literature before. There is not much work in literature on how uncertainty in production process is reflected on production yield levels. Therefore, it is necessary to have further empirical justification for our additive random production yield model. Third, total profit function in section 4 for a deterministic production yield environment results in not being very sensitive to changes in n, total number of production lines. Some possible limitations for this result are: yield levels are all deterministic; and a single knowledge transfer is event is considered. We conjecture that the sensitivity would be higher if we consider a more flexible model with multiple knowledge transfer opportunities. We plan to test our conjecture in our future work. Finally, our yield control mechanism in section 5 assumes that yield difference between the lead production line and any other production line is increasing-concave as far as there is a knowledge and experience gap between them. It would be helpful if we can further generalize this assumption to a larger extent. Furthermore, it would be also useful to derive a mapping between different industries and their corresponding yield difference functions. Moreover, our yield control mechanism assumes only a fixed cost associated with knowledge transfer; extension to include variable costs associated with knowledge transfer would be of interest to make the model more realistic. These aforementioned extensions are in our agenda for our future work.

We have presented an inventory control policy for an inventory system with multiple classes of customers. We developed an equivalent problem in the form of a serial inventory system in which each demand class is assigned to a different installation. Our approach has the following main features to be noted:

- It provides a more transparent model than any existing literature on inventory models with multiple demand classes,
- Different from the existing literature, it is not limited to only two demand classes; it covers any number of demand classes,
- It provides a flexible choice of objective specification; either assigning a service level requirement or a backordering cost for each demand class,
- It allows lot size replenishment orders,
- It suggests numerical algorithms to determine how much inventory to stock,

- It would be easily extended to multi-echelon models, where there are multiple demand classes at each echelon.

We realize that there are important extensions to the current developed models. First, we view our system as a single location inventory system serving a single product to multiple demand classes. Extending our single location model to a multi-echelon system where each location in the lowest echelon faces multiple demand classes is an interesting problem with great practical value. Second, we assume that any demand that is not filled on its due-date is fully backordered. There may exist applications in which, unfilled orders are simply lost or expedited to outside sources by incurring cost. Hence, there is a practical value to model the lost-sales environment as well. We conjecture that our mapping into serial inventory system would allow us to model the lost sales environment as well. Third, we use a FCFS (first-come-first-served) priority fulfillment to model how backorders are cleared upon arrival of replenishment orders. FCFS priority fulfillment is not necessarily optimal, but it is also intuitively clear that it is not a bad backorder clearing method since it only becomes effective only when demand realization during lead-time becomes really high. It is valuable to provide alternative clearing mechanisms to have benchmarks for FCFS priority fulfillment.

Appendix A

Proofs of Propositions

Proposition 1:

The production of an old product stops permanently whenever a new product is introduced.

Proof. Suppose that optimal allocation policy $\mathbf{x}^* = \{x(t) \in [0, K], t \in [0, T]\}$ is decreasing over a small time interval, i.e., for some $\Delta > 0$ and $[\tau - \Delta, \tau + \Delta] \subseteq [0, T]$, $x(t) = \overline{x}$ for $t \in [\tau - \Delta, \tau]$, $x(t) = \underline{x}$ for $t \in [\tau, \tau + \Delta]$, where $\overline{x} > \underline{x}$. Consider, instead, an alternative policy $\mathbf{x^p}$ that is increasing over the same interval, i.e., $x(t) = \underline{x}$ for $t \in (\tau - \Delta, \tau)$ and $x(t) = \overline{x}$ for $t \in (\tau, \tau + \Delta)$. Consider whether the alternative policy $\mathbf{x^p}$ has any improvement on the total profit over $\mathbf{x}^*$. The difference in profits between the two policies becomes:

$$\begin{aligned}
\Pi(\mathbf{x^p}) - \Pi(\mathbf{x}) = {} & p_1 \cdot \overline{x} \cdot \Delta \cdot [y(Q(\tau - \Delta) + \underline{x} \cdot \Delta) - y(Q(\tau - \Delta))] - \\
& p_1 \cdot \underline{x} \cdot \Delta \cdot [y(Q(\tau - \Delta) + \overline{x} \cdot \Delta) - y(Q(\tau - \Delta))] \\
& > 0 \text{ iff } \frac{y(Q(\tau - \Delta) + \underline{x}.\Delta) - y(Q(\tau - \Delta))}{\underline{x}.\Delta} > \\
& \frac{y(Q(\tau - \Delta) + \overline{x}.\Delta) - y(Q(\tau - \Delta))}{\overline{x}.\Delta}
\end{aligned}$$

This last inequality holds true for any $y(Q(t))$ since it is concave and non-decreasing by assumption. This implies that the alternative policy $\mathbf{x^p}$ results in improvement over the policy $\mathbf{x}^*$; this is a contradiction because $\mathbf{x}^*$ was the optimal policy. Hence, the optimal policy $\mathbf{x}^*$ has a non-decreasing trajectory over time, – i.e., $x(\tau_1) \leq x(\tau_2)$ for $\tau_1 < \tau_2 \in [0, T]$. Consider now a small interval $[\tau - \delta, \tau + \delta] \subseteq [0, T]$ around τ. Suppose that the optimal trajectory over this interval is such that $0 < x(t) < K \ \forall t \in [\tau - \delta, \tau + \delta]$. Then, $x(\tau)$ has to satisfy the following necessary condition for optimality:

$$\frac{\partial \Pi\left(\mathbf{x}^*\right)}{\partial x\left(\tau\right)} = -p_0^l \cdot \frac{\partial\left(\int_0^{T_0} x(t)dt\right)}{\partial x\left(\tau\right)} + p_1 \cdot \frac{\partial\left(\int_0^{T_0} x(t) \cdot y\left(Q\left(t\right)\right) dt\right)}{\partial x\left(\tau\right)} +$$
$$K \cdot p_1 \cdot \frac{\partial\left(\int_{T_0}^{T} y\left(Q\left(t\right)\right) dt\right)}{\partial x\left(\tau\right)} = 0$$

Evaluating the partial derivatives yields the following:

$$p_0^l = p_1 \cdot \left[y(Q(\tau)) + \int_\tau^{T_0} x(t) \cdot y'(Q(t))dt\right] + K \cdot p_1 \cdot \int_{T_0}^{T} y'(Q(t))dt. \quad (i)$$

Moreover, $x(\tau + \delta)$ must also satisfy the necessary condition for optimality. Hence, the following holds true:

$$p_0^l = p_1 \left[y(Q(\tau + \delta)) + \int_{\tau+\delta}^{T_0} x(t) \cdot y'(Q(t))dt\right] + K \cdot p_1 \cdot \int_{T_0}^{T} y'(Q(t))dt. \quad (ii)$$

Subtracting the terms in equation (i) from those in (ii) results in:

$$\int_\tau^{\tau+\delta} x(t).y'(Q(t))dt = y(Q(\tau + \delta)) - y(Q(\tau))$$

Dividing both sides by δ and taking the limit as δ approaches to zero result in the following:

$$x(\tau) \cdot y'(Q(\tau)) = y'(Q(\tau))$$

This expression results in a contraction since $x(\tau)$ is not necessarily equal to 1. And, a similar contradiction can be shown when $x(\tau - \delta) \in (0, K)$ is chosen to satisfy the necessary condition. This implies that the optimal trajectory $x(t) \in \{0, K\} \ \forall t \in [\tau - \delta, \tau + \delta]$. Moreover, it has been shown earlier that optimal trajectory is increasing in time. These two results together imply that there exists $\tau \in [0, T]$ such that the optimal trajectory $x(t) = 0 \, \forall t \in [0, \tau]$ and the optimal trajectory $x(t) = K \, \forall t \in (\tau, T]$. ■

Proposition 2:

Optimal capacity allocation policy for problem SLCAP is of bang-bang type; it is optimal to allocate the whole production capacity for a new product immediately if $\frac{p_0^h}{p_1} \leq \frac{1}{T_0} \cdot \int_0^{T_0} y(Q(t))dt$, otherwise it is optimal not to allocate any production capacity for a new product until the old product becomes obsolete.

Proof. For a given optimal policy $\mathbf{x}$ the result in proposition 1 implies that, there exists $\tau \in [0, T]$ such that $x(t) = 0 \; \forall t \in [0, \tau]$ and $x(t) = K \; \forall t \in (\tau, T]$. Total profit function then becomes:

$$\Pi(\mathbf{x}) = \int_0^{\tau} K \cdot (p_0^h - c)dt + \int_{\tau}^{T} K \cdot (y(Q(t - \tau)) \cdot p_1 - c)dt$$

Total profit function is convex in $\tau \in [0, T_0]$ because $\frac{d^2\Pi(\mathbf{x})}{d\tau^2} = K \cdot p_1 \cdot y'(Q(T - \tau)) \geq 0$ holds true. Since the objective is to maximize total profit, optimal production launch time $\tau^* \in \{0, T_0\}$. If $\tau^* = 0$, then total profit becomes $\int_0^T K \cdot (y(Q(t)) \cdot p_1 - c)dt$. If $\tau^* = T_0$, then total profit becomes $\int_0^{T_0} K \cdot (p_0^h - c)dt + \int_{T_0}^{T} K \cdot (y(Q(t - T_0)) \cdot p_1 - c)dt$. Combining these two expressions implies that $\tau^* = 0$ iff $\int_0^T (y(Q(t)) \cdot p_1 - c)dt \geq \int_0^{T_0} (p_0^h - c)dt + \int_{T_0}^{T} (y(Q(t - T_0)) \cdot p_1 - c)dt$. Further simplification implies that $\tau^* = 0$ iff $\frac{p_0^h}{p_1} \leq \frac{1}{T_0} \cdot \int_0^{T_0} y(Q(t))dt$. Similarly, $\tau^* = T_0$ iff $\frac{p_0^h}{p_1} \geq \frac{1}{T_0} \cdot \int_0^{T_0} y(Q(t))dt$. ■

Proposition 3:

For any $n \geq 2$, $n \in Z^+$ and $\gamma \geq 0.5$ optimal production launch time of a new product $\tau^*(n)$ is non-zero iff $\frac{p_0^l}{p_1} > (1 - \gamma) \cdot y(\frac{K}{n} \cdot T) + \gamma \cdot y(0)$.

Proof. For any $n \geq 2$, $n \in Z^+$ and $\gamma \geq 0.5$, let $\tau^*(n)$ denote the optimal production launch time of a new product. Let us first show that $\Pi(n, \tau(n))$ is strictly concave in $\tau(n)$ for all $n \in Z^+$ and $\gamma \geq 0.5$. For all $n \in Z^+$ the second derivative of $\Pi(n, \tau(n))$ with respect to $\tau(n)$ simply yields the following, $\frac{d^2(\Pi(n,\tau(n)))}{d\tau(n)^2} = (\frac{K}{n})^2 \cdot p_1 \cdot [(1-$

$\gamma)^2 \cdot y'(\frac{K}{n} \cdot (T-(1-\gamma) \cdot \tau(n))) - \gamma^2 \cdot y'(\frac{K}{n} \cdot \gamma \cdot \tau(n))]$, where $y'(\cdot)$ denotes the first derivative of $y(\cdot)$. Since $y(\cdot)$ is concave-increasing and $T - (1-\gamma) \cdot \tau > \gamma \cdot \tau$, it holds true that $y'(\frac{K}{n} \cdot \gamma \cdot \tau(n)) > y'(\frac{K}{n} \cdot (T-(1-\gamma) \cdot \tau(n)))$. Furthermore, $(1-\gamma)^2 \leq \gamma^2$ holds true for $\gamma \geq 0.5$. Hence, $(1-\gamma)^2 \cdot y'(\frac{K}{n} \cdot (T-(1-\gamma) \cdot \tau(n))) < \gamma^2 \cdot y'(\frac{K}{n} \cdot \gamma \cdot \tau(n))$ holds true for $\gamma \geq 0.5$, which implies that $\frac{d^2\Pi(n,\tau(n))}{d\tau(n)^2} < 0$ for $\gamma \geq 0.5$. Hence, $\Pi(n, \tau(n))$ is strictly concave in $\tau(n)$ for all $n \in Z^+$ and $\gamma \geq 0.5$. Let us now show that the sufficient condition($\Rightarrow$) holds; suppose that $\tau^*(n) > 0$. This implies that profit function must be increasing over a small interval from zero, $[0, \epsilon]$, $\epsilon > 0$ because profit function is concave in $\tau(n)$ for all $n \in Z^+$ and $\gamma \geq 0.5$ as shown in previous proposition. Hence, we must have $\lim_{\epsilon \to 0} \frac{d\Pi(n,\tau(n))}{d\tau(n)}|_{\tau(n)=\epsilon} > 0$, which implies that $p_0^l - p_1 \cdot [(1-\gamma) \cdot y(\frac{K}{n} \cdot T) + \gamma \cdot y(0)] > 0$. Second, let us now show that the necessary condition($\Leftarrow$) holds; suppose that $p_0^l - p_1 \cdot [(1-\gamma) \cdot y(Q(T)) + \gamma \cdot y(Q(0))]$ holds true. This implies that $\frac{d\Pi(n,\tau(n))}{d\tau(n)}|_{\tau(n)=0} > 0$ since $\frac{d\Pi(n,\tau(n))}{d\tau(n)} = \frac{K}{n} \cdot [p_0^l - p_1 \cdot [(1-\gamma) \cdot y(\frac{K}{n} \cdot (T - (1-\gamma) \cdot \tau(n))) + \gamma \cdot y(\frac{K}{n} \cdot \gamma \cdot \tau(n))]]$. This implies that profit function must be at least increasing over a small interval, hence, optimal production launch time $\tau^* > 0$. ■

Proposition 4:

For $\gamma \geq 0.5$, $\Pi(n = 2, \tau^*(2)) > \Pi(n = 1)$ iff the price ratio $\frac{p_0^l}{p_1} > \frac{2 \cdot \int_0^T y(K \cdot t)dt - \int_0^T y(\frac{K}{2} \cdot t)dt - \int_{T_0}^T y(\frac{K}{2} \cdot (t-(1-\gamma) \cdot T_0))dt}{T_0}$ and optimal knowledge transfer time $\tau^*(2) = T_0$.

Proof. Let us first show that for concave-increasing $y(Q(\cdot))$, $\gamma \geq 0.5$, and $\tau \in [0, T_0]$, the following inequality holds true: $\frac{\int_\tau^T y(Q(t-(1-\gamma) \cdot \tau))dt}{(T-\tau)} \geq -\frac{\partial}{\partial \tau}(\int_\tau^T y(Q(t-(1-\gamma) \cdot \tau))dt)$. Let the following hold true through change of variable, $\int_\tau^T y(Q(t-(1-\gamma) \cdot \tau))dt = \int_{\gamma \cdot \tau}^{T-(1-\gamma) \cdot \tau} y(Q(t))dt$. $y(Q(t))$ is a concave function in $Q(t)$, hence, the area underneath $y(Q(t))$ over the interval $[T-(1-\gamma) \cdot \tau, \gamma \cdot \tau]$ is at least as large as the area of a trapezoid constructed with end points at the boundaries of the same interval. This implies the following inequality: $\int_{\gamma \cdot \tau}^{T-(1-\gamma) \cdot \tau} y(Q(t))dt \geq \frac{y(Q(T-(1-\gamma) \cdot \tau)) + y(Q(\gamma \cdot \tau))}{2} \cdot (T-\tau)$. Subtracting $((1-\gamma) \cdot y(Q(T-(1-\gamma) \cdot \tau)) + \gamma \cdot y(Q(\gamma \cdot \tau))) \cdot (T-\tau)$ from the right hand side of the above inequality results in: $(\gamma - 0.5) \cdot (y(Q(T-(1-\gamma) \cdot \tau)) - y(Q(\gamma \cdot \tau)))$, that is ≥ 0 since $\gamma \geq 0.5$ and $y(Q(t))$ is increasing in $Q(t)$. Hence, the following inequality holds true: $\frac{y(Q(T-(1-\gamma)\cdot\tau))+y(Q(\gamma\cdot\tau))}{2} \cdot (T-\tau) \geq ((1-\gamma)\cdot y(Q(T-(1-\gamma)\cdot\tau)) + \gamma \cdot y(Q(\gamma\cdot\tau))) \cdot (T-\tau)$ Combining the two inequalities implies that $\int_{\gamma\cdot\tau}^{T-(1-\gamma)\cdot\tau} y(Q(t))dt \geq ((1-\gamma)\cdot y(Q(T-(1-\gamma)\cdot\tau))+\gamma\cdot y(Q(\gamma\cdot\tau)))\cdot(T-\tau)$. The term on RHS is equivalent to $(T-\tau)\cdot(-\frac{\partial}{\partial\tau}(\int_\tau^T y(Q(t-(1-\gamma)\cdot\tau))dt)$. Replacing this expression yields the desired inequality, $\frac{\int_{\gamma\cdot\tau}^{T-(1-\gamma)\cdot\tau} y(Q(t))dt}{(T-\tau)} \geq -\frac{\partial}{\partial\tau}(\int_\tau^T y(Q(t-(1-\gamma)\cdot\tau))dt$. After showing this inequality holds true, let us now proof the statement in this proposition holds true. First, let us show that sufficient condition($\Rightarrow$) holds true. Suppose that $\Pi(n=2,\tau^*(2)) > \Pi(n=1)$ holds true. Then, the following inequality simply follows, $\frac{p_0^l}{p_1}\cdot\tau^*(2) > 2\cdot\int_0^T y(K\cdot t)dt - \int_0^T y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt$ (1) Through change of variable the following equality holds true: $2\cdot\int_0^T y(K\cdot t)dt - \int_0^T y(\frac{K}{2}\cdot t)dt = \int_T^{2T} y(\frac{K}{2}\cdot t)dt$. Replacing this equality in (1) results in, $\frac{p_0^l}{p_1}\cdot\tau^*(2) > \int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt$ (2). $RHS > 0$ and $LHS = 0$ in limit when $\tau^* \to 0$. This implies that $\tau^* > 0$ because $LHS > RHS$. Hence, we can divide both sides in (2) by $\tau^*(2)$, $\frac{p_0^l}{p_1} > \frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{\tau^*(2)}$ (3). Consider $\tau^*(2) \in (0, T_0]$, the following inequality holds true because $y(\cdot)$ is an increasing function $\frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt}{T} \geq \frac{\int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{T-\tau^*(2)}$. Multiplying both sides of the previous equality by $\frac{T}{\tau^*}$ and taking some terms to LHS produce the following $\frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{\tau^*(2)} \geq \frac{\int_{\tau^*}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{T-\tau^*(2)}$. Using the result in Lemma 1 on RHS of the previous inequality implies the following: $\frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{\tau^*(2)} \geq -\frac{\partial}{\partial\tau}(\int_\tau^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau))dt)|_{\tau=\tau^*}$ (4). Combining inequalities in (3) and (4) yields that $\frac{p_0^l}{p_1} > \frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{\tau^*(2)}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau^*(2)))dt}{\tau^*(2)} \geq -\frac{\partial}{\partial\tau}(\int_\tau^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau))dt)|_{\tau=\tau^*}$ (5). Moreover, $\Pi(n=2,\tau(2))$ is concave in $\tau(2)$ for $\gamma \geq 0.5$, hence, optimal $\tau^*(2) \in (0, T_0]$ must satisfy first order optimality condition. Namely, $\frac{p_0^l}{p_1} = -\frac{\partial}{\partial\tau}(\int_\tau^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau))dt)|_{\tau=\tau^*(2)}$ for $\tau^*(2) \in (0, T_0)$, and $\frac{p_0^l}{p_1} > -\frac{\partial}{\partial\tau}(\int_\tau^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot\tau))dt)|_{\tau=T_0}$ for $\tau^*(2) = T_0$. Combining these first order optimality conditions with the inequality in (5) yields that $\tau^*(2) = T_0$ and $\frac{p_0^l}{p_1} > \frac{\int_T^{2T} y(\frac{K}{2}\cdot t)dt - \int_{T_0}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot T_0))dt}{T_0}$. Second, let us show that necessary condition($\Leftarrow$) holds true. Suppose that $\tau^*(2) = T_0$ and the following inequality holds true $\frac{p_0^l}{p_1} > \frac{2\cdot\int_0^T y(K\cdot t)dt - \int_0^T y(\frac{K}{2}\cdot t)dt - \int_{T_0}^T y(\frac{K}{2}\cdot(t-(1-\gamma)\cdot T_0))dt}{T_0}$. Multiplying both sides by $K\cdot p_1\cdot T_0$, subtracting $K\cdot T\cdot c$ from both sides, collecting some

terms of RHS on LHS, and replacing T_0 by $\tau^*(2)$ yield that $\int_0^T K \cdot [p_1 \cdot y(\frac{K}{2} \cdot (t - (1-\gamma) \cdot \tau^*(2))) - c]dt + \int_0^{\tau^*(2)} \frac{K}{2} \cdot (p_0^l - c)dt +$
$\int_{\tau^*(2)}^T \frac{K}{2} \cdot [p_1 \cdot y(\frac{K}{2} \cdot (t - (1-\gamma) \cdot \tau^*(2))) - c]dt > \int_0^T K \cdot [p_1 \cdot y(K \cdot t) - c]dt$, which simply means $\Pi(n=2, \tau^*(2)) > \Pi(n=1)$. ■

Proposition 5:

If $\gamma \geq 0.5$ and $\Pi(n=2, \tau^*(2)) > \Pi(n=1)$ hold true, then for all $n \geq 2$, $n \in Z^+$ total profit $\Pi(n, \tau^*(n))$ can be divided into two separate parts; profit obtained from old product $\Pi_{\text{Old}}(n, \tau^*(n))$ and profit obtained from new product $\Pi_{\text{New}}(n, \tau^*(n))$ such that

i) $\Pi_{\text{Old}}(n, \tau^*(n))$ is increasing and discrete-concave in n,

ii)$\Pi_{\text{New}}(n, \tau(n))$ is decreasing and discrete-convex in n.

Proof. Suppose that $\gamma \geq 0.5$ and $\Pi(n=2, \tau^*(2)) > \Pi(n=1)$ hold true. Then, an earlier proposition implies that $\tau^*(2) = T_0$. Let us first show that $\tau^*(n)$ is increasing in $n \in Z^+$ for $\gamma \geq 0.5$. $\Pi(n, \tau(n))$ is strictly concave in $\tau(n)$ for all $n \in Z^+$ and $\gamma \geq 0.5$. Hence, optimal $\tau^*(n) \in [0, T_0]$ must satisfy the following first order condition: $\frac{p_0^l}{p_1} = (1-\gamma) \cdot y(\frac{K}{n} \cdot (T - (1-\gamma) \cdot \tau^*(n))) + \gamma \cdot y(\frac{K}{n} \cdot \gamma \cdot \tau^*(n))$. Let RHS be denoted by $G(\tau^*(n), n)$. Since $G(\tau^*(n), n)$ equals to a constant and $\tau^*(n)$ is a function of n, we can take the implicit derivative of $\tau^*(n)$ with respect to n: $\frac{d\tau^*(n)}{dn} = -\frac{\frac{\partial G(\tau^*(n),n)}{\partial n}}{\frac{\partial G(\tau^*(n),n)}{\partial \tau^*(n)}}$. Evaluating the expressions results in, $\frac{d\tau^*(n)}{dn} = \frac{1}{n} \cdot \frac{(T-(1-\gamma)\cdot\tau^*(n))\cdot(1-\gamma)\cdot y'(\frac{K}{n}\cdot(T-(1-\gamma)\cdot\tau^*(n)))+\gamma^2\cdot\tau^*(n)\cdot y'(\frac{K}{n}\cdot\gamma\cdot\tau^*(n))}{-(1-\gamma)^2\cdot y'(\frac{K}{n}\cdot(T-(1-\gamma)\cdot\tau^*(n)))+\gamma^2\cdot y'(\frac{K}{n}\cdot\gamma\cdot\tau^*(n))}$. The numerator is > 0 because $y'(\cdot) > 0$. Also, the denominator is > 0 since $y'(\frac{K}{n} \cdot (T - (1-\gamma) \cdot \tau^*(n))) < y'(\frac{K}{n} \cdot \gamma \cdot \tau^*(n))$ and $(1-\gamma)^2 \leq \gamma^2$ for $\gamma \geq 0.5$. Hence, $\frac{d\tau^*(n)}{dn} > 0$ holds true. This implies that $\tau^*(n)$ is increasing in $n \in Z^+$. This result implies that $\tau^*(n) = T_0$ holds true for all $n \geq 2$, $n \in Z^+$. Total profit $\Pi(n, \tau^*(n) = T_0)$ can be divided into two exclusive parts; total profit from the old product and total profit from the new product. Let total profit obtained from old product be $\Pi_{\text{Old}}(n, \tau^*(n) = T_0)$. $(n-1)$ production lines with production rate $\frac{K}{n}$ on each of them produce the old product over $[0, \tau^*(n) = T_0]$. Hence, $\Pi_{\text{Old}}(n, \tau^*(n) = T_0) = (n-1) \cdot \frac{K}{n} \cdot (p_0^l - c) \cdot T_0$. Let total profit obtained from new product be $\Pi_{\text{New}}(n, \tau^*(n) = T_0)$. Similarly, these $(n-1)$ production lines produce the new product over $[\tau^*(n) = T_0, T]$ plus the pilot production line with production rate $\frac{K}{n}$ produces the new product over $[0, T]$. Hence, $\Pi_{\text{New}}(n, \tau^*(n) = T_0) = (n-1) \cdot \int_{T_0}^{T} \frac{K}{n} \cdot [p_1 \cdot y(\frac{K}{n} \cdot (t-(1-\gamma) \cdot T_0)) - c]dt + \int_0^T \frac{K}{n} \cdot [p_1 \cdot y(\frac{K}{n} \cdot t) - c]dt$. Let us first show that $\Pi_{\text{Old}}(n, \tau^*(n) = T_0)$ is increasing and discrete-concave in n. Let $\Delta\Pi_{\text{Old}}(n, \tau^*(n) = T_0) = \Pi_{\text{Old}}(n, \tau^*(n) = T_0) - \Pi_{\text{Old}}(n-1, \tau^*(n-1) = T_0)$. Then, $\Delta\Pi_{\text{Old}}(n, \tau^*(n) = T_0) = \frac{1}{n(n-1)} \cdot K \cdot (p_0^l - c) \cdot T_0 > 0$, hence $\Pi_{\text{Old}}(n, \tau^*(n) = T_0)$ is increasing in n. Consider now the forward difference $\Delta\Pi_{\text{Old}}(n+1, \tau^*(n+1) = T_0) - \Delta\Pi_{\text{Old}}(n, \tau^*(n) = T_0) = -\frac{2}{n(n^2-1)} \cdot K \cdot (p_0^l - c) \cdot T_0 < 0$, which implies that $\Pi_{\text{Old}}(n, \tau^*(n) = T_0)$ is discrete-concave in n. Similarly, $\Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0) = \Pi_{\text{New}}(n, \tau^*(n) = T_0) - \Pi_{\text{New}}(n-1, \tau^*(n-1) = T_0)$. We can easily show that $\frac{d\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)}{dT_0} < 0$. This implies that $\Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0) \leq \Delta\Pi_{\text{New}}(n, \tau^*(n) = 0)$. Moreover, it is easy to show that $\Delta\Pi_{\text{New}}(n, \tau^*(n) = 0) < 0$. Hence, $\Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0) < 0$ holds true, hence $\Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0)$ is decreasing in n. Consider now the forward difference $\Delta\Pi_{\text{New}}(n+1, \tau^*(n+1) = T_0) - \Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0)$. Let us now show that $\Delta\Pi_{\text{New}}(n+1, \tau^*(n+1) = T_0) - \Delta\Pi_{\text{New}}(n, \tau^*(n) = T_0) > 0$ for all $n \geq 2$. The following holds true for all $n \geq 2$ because $y(\cdot)$ is increasing and concave:

$$[y(\frac{K}{n-1} \cdot t) - y(\frac{K}{n} \cdot t)] - [y(\frac{K}{n} \cdot t) - y(\frac{K}{n+1} \cdot t)] > 0$$

Moreover, $y(\cdot)$ is a continuous function on $[\gamma \cdot T_0, T - (1-\gamma) \cdot T_0]$, hence, it is integrable over the same interval. This implies that

$$\int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} [y(\frac{K}{n-1} \cdot t) - y(\frac{K}{n} \cdot t)]dt - \int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} [y(\frac{K}{n} \cdot t) - y(\frac{K}{n+1} \cdot t)]dt > 0$$

which is simply equivalent to:

$$\frac{1}{n+1} \cdot \int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1} \cdot t)dt - \frac{2}{n} \cdot \int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n} \cdot t)dt + \frac{1}{n-1} \cdot$$
$$\int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1} \cdot t)dt > \frac{2(n-1)}{n} \cdot \int_{\gamma \cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n} \cdot t)dt - \frac{n}{n+1} \cdot$$

$$\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)dt - \frac{n-2}{n-1}\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)dt$$

Derivative of LHS with respect to γ yields that $-[\frac{1}{n+1}\cdot y(\frac{K\cdot(T-(1-\gamma)\cdot T_0)}{n+1}) - \frac{2}{n}\cdot y(\frac{K\cdot(T-(1-\gamma)\cdot T_0)}{n}) + \frac{1}{n-1}\cdot y(\frac{K\cdot(T-(1-\gamma)\cdot T_0)}{n-1})] < -[\frac{1}{n}\cdot y(\frac{K}{n+1}\cdot t) + \frac{1}{n}\cdot y(\frac{K}{n-1}\cdot t) - \frac{2}{n}\cdot y(\frac{K\cdot(T-(1-\gamma)\cdot T_0)}{n})] = -\frac{1}{n}\cdot[y(\frac{K}{n+1}\cdot t) + y(\frac{K}{n-1}\cdot t) - 2\cdot y(\frac{K\cdot(T-(1-\gamma)\cdot T_0)}{n})] < 0$. Hence, LHS is decreasing in $\gamma\in[0,1]$. This implies that $LHS(\gamma=0)\geq LHS$, hence, $\frac{1}{n+1}\cdot\int_0^{T-T_0} y(\frac{K}{n+1}\cdot t)dt - \frac{2}{n}\cdot\int_0^{T-T_0} y(\frac{K}{n}\cdot t)dt + \frac{1}{n-1}\cdot\int_0^{T-T_0} y(\frac{K}{n-1}\cdot t)dt \geq LHS$. Moreover, RHS is decreasing in T_0, hence $\frac{1}{n+1}\cdot\int_0^{T} y(\frac{K}{n+1}\cdot t)dt - \frac{2}{n}\cdot\int_0^{T} y(\frac{K}{n}\cdot t)dt + \frac{1}{n-1}\cdot\int_0^{T} y(\frac{K}{n-1}\cdot t)dt \geq LHS$ holds true. Taking LHS to the other side yields that: $\frac{1}{n+1}\cdot\int_0^{T} y(\frac{K}{n+1}\cdot t)dt + \frac{n}{n+1}\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)dt - \frac{2}{n}\cdot\int_0^{T-T_0} y(\frac{K}{n}\cdot t)dt - \frac{2(n-1)}{n}\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt + \frac{1}{n-1}\cdot\int_0^{T} y(\frac{K}{n-1}\cdot t)dt + \frac{n-2}{n-1}\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)dt > 0$. Moreover, multiplying both sides by $K\cdot p_1$ preserves the inequality and yields $\Delta\Pi_{\text{New}}(n+1,\tau^*(n+1)=T_0) - \Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0) > 0$. This implies that $\Pi_{\text{New}}(n,\tau^*(n)=T_0)$ is discrete-convex in n. ∎

Proposition 6:

If $\gamma\geq 0.5$ and $\Pi(n=2,\tau^*(2)) > \Pi(n=1)$ hold true, then

i) total profit $\Pi(n,\tau^*(n))$ is unimodal in n, $n\in Z^+$,

ii) total profit $\Pi(n,\tau^*(n))$ is increasing for all $n < n^*\in Z^+$,

ii) total profit $\Pi(n,\tau^*(n))$ is decreasing for all $n > n^*\in Z^+$.

Proof. We begin with showing *i)*. Suppose that $\gamma\geq 0.5$ and $\Pi(n=2,\tau^*(2)) > \Pi(n=1)$ hold true. It is shown earlier that $\tau^*(2)=T_0$ and $\tau^*(n)$ is increasing in $n\in Z^+$. Hence, $\tau^*(n)=T_0$ holds true for all $n\geq 2, n\in Z^+$. Consider $\Delta\Pi(n,\tau^*(n)) = \Pi(n,\tau^*(n)=T_0) - \Pi(n-1,\tau^*(n-1)=T_0)$. More explicitly $\Delta\Pi(n,\tau^*(n)) = [K\cdot p_0^l\cdot T_0 + \Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)]\cdot\frac{1}{n\cdot(n-1)}$. To show unimodality of $\Pi(n,\tau^*(n)=T_0)$, it is enough to show that it has at most one change in its sign. Given its explicit form, it holds true that $\Delta\Pi(n,\tau^*(n))$ has at most one change in its sign if $\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)$ is monotone in n. Let the forward difference of $-\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)$ be represented by $R(n) = n\cdot[-\Delta\Pi_{\text{New}}(n+1,\tau^*(n+$

$1) = T_0)\cdot(n+1)+\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot(n-1)]$. Let us now show that $R(n)>0$ for all $n\geq 2, n\in Z^+$. First, $y(\cdot)$ is increasing and concave, hence the following inequality holds true for all $n\geq 2, n\in Z^+$ and $t\in[0,T]$: $\frac{n}{n+1}\cdot y(\frac{K}{n}\cdot t)-\frac{n-1}{n+1}\cdot y(\frac{K}{n-1}\cdot t) > y(\frac{K}{n+1}\cdot t)-\frac{n}{n+1}\cdot y(\frac{K}{n}\cdot t)$. Collecting terms on LHS and then multiplying both sides by $(n+1)$ yield $2\cdot n\cdot y(\frac{K}{n}\cdot t)-(n-1)\cdot y(\frac{K}{n-1}\cdot t)-(n+1)\cdot y(\frac{K}{n+1}\cdot t)>0$. Since LHS is a continuous function of t on $[\gamma\cdot T_0, T-(1-\gamma)\cdot T_0]$, it is integrable over the same interval. This implies that $2\cdot n\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt-(n-1)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)-(n+1)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)>0$. Rearranging terms yields, $2\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt-\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)-\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)+2\cdot(n-1)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt-n\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)-(n-2)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)>0$ the sum of the first three terms on LHS obtains its highest value at $\gamma=1$, hence, $2\cdot\int_{T_0}^{T} y(\frac{K}{n}\cdot t)dt-\int_{T_0}^{T} y(\frac{K}{n+1}\cdot t)-\int_{T_0}^{T} y(\frac{K}{n-1}\cdot t)+2\cdot(n-1)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt-n\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)-(n-2)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)>0$ holds true. Moreover, both the sum of the last three terms and the sum of all terms on LHS are decreasing in T_0. Hence, replacing $T_0=0$ for the first three terms still preserves the inequality: $2\cdot\int_{0}^{T} y(\frac{K}{n}\cdot t)dt-\int_{0}^{T} y(\frac{K}{n+1}\cdot t)-\int_{0}^{T} y(\frac{K}{n-1}\cdot t)$ $+2\cdot(n-1)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n}\cdot t)dt-n\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n+1}\cdot t)-(n-2)\cdot\int_{\gamma\cdot T_0}^{T-(1-\gamma)\cdot T_0} y(\frac{K}{n-1}\cdot t)>0$. Multiplying both sides by $n\cdot K\cdot p_1$ yields that $R(n)>0$, for all $n\geq 2, n\in Z^+$. This implies that $-\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)$ is increasing in $n\geq 2, n\in Z^+$. This further implies that $\Delta\Pi(n,\tau^*(n)=T_0)$ has at most one change in its sign, hence, $\Pi(n,\tau^*(n)=T_0)$ is unimodal in $n\in Z^+$.

Next, we show that *ii)* and *iii)* hold true. Suppose that $\gamma\geq 0.5$ and $\Pi(n=2,\tau^*(2))>\Pi(n=1)$ hold true. It is shown earlier that $\tau^*(2)=T_0$ and $\tau^*(n)$ is increasing in $n\in Z^+$. Hence, $\tau^*(n)=T_0$ holds true for all $n\geq 2$, $n\in Z^+$. Consider $\Delta\Pi(n,\tau^*(n)=T_0)$. It is easy to show that $\Delta\Pi(n,\tau^*(n)=T_0)>0$ iff $K\cdot p_0^l\cdot T_0>-\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)$ for all $n\geq 2, n\in Z^+$. Moreover, we showed earlier that $-\Delta\Pi_{\text{New}}(n,\tau^*(n)=T_0)\cdot n\cdot(n-1)$ is increasing in $n\geq 2, n\in Z^+$. This implies that there exists an $n^*\in Z^+$ such that $\Delta\Pi(n,\tau^*(n)=T_0)>0$ on $[1,n^*)$. Hence, $\Pi(n,\tau^*(n))$ is increasing for all $n<n^*, n\in Z^+$. Similarly, $\Delta\Pi(n,\tau^*(n)=T_0)<0$ for all $n>n^*, n\in Z^+$, hence, $\Pi(n,\tau^*(n))$ is decreasing for all $n>n^*, n\in Z^+$. ■

Proposition 7:

If $E[Y_h(Q(t),\hat{\epsilon})]$ is increasing and discrete-concave in $m \in Z^+$ and similarity index between pilot production lines and the remaining production lines $\gamma \geq 0.5$, then the total expected profit $E[\Pi(\tau_m(n), m)]$ is discrete-concave for all $m < n \in Z^+$.

Proof. Suppose that $E[Y_h(Q(t),\hat{\epsilon})]$ is increasing and discrete-concave in $m \in Z^+$ and similarity index $\gamma \geq 0.5$. Let $m, n \in Z^+$ such that $m < n$. $E[\Pi(\tau_m(n), m)]$ has the following expression:

$$\begin{aligned} E[\Pi(\tau_m(n), m)] &= m \cdot \int_0^T K \cdot p_1 \cdot y(K \cdot t)dt + (n-m) \cdot [K \cdot p_0^l \cdot \tau_m(n) + \\ &\int_{\gamma \cdot \tau_m(n)}^{T-(1-\gamma)\cdot\tau_m(n)} K \cdot p_1 \cdot y(K \cdot t)dt] + K \cdot p_1 \cdot \frac{m-1}{m+1} \cdot u \cdot \\ &(T - \tau_m(n)) \cdot [m + \gamma \cdot (n-m)] \end{aligned}$$

Let $E[\hat{\Pi}(\tau_m(n), m)]$ be the interpolation of the function $E[\Pi(\tau_m(n), m)]$ on the real domain, $R^+ \to R^+$. The function $E[\hat{\Pi}(\tau_m(n), m)]$ is continuous and differentiable on $[0, \infty)$. We observe by differentiation that $\frac{d^2 E[\hat{\Pi}(\tau_m(n),m)]}{d\tau_m(n)^2} = -(n-m) \cdot K^2 \cdot p_1 \cdot [-(1-\gamma)^2 \cdot y'(K(T-(1-\gamma)\cdot\tau_m(n))) + \gamma^2 \cdot y(K \cdot \gamma \cdot \tau_m(n))]$, which is negative for $\gamma \geq 0.5$. Hence, $E[\hat{\Pi}(\tau_m(n), m)]$ is strictly concave in $m \in [0, \infty)$. Applying the Envelope Theorem implies that the derivative of $E[\hat{\Pi}(\tau_m(n), m)]$ with respect to m is given by the partial derivative of $E[\hat{\Pi}(\tau_m(n), m)]$ with respect to m, holding $\tau_m(n)$ fixed at the optimal choice, $\frac{dE[\hat{\Pi}(\tau_m(n),m)]}{dm} = \frac{\partial E[\hat{\Pi}(\tau_m(n),m)]}{\partial m}|_{\tau_m(n)=\tau_m^*(n)}$. Hence, $\frac{dE[\hat{\Pi}(\tau_m(n),m)]}{dm} = -K \cdot p_0^l \cdot \tau_m^*(n) + K \cdot p_1 \cdot [\int_0^T y(K \cdot t)dt - \int_{\gamma\cdot\tau_m^*(n)}^{T-(1-\gamma)\cdot\tau_m^*(n)} y(K \cdot t)dt + \frac{2\cdot u}{(m+1)^2} \cdot (m + (n-m)\cdot\gamma) \cdot (T - \tau_m^*(n)) + \frac{m-1}{m+1} \cdot u \cdot (1-\gamma) \cdot (T - \tau_m^*(n))]$.

Derivative of this last expression with respect to m yields that $\frac{d^2 E[\hat{\Pi}(\tau_m(n),m)]}{dm^2} = -4 \cdot K \cdot p_1 \cdot \frac{(n+1)\cdot\gamma-1}{(m+1)^3}$, which is negative. Thus, $E[\hat{\Pi}(\tau_m(n), m)]$ is concave in $m \in R^+$. Since $E[\hat{\Pi}(\tau_m(n), m)]$ is both continuous and concave in real domain, it implies that $E[\Pi(\tau_m(n), m)]$ is discrete-concave in $m \in Z^+$. ■

Proposition 8:

Given $\Delta Y(t) = y_1(Q(t)) - y_2(Q(t), N, \tau) \geqslant 0, t \in [0, T]$ under control policy (N, τ) and given $h = p_1$. Then, (YCP) is equivalent to the following minimization problem:

$$(YCP/E) \quad \text{Min} \quad N \cdot F + \int_0^T h \cdot \Delta Y(t) dt$$

$$s.t \qquad N \geq 0, N \in \mathbf{Z} \text{ and } 0 \leq \tau_i < \tau_{i+1} \leq T, \;\; i \in \{|1, N-1|\}.$$

Proof. The objective function in (YCP) can be reformulated as follows:

$$-N \cdot F + 2p_1 \cdot \int_0^T y_1(t)dt - p_1 \cdot \int_0^T [y_1(t) - y_2(t, (N, \tau))]dt - 2T \cdot c$$

Considering the negative of the above expression and eliminating the terms independent of the control policy (N, τ) imply the following equivalent problem for (YCP) where $y_2(t, (N, \tau))$ indicates that the yield level in production line 2 at time t that depends on control policy (N, τ):

$$\text{Min} \quad N \cdot F + p_1 \int_0^T [y_1(t) - y_2(t, (N, \tau))]dt$$

$$s.t \qquad N \geq 0, N \in \mathbf{Z} \text{ and } 0 \leq \tau_i < \tau_{i+1} \leq T, \;\; i \in \{|1, N-1|\}$$

Substituting $\Delta Y(t) = y_1(t) - y_2(t, (N, \tau))$ and $h = p_1$ produce the following expression:

$$\text{Min} \quad N \cdot F + h \int_0^T \Delta Y(t) dt$$

$$s.t \qquad N \geq 0, N \in \mathbf{Z} \text{ and } 0 \leq \tau_i < \tau_{i+1} \leq T, \;\; i \in \{|1, N-1|\}$$

■

Proposition 9:

Under control policy (N, τ), where $N \geq 0, N \in \mathbf{Z}$ and $\tau = (\tau_1, \tau_2, ..., \tau_N)$ such that $0 \leq \tau_i < \tau_{i+1} \leq T$, for $i \in \{|1, N-1|\}$, if $\Delta Y(t) = y_1(Q(t)) - y_2(Q(t), N, \tau)$ is real-valued continuous function on $[0, T]$ such that $\frac{d\Delta Y(t)}{dt} \geqslant 0$ and $\frac{d^2 \Delta Y(t)}{dt^2} < 0$ for $t \in [0, T]$, then the objective function in (YCP/E) can be replaced by the following expression, $C(N, \tau)$:

$$C(N, \tau) = N \cdot F + \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot [\Delta Y(t) - \Delta Y(\tau_i)] dt.$$

Proof. Let Δ_i denote the increase in yield level in production line 2 after knowledge transfer at time τ_i. At each knowledge transfer at time τ_i, $i \in \{|1, N|\}$, the yield function in production line 2 is shifted up by an amount $y_2(\tau_i) - y_1(\tau_i)$. Hence, we have $\Delta_1 = y_1(\tau_1) - y_2(\tau_1)$ and $\Delta_i = [y_1(\tau_i) - y_2(\tau_i)] - [y_1(\tau_{i-1}) - y_2(\tau_{i-1})]$, $i \in \{2, ..., N\}$. Defining $k = max\{j : \in \tau_j < t\}$,then the cumulative increase in yield level in production line 2 by time $t \in [0, T]$ becomes $\sum_{i=1}^{k} \Delta_i$. Hence, the yield difference at time t under control policy (N, τ) can be represented by $\Delta Y(t) = y_1(t) - y_2(t) - \sum_{i=1}^{k} \Delta_i$, $t \in [0, T]$. Then, for $N \geq 0, N \in \mathbf{Z}$ and $0 \leq \tau_i < \tau_{i+1} \leq T$, $i \in \{|1, N-1|\}$

$$\begin{aligned}
\int_0^T h \cdot \Delta Y(t) dt &= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot \Delta Y(t) dt, \ \tau_0 = 0 \text{ and } \tau_{N+1} = T, \\
&= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot [y_1(t) - y_2(t) - \sum_{i=1}^{k} \Delta_i] dt, \\
&= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} h \cdot [(y_1(t) - y_2(t)) - (y_1(\tau_i) - y_2(\tau_i))] dt,
\end{aligned}$$

Replacing this new expression in the objective function of (YCP/E), the following equivalent objective function, $C(N, \tau)$ is obtained,

$$C(N, \tau) = N \cdot F + h \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} [\Delta Y(t) - \Delta Y(\tau_i)] dt$$

■

Proposition 10:

Optimal number of knowledge-transfer events $N \geq 0, N \in \mathbf{Z}$ in (YCP/E) is bounded by the following expression:

$$N \leqslant N_{\max} = \frac{\int_0^T h \cdot \Delta Y(t)dt}{F}.$$

Proof. Let $C(N = 0, \tau)$ represent the total cost of yield discrepancy when there is no knowledge transfer between the two lines. And also, let $C(N = \infty, \tau)$ represent the total cost of yield discrepancy when there is a continuous knowledge transfer between the two lines over the period $[0, T]$. If $N = 0$, then $C(N = 0, \tau) = \int_0^T h \cdot \Delta Y(t)dt$. If $N = \infty$, then the yield level in production line 2 exactly follows the yield level in production line 1. Hence no opportunity cost exists, $C(N = \infty, \tau) = 0$. The cost of maximum number of meetings achievable should be less than or equal to the maximum benefit obtained by decreasing the total cost of yield discrepancy, — i.e., $N \cdot F \leqslant C(N = \infty, \tau) - C(N = 0, \tau)$. Hence, the following relation holds true,

$$N \leqslant N_{\max} = \frac{\int_0^T h \cdot \Delta Y(t)dt}{F}$$

■

Proposition 11:

Suppose that:

i) $y(t) = y_1(Q(t)) - y_2(Q(t))$ is a piecewise-linear-concave function, that is piecewise continuous and bounded over $[0, T]$; and

ii) $\{f_1, f_2, ..., f_m\}$ is the set of distinct components of $y(t)$ such that $y(t) = \max\{f_1, f_2, ..., f_m\}$ and $0 \leq f'_{i+1} \leq f'_i, i \in [|1, N-1|]$.

If $\frac{f'_1}{f'_m} \leq 1/\cos(\frac{\pi}{N_{\max}+1})^2$, then there exists a unique solution for (YCP/E).

Proof. Suppose that $f'_1 \leq f'_m \cdot 1/\cos(\frac{\pi}{N_{\max}+1})^2$. Since $N \leq N_{\max}$, $\cos(\frac{\pi}{N_{\max}+1}) \leq \cos(\frac{\pi}{N+1})$ holds true, then $f'_1 \leq f'_m \cdot 1/\cos(\frac{\pi}{N_{\max}+1})^2 \Longrightarrow f'_1 \leq f'_m \cdot 1/\cos(\frac{\pi}{N_{\max}+1})^2$. Moreover, $0 \leq f'_{i+1} \leq f'_i, i \in [|1, N-1|] \Longrightarrow f'_1 \leq y'(\tau_i) \leq f'_m$, $i \in [|1, N|]$. Hence,

$f_1' \leq f_m' \cdot 1/\cos(\frac{\pi}{N+1})^2 \Longrightarrow y'(\tau_{i-1}) \leq y'(\tau_i) \cdot 1/\cos(\frac{\pi}{N+1})^2$. Let $H = (h_{i,j})$ denote the Hessian matrix of the objective function in (YCP/E). The elements of the Hessian matrix when $y(t)$ is piecewise linear concave become as follows: $h_{i,\,i} = 2 \cdot y'(\tau_i)$, $h_{i,\,i-1} = -y'(\tau_{i-1})$, $h_{i,\,i+1} = -y'(\tau_i)$, $h_{1,\,0} = h_{N,\,N+1} = 0$, $i \in [|1, N|]$. Hence, H is a tridiagonal $N \times N$ real matrix with positive diagonal entries since $y'(t) > 0$, $t \in [0, T]$. It is shown in Johnson and company in 1996 that $h_{i,\,i-1} \cdot h_{i-1,\,i} \leq \frac{1}{4} \cdot h_{i,\,i} \cdot h_{i-1,\,i-1} \cdot 1/\cos(\frac{\pi}{N+1})^2, i \in [|2, N|]$ implies $det(H) > 0$ when H is a tridiagonal matrix with positive diagonal entries. This condition is satisfied by the Hessian matrix $H = (h_{i,j})$ of the objective function in (YCP/E) and inherited by all the principal submatrices of H. This implies that all principal minors of H are positive, hence the Hessian matrix H is positive definite. Hence, the objective function in (YCP/E) is strictly jointly convex in decision variables, $\{\tau_1, \tau_2, ..., \tau_N\}$. Hence, there exists a unique solution for (YCP/E). ■

Proposition 12

For $s_1 \in Z^+ \cup \{0\}, s_2 \in Z$, the following inequality holds true,

$$P(B_2(s_2) > 0) > P(B_1(s_2 + 1, s_1 - 1) > 0)$$

Proof. Let us first consider $P(B_2(s_2) > 0)$,

$$\begin{aligned}
P(B_2(s_2) > 0) &= \sum_{i=1}^{\infty} P(D_1^L + D_2^L = s_2 + i) \\
&= \sum_{i=1}^{\infty} P(D_1^L + D_2^L = s_2 + 1 + i) + P(D_1^L + D_2^L = s_2 + 1) \\
&= P(B_2(s_2 + 1) > 0) + P(D_1^L + D_2^L = s_2 + 1)
\end{aligned}$$

Let us now consider $P(B_1(s_2 + 1, s_1 - 1) > 0)$,

$$P(B_1(s_2 + 1, s_1 - 1) > 0) = P(B_{2,1}(s_2 + 1) > s_1 - 1)$$

By definition, for $i \in Z^+ \cup \{0\}$, $P(B_2(s_2+1)=i) \geq P(B_{2,1}(s_2+1)=i)$ holds true. This implies $P(B_2(s_2+1)>0) \geq P(B_{2,1}(s_2+1)>s_1-1)$, for any $s_1 \in Z^+, s_2 \in Z$. Hence, the following desired result $P(B_2(s_2+1)>0)+P(D_1^L+D_2^L=s_2+1) > P(B_{2,1}(s_2+1)>s_1)$ holds true. ∎

Proposition 13

If $\widehat{s}_1(s_2) \in Z^+$ denotes the optimal stockage level at installation 1 for a given $s_2 \in Z$ stockage level at installation 2, then the following holds true

$$\widehat{s}_1(s_2+1) \in \{\widehat{s}_1(s_2)-1, \widehat{s}_1(s_2)\}$$

Proof. Suppose that $\widehat{s}_1(s_2)$ denotes the optimal stockage level at installation for a given stockage level at installation 2, s_2. Hence, $\widehat{s}_1(s_2)$ satisfies the following $\widehat{s}_1(s_2)=\min\{s_1|s_2 : P(IL_1(s_2,s_1)>0) \geq \beta_1\}$, which implies that $P(IL_1(s_2,\widehat{s}_1(s_2))>0) \geq \beta_1$. Increasing stockage level at installation 2 increases the fill-rate at installation 1, hence, $P(IL_1(s_2,\widehat{s}_1(s_2))>0) \leq P(IL_1(s_2+1,\widehat{s}_1(s_2))>0)$, which implies that $\widehat{s}_1(s_2+1) \leq \widehat{s}_1(s_2)$. Suppose now that $\widehat{s}_1(s_2+1)+1 < \widehat{s}_1(s_2)$ holds true. This implies that $P(IL_1(s_2,\widehat{s}_1(s_2+1)+1)>0) < \beta_1$. However, $P(IL_1(s_2+1,\widehat{s}_1(s_2+1))>0) \geq \beta_1$ holds true. Furthermore, moving one unit from installation 2 to installation 1 results in a fill-rate at installation 1 that is at least as large as before the move, i.e, $P(IL_1(s_2,\widehat{s}_1(s_2+1)+1)>0) \geq P(IL_1(s_2+1,\widehat{s}_1(s_2+1))>0) \geq \beta_1$. This implies that $P(IL_1(s_2,\widehat{s}_1(s_2+1)+1)>0) \geq \beta_1$. This result is a contradiction. Hence, $\widehat{s}_1(s_2+1)+1 \geq \widehat{s}_1(s_2)$ holds true. Since we showed that $\widehat{s}_1(s_2+1) \leq \widehat{s}_1(s_2)$ and $\widehat{s}_1(s_2+1)+1 \geq \widehat{s}_1(s_2)$, these results together imply that $\widehat{s}_1(s_2+1) \in \{\widehat{s}_1(s_2)-1, \widehat{s}_1(s_2)\}$.
∎

Proposition 14

If the solution for Single-Pass-Algorithm is increased by one unit at some installation $i \in [|1,N|]$, i.e, $\widehat{s}_i \rightarrow \widehat{s}_i+1$, then the solution for Single-Pass-Algorithm for all downstream installations changes at most one installation.

Proof. Let $\widehat{s} = \{\widehat{s}_j : j \in [|1, N|]\}$ be the solution for Single-Pass-Algorithm. Let the stockage level at some installation $i \in [|1, N|]$ be increased by one unit so that the new stockage level at installation i becomes $\widehat{s}_i + 1$. The fill-rate at any upstream installation $j \in [|i+1, N|]$ does not depend on stockage level at installation i, hence, the solution for Single-Pass-Algorithm for these upstream installations does not change.

Let us first consider the change in stockage level at installation $i-1$ and let us show that the change is at most one unit. To achieve this, let the new stockage level at installation $i-1$ be $\widetilde{s}_{i-1}$. Increasing the stockage level at installation i improves the fill-rate at installation $i-1$, hence, $\widetilde{s}_{i-1} \leq s_{i-1}$ holds true. Suppose now that $\widetilde{s}_{i-1} < s_{i-1} - 1$. This implies that $P(IL_{i-1}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widetilde{s}_{i-1} + 1) > 0) < \beta_{i-1}$. Also, it holds true that $P(IL_{i-1}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i + 1, \widetilde{s}_{i-1}) > 0) \geq \beta_{i-1}$. Moving one unit from installation i to installation $i-1$ improves fill-rate at installation $i-1$. Hence, $P(IL_{i-1}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widetilde{s}_{i-1} + 1) > 0) \geq \beta_{i-1}$ holds true. This results in contradiction, hence, $\widetilde{s}_{i-1} \geq s_{i-1} - 1$. Together with the previous result, it holds true that $\widetilde{s}_{i-1} \in \{s_{i-1} - 1, s_{i-1}\}$. This implies that at most one change occurs at the most immediate downstream installation.

Let us now suppose that there is at most one change in downstream installations $i-1, i-2, \cdots, i-k$, for some $k \in [|1, i-2|]$. This implies that either $(\widehat{s}_{i-1}, \widehat{s}_{i-2}, \cdots, \widehat{s}_{i-k})$ or $(\widehat{s}_{i-1}, \widehat{s}_{i-2}, \cdots, \widehat{s}_{i-p} - 1, \cdots, \widehat{s}_{i-k})$ for some $p \leq k$ holds true. Let us treat these two cases separately.

First, let the new stockage levels at installations $i-1, \cdots, i-k$ be $(\widehat{s}_{i-1}, \widehat{s}_{i-2}, \cdots, \widehat{s}_{i-k})$. Represent the new stockage level at installation $i-(k+1)$ with $\widetilde{s}_{i-(k+1)}$. Increasing the stockage level at installation i improves the fill-rate at installation $i-(k+1)$, hence, $\widetilde{s}_{i-(k+1)} \leq \widehat{s}_{i-(k+1)}$ holds true. Suppose that $\widetilde{s}_{i-(k+1)} < \widehat{s}_{i-(k+1)} - 1$. This implies that $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)} + 1) > 0) < \beta_{i-(k+1)}$. It also holds true that $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i + 1, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)}) > 0) \geq \beta_{i-(k+1)}$. Moving one unit of stock from installation i to installation $i-(k+1)$ improves the fill-rate at installation $i-(k+1)$. Hence, $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)} + 1) > 0) \geq \beta_{i-(k+1)}$ holds true. This results in contradiction. Hence,

$\widetilde{s}_{i-(k+1)} \geq \widehat{s}_{i-(k+1)} - 1$ holds true. $\widetilde{s}_{i-(k+1)} \leq \widehat{s}_{i-(k+1)}$ and $\widetilde{s}_{i-(k+1)} \geq \widehat{s}_{i-(k+1)} - 1$ together imply that $\widetilde{s}_{i-(k+1)} \in \{\widehat{s}_{i-(k+1)} - 1, \widehat{s}_{i-(k+1)}\}$. This implies that there is at most one change in downstream installations $i-1, i-2, \cdots, i-k, i-(k+1)$.

Second, let the new stockage levels at installations $i-1, \cdots, i-k$ be $(\widehat{s}_{i-1}, \widehat{s}_{i-2}, \cdots, \widehat{s}_{i-p} - 1, \cdots, \widehat{s}_{i-k})$ for some $p \leq k$. Increasing the stockage level at installation i improves the fill-rate at installation $i-(k+1)$, hence, $\widetilde{s}_{i-(k+1)} \leq \widehat{s}_{i-(k+1)}$ holds true. Suppose that $\widetilde{s}_{i-(k+1)} < \widehat{s}_{i-(k+1)}$ holds true. This implies that $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)}) > 0) < \beta_{i-(k+1)}$. It also holds true that $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i + 1, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-p} - 1, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)}) > 0) \geq \beta_{i-(k+1)}$. Since $p < i$, moving one unit of stock from installation i to installation $i-p$ improves the fill-rate at installation $i-(k+1)$. Hence, $P(IL_{i-(k+1)}(\widehat{s}_N, \widehat{s}_{N-1}, \cdots, \widehat{s}_i, \widehat{s}_{i-1}, \cdots, \widehat{s}_{i-p}, \cdots, \widehat{s}_{i-k}, \widetilde{s}_{i-(k+1)}) > 0) \geq \beta_{i-(k+1)}$. This results in a contradiction. Hence, $\widetilde{s}_{i-(k+1)} \geq \widehat{s}_{i-(k+1)}$ holds true. $\widetilde{s}_{i-(k+1)} \leq \widehat{s}_{i-(k+1)}$ and $\widetilde{s}_{i-(k+1)} \geq \widehat{s}_{i-(k+1)}$ together imply that $\widetilde{s}_{i-(k+1)} = \widehat{s}_{i-(k+1)}$. This implies that there is at most one change in downstream installations $i-1, i-2, \cdots, i-k, i-(k+1)$. This completes the induction argument. Hence, if the solution for Single-Pass-Algorithm is increased by one unit at some installation $i \in [|1, N|]$, the solution for Single-Pass-Algorithm for all downstream installations changes at most one installation. ∎

Proposition 15

For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation 3 is less than or equal to one unit.

Proof. Suppose that stockage level at installation 3 is increased by one unit so that the new set of stockage levels becomes $(s_3 + 1, s_2, s_1)$. The decrease in expected backorders that belong to installation 3 simply becomes $\Delta_{3,3}(s_3) = \frac{\lambda_3}{\lambda} \cdot P(B_3(s_3) > 0)$. The decrease in expected backorders that belong to installation 2 becomes $\Delta_{2,2}(s_3) = \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot [E[B_2(s_3, s_2)] - E[B_2(s_3+1, s_2)]]$. It is easy to show that $E[B_2(s_3, s_2)] - E[B_2(s_3+1, s_2)] \leq E[B_2(s_3, 0)] - E[B_2(s_3+1, 0)]$. Moreover, it holds true that $E[B_2(s_3, 0)] = E[B_{3,2}(s_3)]$ and $E[B_2(s_3+1, 0)] = E[B_{3,2}(s_3+1)]$. This implies that $\Delta_{2,2}(s_3) \leq \frac{\lambda_2}{\lambda} \cdot [E[B_3(s_3) - E[B_3(s_3+1)]] = \frac{\lambda_2}{\lambda} \cdot P(B_3(s_3) > 0)$. The decrease in expected backorders at installation 1 becomes $\Delta_1(s_3) = E[B_1(s_3, s_2, s_1)] - E[B_1(s_3 + 1, s_2, s_1)]$. Moreover, it is easy to show that $E[B_1(s_3, s_2, s_1)] - E[B_1(s_3 + 1, s_2, s_1)] \leq E[B_1(s_3, s_2, 0)] - E[B_1(s_3 + 1, s_2, 0)]$. Since $E[B_1(s_3, s_2, 0)] = E[B_{2,1}(s_3, s_2)]$ and $E[B_1(s_3 + 1, s_2, 0)] = E[B_{2,1}(s_3 + 1, s_2)]$ hold true, we can write $E[B_1(s_3, s_2, s_1)] - E[B_1(s_3 + 1, s_2, s_1)] \leq E[B_{2,1}(s_3, s_2)] - E[B_{2,1}(s_3 + 1, s_2)] = \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot [E[B_2(s_3, s_2)] - E[B_2(s_3 + 1, s_2)]]$. Furthermore, it also holds true that $E[B_2(s_3, s_2)] - E[B_2(s_3 + 1, s_2)] \leq E[B_2(s_3, 0)] - E[B_2(s_3+1, 0)]$. And, $E[B_2(s_3, 0)] = E[B_{3,2}(s_3)]$ and $E[B_2(s_3+1, 0)] = E[B_{3,2}(s_3 + 1)]$ hold true. Replacing these expressions into the earlier inequality results in $\Delta_1(s_3) = E[B_1(s_3, s_2, s_1)] - E[B_1(s_3+1, s_2, s_1)] \leq \frac{\lambda_1}{\lambda} \cdot [E[B_3(s_3)] - E[B_3(s_3 + 1)]]$. Combining all the above derived expressions for the decrease in expected backorders implies that $\Delta_{3,3}(s_3) + \Delta_{2,2}(s_3) + \Delta_1(s_3) \leq [E[B_3(s_3)] - E[B_3(s_3 + 1)]] = P(B_3(s_3) > 0)$. Since $P(B_3(s_3) > 0) \leq 1$, this implies that the total decrease in expected backorders when one unit of stock is added at installation 3 is less than or equal to one unit. ■

Proposition 16

For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation 2 is less than or equal to one unit.

Proof. Suppose that stockage level at installation 2 is increased by one unit so that the new set of stockage levels becomes $(s_3, s_2 + 1, s_1)$. This increase in stockage level does not affect the expected backorders at the upstream installation 3, however, it decreases the expected backorders that belong to installation 2 and 1. The decrease in expected backorders at installation 2 becomes $E[B_{2,2}(s_3, s_2)] - E[B_{2,2}(s_3, s_2+1)] = \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot [E[B_2(s_3, s_2)] - E[B_2(s_3, s_2 + 1)]]$. The decrease in expected backorders at installation 1, on the other hand, becomes $E[B_1(s_3, s_2, s1)] - E[B_1(s_3, s_2 + 1, s1)]$. It is easy to show that $E[B_1(s_3, s_2, s1)] - E[B_1(s_3, s_2 + 1, s1)] \leq E[B_1(s_3, s_2, 0)] - E[B_1(s_3, s_2 + 1, 0)]$ (i). Since $E[B_1(s_3, s_2, 0)] = E[B_{2,1}(s_3, s_2)]$ and $E[B_1(s_3, s_2 + 1, 0)] = E[B_{2,1}(s_3, s_2 + 1)]$, it holds true that $E[B_1(s_3, s_2, s1)] - E[B_1(s_3, s_2 + 1, s1)] \leq \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot [E[B_2(s_3, s_2)] - E[B_2(s_3, s_2 + 1)]]$ (ii). Combining the two inequalities in (i) and (ii) implies that total decrease in expected backorders is simply less than or equal to

$E[B_2(s_3, s_2)] - E[B_2(s_3, s_2 + 1)]]$, which is equivalent to $P(B_{3,2}(s_3) > s_2)$. Hence, the total decrease in expected backorders when one unit of stock is added at installation 2 is less than or equal to one unit. ■

Proposition 17

For a problem with three demand classes, the decrease in total expected backorders through adding one unit of stock at installation 1 is less than or equal to one unit.

Proof. Suppose that stockage level at installation 1 is increased by one unit so that the new set of stockage levels becomes $(s_3, s_2, s_1 + 1)$. This increase in stockage level does not affect the expected backorders at the upstream installation 3 and 2. And, it only decreases the expected backorders at installation 1. The decrease in expected backorders at installation 1 becomes $E[B_1(s_3, s_2, s_1)] - E[B_1(s_3, s_2, s_1 + 1)] = P(B_{2,1}(s_3, s_2) > s_1)$. This simply implies that increasing the stockage level at installation 1 by one unit decreases the expected backorders at installation 1 at most by one unit. ■

Proposition 18

For a given set of stockage levels $s_3 \in Z, s_2$ and $s_1 \in Z^+ \cup \{0\}$,

$$i)\ TC(s_3, s_2 + 1, s_1 - 1) < TC(s_3, s_2, s_1)$$

$$ii)\ TC(s_3 + 1, s_2 - 1, s_1) < TC(s_3, s_2, s_1)$$

Proof. We begin with showing the first inequality. Consider total expected on-hand inventory level function $TC(s_3, s_2, s_1)$ for a set of stockage levels $\{s_3, s_2, s_1\}$, which can be written in the following form: $TC(s_3, s_2, s_1) = s_3 + s_2 + s_1 - E[D^L] + E[B_{3,3}(s_3)] + E[B_{2,2}(s_3, s_2)] + E[B_1(s_3, s_2, s_1)]$. Now, consider stockage levels $\{s_3, s_2 + 1, s_1 - 1\}$, which results in $TC(s_3, s_2 + 1, s_1 - 1) = s_3 + s_2 + s_1 - E[D^L] + E[B_{3,3}(s_3)] + E[B_{2,2}(s_3, s_2 + 1)] + E[B_1(s_3, s_2 + 1, s_1 - 1)]$. Moreover, we can also write down

$E[B_{2,2}(s_3, s_2+1)]$ and $E[B_1(s_3, s_2+1, s_1-1)]$ in the following forms:

$$
\begin{aligned}
E[B_{2,2}(s_3, s_2+1)] &= E[B_{2,2}(s_3, s_2)] - \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot P(B_2(s_3, s_2) > 0) \\
E[B_1(s_3, s_2+1, s_1-1)] &= E[B_1(s_3, s_2, s_1)] + \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot P(B_1(s_3, s_2+1, s_1-1) > 0)
\end{aligned}
$$

Replacing these above expressions in $TC(s_3, s_2+1, s_1-1)$ simply yields $TC(s_3, s_2+1, s_1-1) = TC(s_3, s_2, s_1) - \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot [P(B_2(s_3, s_2) > 0) - P(B_1(s_3, s_2+1, s_1-1) > 0)]$. Moreover, it holds true that $P(B_2(s_3, s_2) > 0) - P(B_1(s_3, s_2+1, s_1-1) > 0) > 0$, which implies that $TC(s_3, s_2+1, s_1-1) > TC(s_3, s_2, s_1)$. To show the second inequality, consider stockage levels $\{s_3+1, s_2-1, s_1\}$, which results in $TC(s_3+1, s_2-1, s_1) = s_3+s_2+s_1 - E[D^L] + E[B_{3,3}(s_3+1)] + E[B_{2,2}(s_3+1, s_2-1)] + E[B_1(s_3+1, s_2-1, s_1)]$. We can write down $E[B_{3,3}(s_3+1)]$, $E[B_{2,2}(s_3+1, s_2-1)]$, and $E[B_1(s_3+1, s_2-1, s_1)]$ in the following forms:

$$
\begin{aligned}
E[B_{3,3}(s_3+1)] &= E[B_{3,3}(s_3)] - \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot P(B_3(s_3) > 0) \\
E[B_{2,2}(s_3+1, s_2-1)] &= E[B_{2,2}(s_3, s_2)] + \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot \\
&\quad P(B_2(s_3+1, s_2-1) > 0)
\end{aligned}
$$

$$
\begin{aligned}
E[B_1(s_3+1, s_2-1, s_1)] = E[B_1(s_3, s_2, s_1)] + \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot \\
P(B_1(s_3+1, s_2-1, s_1) > 0)
\end{aligned}
$$

Replacing the previous expressions in $TC(s_3+1, s_2-1, s_1)$ results in

$$
\begin{aligned}
TC(s_3+1, s_2-1, s_1) = TC(s_3, s_2, s_1) - \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot [P(B_3(s_3) > 0) - \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot \\
P(B_2(s_3+1, s_2-1) > 0) - \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot P(B_1(s_3+1, s_2-1, s_1) \\
> 0)]
\end{aligned}
$$

where $P(B_3(s_3) > 0) > P(B_2(s_3+1, s_2-1) > 0) > P(B_1(s_3+1, s_2-1, s_1) > 0)$, which implies that $TC(s_3+1, s_2-1, s_1) < TC(s_3, s_2, s_1)$. $\blacksquare$

Appendix B

Comparing with Threshold Clearing Mechanism

We next compare our serial inventory mapping to the model developed by Deshpande et. al (2003) for two demand classes; high-priority and low-priority demand classes. Furthermore, we show that our model with two demand classes produces the same performance measures in distribution as in Deshpande et. al (2003), hence, the same inventory costs.

Deshpande et. al (2003) develop an inventory rationing mechanism that combines the well known critical level policy with a *Threshold Clearing* mechanism to clear backorders upon arrival of replenishment orders. The main assumptions in their model are: there are two demand classes; class 1 (high priority demand class) and class 2 (low priority demand class); each demand class has an independent Poisson process; there is a positive fixed replenishment leadtime; and inventory is replenished with a continuous-review (Q,R) policy with rationing. Some important properties of continuous review (Q, R) policy with rationing level K are as follows:

Inventory position is continuously observed and a replenishment order of size Q is placed whenever inventory position drops to or below level R. When on-hand inventory level drops to the level K, class 2 demands (low priority) begin to backlog. When on-hand inventory level drops to zero, class 1 demands (high priority) also begin to backlog. When an inventory replenishment order Q arrives, inventory level is raised up if there are no backorders. If, however, there are backorders upon arrival of a replenishment order, then incoming replenishment order must be rationed between backorders from class 1 and class 2 according to a chosen backlog clearing policy. Hence, a continuous-review (Q, R) policy with rationing in a backlog environment needs to have a pre-specified backlog clearing policy to handle incoming replenishment orders.

Deshpande et. al (2003) develop their *Threshold Clearing* mechanism to handle incoming replenishment orders for an inventory rationing model with two demand classes. Before we describe this clearing mechanism in detail, we provide the following notation:

$$
\begin{aligned}
\tau &= \text{ replenishment leadtime} \\
K &= \text{ inventory rationing level} \\
OH(t) &= \text{ on-hand inventory level at time } t \\
IP(t) &= \text{ inventory position at time } t \\
D_1(t, t+\tau) &= \text{ demand from class 1 during time interval } (t, t+\tau] \\
D_2(t, t+\tau) &= \text{ demand from class 2 during time interval } (t, t+\tau] \\
D(t, t+\tau) &= D_1(t, t+\tau) + D_2(t, t+\tau) \\
BO_1(t) &= \text{ class 1 backorders at time } t \\
BO_2(t) &= \text{ class 2 backorders at time } t \\
t_K &= \text{ the time of the } IP(t) - K\text{th demand arrival in the interval } (t, t+\tau]
\end{aligned}
$$

For a given value y of the inventory position at time t, IP(t), the following equations represent the dynamics of the random variables for on-hand inventory, class 1 backorders, and class 2 backorders, as stated in Deshpande et. al (2003):

$$
\begin{aligned}
OH(t+\tau) &= y - D(t, t+\tau) && \text{if } D(t, t+\tau) \leq y - K, \\
OH(t+\tau) &= [K - D_1(t_K, t+\tau)]^+ && \text{if } D(t, t+\tau) > y - K,
\end{aligned}
$$

$$BO_1(t+\tau) = [D_1(t_K, t+\tau) - K]^+$$
$$BO_2(t+\tau) = D_2(t_K, t+\tau)$$

In steady state the above set of equations can be represented as follows, where the inventory position IP has value of y and we use the following representation for the random variables in steady-state: $OH(t+\tau) \approx OH$, $BO_1(t+\tau) \approx BO_1$, $BO_2(t+\tau) \approx BO_2$, $D(t, t+\tau) \approx D_\tau$, $D_1(t_K, t+\tau) \approx D_1$, $D_2(t_K, t+\tau) \approx D_2$:

$$\begin{aligned} OH &= y - D_\tau && \text{if } D_\tau \leq y - K, \\ OH &= [K - D_1]^+ && \text{if } D_\tau > y - K, \end{aligned}$$

$$BO_1 = [D_1 - K]^+$$
$$BO_2 = D_2$$

Using the above inventory dynamics in steady-state, the distribution of backorders for class 1 and class 2 can be easily derived under *Threshold Clearing* mechanism. The details are as follows, where IP refers to random variable for inventory position in steady-state:

$$\begin{aligned} P(BO_2 = i) =& P(D_2 = i) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} P(D_2 = i | IP = y) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} P(D_2 = i | IP = y) + \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{R+Q} P(D_2 = i | IP = y) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} P(D_2 = i | IP = y) + \\ & \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{R+Q} \sum_{x=i+y-K}^{\infty} P(D_2 = i | IP = y, D_\tau = x) \cdot P(D_\tau = x) \end{aligned}$$

$$BO_1(t+\tau) = [D_1(t_K, t+\tau) - K]^+$$
$$BO_2(t+\tau) = D_2(t_K, t+\tau)$$

In steady state the above set of equations can be represented as follows, where the inventory position IP has value of y and we use the following representation for the random variables in steady-state: $OH(t+\tau) \approx OH$, $BO_1(t+\tau) \approx BO_1$, $BO_2(t+\tau) \approx BO_2$, $D(t, t+\tau) \approx D_\tau$, $D_1(t_K, t+\tau) \approx D_1$, $D_2(t_K, t+\tau) \approx D_2$:

$$\begin{aligned} OH &= y - D_\tau && \text{if } D_\tau \leq y - K, \\ OH &= [K - D_1]^+ && \text{if } D_\tau > y - K, \end{aligned}$$

$$BO_1 = [D_1 - K]^+$$
$$BO_2 = D_2$$

Using the above inventory dynamics in steady-state, the distribution of backorders for class 1 and class 2 can be easily derived under *Threshold Clearing* mechanism. The details are as follows, where IP refers to random variable for inventory position in steady-state:

$$\begin{aligned} P(BO_2 = i) =& P(D_2 = i) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} P(D_2 = i | IP = y) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} P(D_2 = i | IP = y) + \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{R+Q} P(D_2 = i | IP = y) \\ =& \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} P(D_2 = i | IP = y) + \\ & \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{R+Q} \sum_{x=i+y-K}^{\infty} P(D_2 = i | IP = y, D_\tau = x) \cdot P(D_\tau = x) \end{aligned}$$

$$=\frac{1}{Q}\cdot\sum_{y=R+1}^{K-1}\sum_{n=0}^{i}\binom{K-y}{i-n}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{i-n}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{K-y-(i-n)}\cdot$$
$$\frac{e^{-\lambda_2\cdot\tau}\cdot(\lambda_2\cdot\tau)^n}{n!}+\frac{1}{Q}\cdot\sum_{y=Max(K,R+1)}^{R+Q}\sum_{x=i+y-K}^{\infty}\binom{x-(y-K)}{i}$$
$$(\frac{\lambda_2}{\lambda_1+\lambda_2})^{i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{x-(y-K)-i}\cdot P(D_\tau=x)$$

When there are only two demand classes, the distribution of class 2 backorders in our model becomes as follows, where we note that class 2 backorders refer to backorders at installation 2 that are created by the exogenous demand at installation 2 in the equivalent serial inventory system:

$$P(B_{2,2}=i)=\frac{1}{Q}\cdot\sum_{y=s_2+1}^{s_2+Q}\sum_{j=i}^{\infty}\binom{j}{i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{j-i}\cdot P(D_L=j+y),\quad \text{for all } i\in Z^+$$

We show that the expression for $P(BO_2=i)$ is equivalent to the expression for $P(B_{2,2}=i)$ for all $i\in Z^+\cup\{0\}$. The details of this equivalency is provided in Section $B.1$. Similarly, $P(BO_1=i)$ and $P(B_{1,1}=i)$ have the following expressions, where by the same token an equivalency can be built between them.

$$\begin{aligned}P(BO_1=i)=&P([D_1-K]^+=i)\\=&\frac{1}{Q}\cdot\sum_{y=R+1}^{R+Q}P(D_1=i|IP=y)\\=&\frac{1}{Q}\cdot\sum_{y=R+1}^{R+Q}\sum_{x=y+i}^{\infty}P(D_1=i|IP=y,D_L=x)\cdot P(D_L=x)\\=&\frac{1}{Q}\cdot\sum_{y=R+1}^{R+Q}\sum_{x=y+i}^{\infty}\binom{x-(y-K)}{K+i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{K+i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{x-y-i}\cdot P(D_L=x)\end{aligned}$$

$$P(B_{1,1}=i)=\frac{1}{Q}\cdot\sum_{y=s_2+1}^{s_2+Q}\sum_{j=s_1+i}^{\infty}\binom{j}{s_1+i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{s_1+i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-(s_1+i)}\cdot P(D_L=j+y)$$

We show in Section $B.2$ that the expression for $P(BO_1=i)$ is equivalent to the expression for $P(B_{1,1}=i)$ for all $i\in Z^+\cup\{0\}$ under the following mapping: $s_1=K$

and $s_2 = R - K$. Next, we show that distribution of on-hand inventory level under *Threshold Clearing* mechanism is equivalent to that in our model. The following probability for on-hand inventory level holds true for *Threshold Clearing* mechanism:

$$P(OH = i) = \frac{1}{Q}\sum_{y=r+1}^{r+Q} P(OH = i|IP = y)$$

$$= \begin{cases} \frac{1}{Q}\sum_{y=r+1}^{r+Q} P(D_\tau = y - i), \text{if } i \in [K, y], \\ \frac{1}{Q}\sum_{y=r+1}^{r+Q}\sum_{x=y-i}^{\infty} \binom{x-(y-K)}{K-i}(\frac{\lambda_1}{\lambda_1+\lambda_2})^{K-i}(\frac{\lambda_2}{\lambda_1+\lambda_2})^{x-y+i}P(D_\tau = x), \text{if } i \in (0, K), \\ \frac{1}{Q}\sum_{y=r+1}^{r+Q}\sum_{x=y}^{\infty}\sum_{z=K}^{x-y+K} \binom{x-(y-K)}{z}(\frac{\lambda_1}{\lambda_1+\lambda_2})^{z}(\frac{\lambda_2}{\lambda_1+\lambda_2})^{x-y+K-z}P(D_\tau = x), \text{if } i = 0, \end{cases}$$

Similarly, distribution of on-hand inventory level in our model can be derived easily. Furthermore, we show that on hand inventory levels are the same in distribution for both models with only two demand classes. The details of this result is given in Section $B.3$. Since our model with two demand classes yields inventory levels that are the same in distribution as the one in Desphande et. al (2003), our model with only two demand classes results in the same performance measures and the same inventory costs as the ones in Deshpande et. al (2003). In addition to achieving the same performance with the most recent study in literature, our model, furthermore, provides a more transparent analysis and can handle any number of demand classes with an easy extension to multi-echelon setting, which are the main advantages of our model over the existing studies in inventory management literature.

B.1 Distributions of Backorders for Class 2

We first replace $s_1 = K$ and $s_2 = R - K$ in the expression for $P(B_{2,2} = i)$. This results in the following expression:

$$P(B_{2,2} = i) = \frac{1}{Q} \cdot \sum_{y=R-K+1}^{R-K+Q}\sum_{j=i}^{\infty}\binom{j}{i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^i \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{j-i} \cdot P(D_L = j + y)$$

We change the limits of summation so that $\sum\limits_{y=R-K+1}^{R-K+Q} P(D_L = j + y)$ is equivalent to $\sum\limits_{y=R+1}^{R+Q} P(D_L = j + y - K)$. We replace this equivalent expression back in $P(B_{2,2} = i)$, which becomes:

$$P(B_{2,2} = i) = \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^{j-i} \cdot P(D_L = j + y - K)$$

The first summation in the above expression is from $y = R + 1$ to $y = R + Q$. We can divide this summation into two parts with a variable $K \in [0, R + Q]$. The first part is the summation of the same expression from $y = R + 1$ to $y = K - 1$. And, the second part is the summation of the same expression from $y = Max(K, R + 1)$ to $y = R + Q$ as depicted below:

$$\begin{aligned} P(B_{2,2} = i) = & \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^{j-i} \cdot P(D_L = j + y - K) + \\ & \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{K-1} \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^{j-i} \cdot P(D_L = j + y - K) \end{aligned}$$

We introduce a new variable x and replace $x = j + y - K$ in the second summation expression in $P(B_{2,2} = i)$, which yields the following expression:

$$\begin{aligned} P(B_{2,2} = i) = & \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} \sum_{j=i}^{\infty} \binom{j}{i} \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^i \cdot (\frac{\lambda_1}{\lambda_1 + \lambda_2})^{j-i} \cdot P(D_L = j + y - K) + \\ & \frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{K-1} \sum_{x=i+y-K}^{\infty} \binom{x - (y - K)}{i} \cdot (\frac{\lambda_2}{\lambda_1 + \lambda_2})^i \cdot \\ & (\frac{\lambda_1}{\lambda_1 + \lambda_2})^{x-(y-K)-i} \cdot P(D_L = x) \end{aligned}$$

Similarly, we introduce a new variable n and replace $n = j - i$ in the first summation expression in $P(B_{2,2}=i)$, which yields the following:

$$P(B_{2,2}=i) = \frac{1}{Q}\cdot\sum_{y=R+1}^{K-1}\sum_{n=0}^{\infty}\binom{n+i}{i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^n\cdot P(D_L=n+i+y-K)$$
$$+\frac{1}{Q}\cdot\sum_{y=Max(K,R+1)}^{K-1}\sum_{x=i+y-K}^{\infty}\binom{x-(y-K)}{i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot$$
$$(\frac{\lambda_1}{\lambda_1+\lambda_2})^{x-(y-K)-i}\cdot P(D_L=x)$$

We can replace $\sum_{n=0}^{\infty}\binom{n+i}{i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^n\cdot P(D_L=n+i+y-K)$ with $(\frac{\lambda_1+\lambda_2}{\lambda_1})^{i-(K-y)}\cdot e^{-(\lambda_1+\lambda_2)\cdot L}\cdot\sum_{n=0}^{\infty}\binom{n+i}{i}\cdot\frac{(\lambda_1\cdot L)^{n+i-(K-y)}}{(n+i-(K-y))!}$. Moreover, for $K>y$, we have the following equality $\sum_{n=0}^{\infty}\binom{n+i}{i}\cdot\frac{(\lambda_1\cdot L)^{n+i-(K-y)}}{(n+i-(K-y))!}=\sum_{n=0}^{i}\binom{K-y}{i-n}\cdot\frac{(\lambda_1\cdot L)^n}{n!}\cdot e^{\lambda_1\cdot L}$. Replacing this last equality into $P(B_{2,2}=i)$ provides us with the following:

$$P(B_{2,2}=i) = \frac{1}{Q}\cdot\sum_{y=R+1}^{K-1}\sum_{n=0}^{i}\binom{K-y}{i-n}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot(\frac{\lambda_1+\lambda_2}{\lambda_1})^{i-(K-y)}\cdot e^{-(\lambda_1+\lambda_2)\cdot L}\cdot$$
$$\frac{(\lambda_1\cdot L)^n}{n!}\cdot e^{\lambda_1\cdot L}+\frac{1}{Q}\cdot\sum_{y=Max(K,R+1)}^{K-1}\sum_{x=i+y-K}^{\infty}\binom{x-(y-K)}{i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot$$
$$(\frac{\lambda_1}{\lambda_1+\lambda_2})^{x-(y-K)-i}\cdot P(D_L=x)$$

Further simplification of the first summation part in the above expression results in:

$$P(B_{2,2}=i) = \frac{1}{Q}\cdot\sum_{y=R+1}^{K-1}\sum_{n=0}^{i}\binom{K-y}{i-n}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{(K-y)-i}\cdot e^{-\lambda_2\cdot L}\cdot\frac{(\lambda_1\cdot L)^n}{n!}$$
$$+\frac{1}{Q}\cdot\sum_{y=Max(K,R+1)}^{K-1}\sum_{x=i+y-K}^{\infty}\binom{x-(y-K)}{i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot$$
$$(\frac{\lambda_1}{\lambda_1+\lambda_2})^{x-(y-K)-i}\cdot P(D_L=x)$$

Moreover, we have the following equality $(\frac{\lambda_2}{\lambda_1+\lambda_2})^i\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{(K-y)-i}\cdot e^{-\lambda_2\cdot L}\cdot\frac{(\lambda_1\cdot L)^n}{n!}=(\frac{\lambda_2}{\lambda_1+\lambda_2})^{i-n}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{(K-y)-i+n}\cdot(\frac{\lambda_2}{\lambda_1})^n\cdot e^{-\lambda_2\cdot L}\cdot\frac{(\lambda_1\cdot L)^n}{n!}$, which further simplifies to $(\frac{\lambda_2}{\lambda_1+\lambda_2})^{i-n}\cdot$

$(\frac{\lambda_1}{\lambda_1+\lambda_2})^{(K-y)-i+n} \cdot \frac{(\lambda_2 \cdot L)^n \cdot e^{-\lambda_2 \cdot L}}{n!}$. Replacing this last equality into $P(B_{2,2}=i)$ yields the following:

$$P(B_{2,2}=i) = \frac{1}{Q} \cdot \sum_{y=R+1}^{K-1} \sum_{n=0}^{i} \binom{K-y}{i-n} (\frac{\lambda_2}{\lambda_1+\lambda_2})^{i-n} (\frac{\lambda_1}{\lambda_1+\lambda_2})^{(K-y)-i+n} \frac{(\lambda_2 L)^n e^{-\lambda_2 L}}{n!} +$$
$$\frac{1}{Q} \cdot \sum_{y=Max(K,R+1)}^{K-1} \sum_{x=i+y-K}^{\infty} \binom{x-(y-K)}{i} (\frac{\lambda_2}{\lambda_1+\lambda_2})^{i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{x-(y-K)-i}$$
$$P(D_L = x)$$

This shows that $P(B_{2,2}=i) = P(BO_2 = i)$ for all $i \in Z^+ \cup \{0\}$. Hence, steady-state distribution of class 2 backorders are the same for the random clearing mechanism we developed and the threshold clearing mechanism developed by Desphande et al. (2003) if we use the following mapping, $s_1 = K$ and $s_2 = R - K$.

B.2 Distributions of Backorders for Class 1

We replace $s_1 = K$ and $s_2 = R - K$ in the previous expression for $P(B_{1,1} = i)$. This results in the following expression:

$$P(B_{1,1}=i) = \frac{1}{Q} \cdot \sum_{y=R-K+1}^{R-K+Q} \sum_{j=K+i}^{\infty} \binom{j}{K+i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{K+i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-(K+i)} \cdot P(D_L = j+y)$$

We change the limits of summation so that $\sum_{y=R-K+1}^{R-K+Q} P(D_L = j+y)$ is equivalent to $\sum_{y=R+1}^{R+Q} P(D_L = j+y-K)$. We replace this expression back in $P(B_{1,1}=i)$, which becomes:

$$P(B_{1,1}=i) = \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} \sum_{j=K+i}^{\infty} \binom{j}{K+i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{K+i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-(K+i)} \cdot$$
$$P(D_L = j+y-K)$$

We introduce a new variable x and replace $x = j + y - K$ in $P(B_{2,2} = i)$, which yields the following expression:

$$P(B_{1,1}=i)=\frac{1}{Q}\cdot\sum_{y=R+1}^{R+Q}\sum_{x=i+y}^{\infty}\binom{x-(y-K)}{K+i}\cdot(\frac{\lambda_1}{\lambda_1+\lambda_2})^{K+i}\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{x-y-i}\cdot P(D_L=x)$$

This shows that $P(B_{1,1}=i)=P(BO_1=i)$ for all $i\in Z^+\cup\{0\}$. Hence, steady-state distribution of class 1 backorders are the same for random clearing mechanism and Threshold clearing mechanism if we use the following mapping, $s_1=K$ and $s_2=R-K$.

B.3 Distributions of On-hand Inventory Level

To derive the distributions of on-hand inventory level, we need to consider two cases that are feasible, where IL^+ denotes the on-hand inventory level in steady-state: $(IL_2^+>0, IL_1^+=s_1)$ and $(IL_2^+=0, 0\le IL_1^+\le s_1)$. We first consider the region in which $i\ge s_1$. If $IL_1^++IL_2^+=i\ge s_1$, then we can conclude that $IL_2^+=i-s_1$. Hence,

$$P(IL_1^++IL_2^+=i)=P(IL_2^+=i-s_1)=\frac{1}{Q}\cdot\sum_{y=s_2+1}^{s_2+Q}P(D_L=y-i+s_1),\quad \text{for } i\ge s_1$$

Next, we consider the case in which $i<s_1$. If $IL_1^++IL_2^+=i<s_1$, then we can conclude that $IL_1^+=i<s_1$. Hence,

$$\begin{aligned}
P(IL_1^++IL_2^+=i)&=\frac{1}{Q}\cdot\sum_{y=s_2+1}^{s_2+Q}P(IL_1^+=i|IP=y)\\
&=\begin{cases}\frac{1}{Q}\sum\limits_{y=s_2+1}^{s_2+Q}P(B_{2,1}=s_1-i), & \text{for } i\in(0,s_1),\\ \frac{1}{Q}\sum\limits_{y=s_2+1}^{s_2+Q}\sum\limits_{n=s_1}^{\infty}P(B_{2,1}=n), & \text{for } i=0,\end{cases}\\
&=\begin{cases}\frac{1}{Q}\sum\limits_{y=s_2+1}^{s_2+Q}\sum\limits_{j=s_1-i}^{\infty}\binom{j}{s_1-i}(\frac{\lambda_1}{\lambda_1+\lambda_2})^{s_1-i}(\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-s_1+i}P(D_L=j+y),\\ \frac{1}{Q}\sum\limits_{y=s_2+1}^{s_2+Q}\sum\limits_{n=s_1}^{\infty}\sum\limits_{j=n}^{\infty}\binom{j}{n}(\frac{\lambda_1}{\lambda_1+\lambda_2})^{n}(\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-n}P(D_L=j+y),\end{cases}
\end{aligned}$$

We would like to show that the expression for $P(IL_1^+ + IL_2^+ = i)$ is equivalent to the expression for $P(OH = i)$ for all $i \in Z^+ \cup \{0\}$. We first replace $s_1 = K$ and $s_2 = R - K$ in the previous expressions for $P(IL_1^+ + IL_2^+ = i)$. This results in the following expressions:

$$P(IL_1^+ + IL_2^+ = i) = \frac{1}{Q} \cdot \sum_{y=R-K+1}^{R-K+Q} P(D_L = y - i + K), \quad \text{for } i \geq K$$

$$P(IL_1^+ + IL_2^+ = i) = \begin{cases} \frac{1}{Q} \cdot \sum\limits_{y=R-K+1}^{R-K+Q} \sum\limits_{j=K-i}^{\infty} \binom{j}{K-i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{K-i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-K+i} \cdot \\ P(D_L = j + y), \text{ for } i \in (0, K), \\ \frac{1}{Q} \cdot \sum\limits_{y=R-K+1}^{R-K+Q} \sum\limits_{n=K}^{\infty} \sum\limits_{j=n}^{\infty} \binom{j}{n} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{n} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-n} \cdot P(D_L = j + y), \\ \text{for } i = 0, \end{cases}$$

Moreover, $\sum\limits_{y=R-K+1}^{R-K+Q} P(D_L = y - i + K)$ is equivalent to $\sum\limits_{y=R+1}^{R+Q} P(D_L = y - i)$. And, $\sum\limits_{y=R-K+1}^{R-K+Q} P(D_L = j + y)$ is equivalent to $\sum\limits_{y=R+1}^{R+Q} P(D_L = j + y - K)$. Replacing these in expressions for $P(IL_1^+ + IL_2^+ = i)$ results in the following:

$$P(IL_1^+ + IL_2^+ = i) = \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} P(D_L = y - i), \quad \text{for } i \geq K$$

$$P(IL_1^+ + IL_2^+ = i) = \begin{cases} \frac{1}{Q} \cdot \sum\limits_{y=R+1}^{R+Q} \sum\limits_{j=K-i}^{\infty} \binom{j}{K-i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{K-i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-K+i} \cdot \\ P(D_L = j + y - K), \text{ for } i \in (0, K), \\ \frac{1}{Q} \cdot \sum\limits_{y=R+1}^{R+Q} \sum\limits_{n=K}^{j} \sum\limits_{j=n}^{\infty} \binom{j}{n} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{n} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-n} \cdot P(D_L = j + y - K), \\ \text{for } i = 0, \end{cases}$$

Furthermore, we introduce a new variable x such that $x = j + y - K$. Replacing x into above expression for $P(IL_1^+ + IL_2^+ = i)$ yields the following:

$$P(IL_1^+ + IL_2^+ = i) = \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} P(D_L = y - i), \quad \text{for } i \geq K$$

$$P(IL_1^+ + IL_2^+ = i) = \begin{cases} \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} \sum_{x=y-i}^{\infty} \binom{x-(y-K)}{K-i} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{K-i} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{x-y+i} \cdot P(D_L = x), \\ \text{for } i \in (0, K), \\ \frac{1}{Q} \cdot \sum_{y=R+1}^{R+Q} \sum_{n=K}^{x-y+K} \sum_{j=n}^{\infty} \binom{x-(y-K)}{n} \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2})^{n} \cdot (\frac{\lambda_2}{\lambda_1+\lambda_2})^{j-n} \cdot P(D_L = x), \\ \text{for } i = 0, \end{cases}$$

This shows that $P(IL_1^+ + IL_2^+ = i) = P(OH = i)$ for all $i \in Z^+ \cup \{0\}$. Hence, steady-state distribution of on-hand inventory level is the same for the random clearing mechanism that we developed and the threshold clearing mechanism developed by Deshpande et al. (2003) if we use the following mapping, $s_1 = K$ and $s_2 = R - K$.

Bibliography

[1] Adler, P. S. 1990. "Shared Learning," *Management Science*, Vol. 36, No. 8, 938-957.

[2] Argote, L. and P. Ingram, 2000. "Knowledge Transfer: A Basis for Competitive Advantage in Firms," *Organizational Behavior and Human Decision Processes*, Vol. 82, No. 1, 150-169.

[3] Argote, L., Beckman, S.L., and D. Epple, 1990. "The Persistence and Transfer of Learning in Industrial Settings," *Management Science*, Vol. 36, No. 2, 140-154.

[4] Arrow, K. J. 1962. "The Economics Implications of Learning by Doing," *Review of Economic Studies*, Vol. 29, No. 1, 155-173.

[5] Atkins, D. and Katircioglu, K. 1996. "Managing Inventory for Multiple Customers Requiring Different Levels of Service," Working Paper, University of British Columbia, Vancouver, Canada, B.C.

[6] Baloff, N. 1971. "Extension of the Learning Curve-Some Empirical Results," *Operations Research Quarterly*, Vol. 22, No. 4, 329-340.

[7] Bohn, R. 1995. "Noise and Learning in Semiconductor Manufacturing," *Management Science*, Vol. 41, No. 1, 972-986.

[8] Bohn, R. E. and C. Terwiesch, 1999. "The Economics of Yield-Driven Processes," *Journal of Operations Management*, Vol. 18, 41-59.

[9] Carrillo, J. E. and C. Gaimon, 2000. "Improving Manufacturing Performance Through Process Change and Knowledge Creation," *Management Science*, Vol. 46(2), 265-288.

[10] Clark, D. 2002. "In Setting Up Its New Plants, Chip Maker Clones Older Ones Down to the Paint on the Wall," *Wall Street Journal*, 10/28/2002, Page B1.

[11] Cohen, M., R. Kleindorfer, and H. Lee, 1988. "Service Constrained (s, S) Inventory Systems with Priority Demand Classes and Lost Sales," *Management Science*, **34**, 482-499.

[12] Cohen, M., R. Kleindorfer, and H. Lee, 1989. "Near Optimal Service Constrained Stocking Policies for Spare Parts," *Operations Research*, **37**, 104-116.

[13] Darr, E. D. and T. R. Kurtzberg, 2000. "An Investigation of Partner Similarity Dimensions on Knowledge Transfer," *Organizational Behavior and Human Decision Processes*, Vol. 82, No. 1, 28-44.

[14] Dekker, R., R.M. Hill, and M.J. Kleijn, 1997. "On the $(S-1, S)$ Lost Sales Inventory Model with Priority Demand Classes," Technical Report 9743/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

[15] Deshpande, V. M.A. Cohen, and K. Donohue, 2003. "A Threshold Inventory Rationing Policy for Service-Differentiated Demand Classes," *Management Science*, **49(6)**, 683-703.

[16] Dorroh, J. R., Gulledge, T. R., and N. K. Womer, 1994. "Investment in Knowledge: A Generalization of Learning By Experience," *Management Science*, Vol. 40, No. 8, 947-958.

[17] Dutton, J. M. and A. Thomas, 1984. "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review*, Vol. 9, 235-247.

[18] Evans, R.V., 1968. "Sales and Restocking Policies in a Single Inventory System," *Management Science*, **14**, 463-472.

[19] Fine, C. H., 1986. "Quality Improvement and Learning in Productive Systems," *Management Science*, Vol. 32, 1301-1305.

[20] Frank, K.C., R.Q. Zhang, and I. Duenyas, 1999. "Optimal Policies for Inventory Systems with Priority Demand Classes," Technical Report 99-01, Department of Industrial Engineering and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

[21] Graves, S.C., 1985. "A Multi-Echelon Inventory Model for a Repairable Item with One-for-one Replenishment," *Management Science*, Vol. 31, No. 10, 1247-1256.

[22] Ha, A.Y., 1997a. "Inventory Rationing in a Make-to-stock Production System with Several Demand Classes and Lost Sales," *Management Science*, **43**, 1093-1103.

[23] Ha, A.Y., 1997b. "Stock Rationing Policy for a Make-to-Stock Production System with Two Priority Classes and Backordering," *Naval Research Logistics*, **44**, 457-472.

[24] Hatch, N. W. and D. C. Mowery, 1998. "Process Innovation and Learning by Doing in Semiconductor Manufacturing," *Management Science*, Vol. 44, No. 11(1), 1461-1477.

[25] Haynsworth, H. and B. Price, 1989. "A Model for Use in the Rationing of Inventory during Lead Time," *Naval Research Logistics*, **36**, 491-506.

[26] Johnson, C.R., Neumann, M., and M.J. Tsatsomeros, 1996. "Conditions for the Positivity of Determinants," *Linear and Multilinear Algebra,* Vol. 40, 241-248.

[27] Joskow, P. L. and G. A. Rozanski, 1979. "Effects of Learning by Doing on Nuclear-Plant Operating Reliability," *Review of Economics and Statistics*, Vol. 61, No. 2, 161-168.

[28] Kaplan, A., 1969. "Stock Rationing," *Management Science*, **15**, 260-267.

[29] KLA-Tencor, 2000. "Accelerating Yield," *www.kla-tencor.com*, 1-5.

[30] Lapre, M. A. and L. N. V. Wassenhove, 2001. "Creating and Transferring Knowledge for Productivity Improvement in Factories," *Management Science*, Vol. 47, No. 10, 1311-1325.

[31] Li, G. and S. Rajagopalan, 1998. "Process Improvement, Quality, and Learning Effects," *Management Science*, Vol. 44, No. 11, 1517- 1532.

[32] Lieberman, M. B. 1984. "The Learning Curve and Pricing in the Chemical Processing Industries," *RAND Journal of Economics*, Vol. 15, No. 2, 213-228.

[33] McDonald, C., 1998. "The Evolution of Intel's Copy Exactly! Technology Transfer Model," *Intel Technology Journal*, Q4.

[34] McIvor, R., Martin, J.D., Matsuo, H., and Ng, J. 1997. "Profiting from Process Improvement in the New Semiconductor Manufacturing Environment," *Technology and Operations Review*, Vol. 1(2), 1997.

[35] Melchiors, P., R. Dekker, and M.J. Kleijn, 2000. "Inventory Rationing in an (s, Q) Inventory Model with Lost Sales and Two Demand Classes," Journal of Operations Research Society, **51**, 111-122.

[36] Mody, A. A. 1989. "Firm Strategies for Costly Engineering Learning," *Management Science*, Vol. 35, No. 4, 496-512.

[37] Moon, I. and S. Kang, 1998. "Rationing Policies for Some Inventory Systems," *Journal of the Operations Research Society*, **49**, 509-518.

[38] Nahmias, S. and S. Demmy, 1981. "Operating Characteristics of an Inventory System with Rationing," *Management Science*, **27**, 1236-1245.

[39] National Instruments, 1999. "Visteon Test System Standardization," *www.zone.ni.com/devzone*, 1-4.

[40] Repenning, N. P., 2000. "Drive out Fear (Unless You Can Drive It in): The Role of Agency and Job Security in Process Improvement," *Management Science*, Vol. 46, No. 11, 1385-1396.

[41] Szulanski, G. 2000. "The Process of Knowledge Transfer: A Diachronic Analysis of Stickiness," *Organizational Behavior and Human Decision Processes*, Vol. 82, No. 1, 9-27.

[42] Tang, C.S., 1991. "Designing an Optimal Production System with Inspection," *European Journal of Operational Research*, Vol. 52, No. 1, 45-54.

[43] Terwiesch, C. and R. Bohn, 2001. "Learning and Process Improvement During Production Ramp-up," *International Journal of Production Economics*, Vol. 70(1), 1-19.

[44] Terwiesch, C. and Y. Xu, 2004. "The Copy-Exactly Ramp-up Strategy: Trading-off Learning with Process Change," *IEEE Transactions on Engineering Management*, Vol. 51, Issue 1, 70-84.

[45] Topkis, D., 1968. "Optimal Ordering and Rationing Policies in a Non-stationary Dynamic Inventory Model with n Demand Classes," *Management Science*, **15**, 160-176.

[46] Veinott JR, A.F., 1965. "Optimal Policy in a Dynamic, Single Product, Nonstationary Inventory Model with Several Demand Classes," *Operations Research*, **13**, 761-778.

[47] Wright, T. 1936. "Factors Affecting the Cost of Airplanes," *Journal of Aeronautical Sciences*, Vol. 3, No. 4, 122-128.

[48] Zangwill, W. I. and P. B. Kantor, 1998. "Toward a Theory of Continuous Improvement and the Learning Curve," *Management Science*, Vol. 44, No. 7, 910-920.

[49] Zipkin, P.H., "Foundations of Inventory Management," 2003.