THREE ESSAYS ON PRODUCT RECALL DECISION OPTIMIZATION

THREE ESSAYS ON PRODUCT RECALL DECISION OPTIMIZATION

By

LIUFANG YAO
BSE., MNG.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

McMaster University
©Copyright by Liufang Yao, June 2017

DOCTOR OF PHILOSOPHY (2017)	**McMaster University** **Hamilton, Ontario**
MASTER OF ENGINEERING (2011)	**Huazhong University of Science and Technology** **Wuhan, P. R. China**
BACHELOR OF ENGINEERING (2008)	**Huazhong University of Science and Technology** **Wuhan, P. R. China**

TITLE:	**Three Essays on Product Recall Decision Optimization**
AUTHOR:	**Liufang Yao**
SUPERVISOR:	**Professor Mahmut Parlar**
SUPERVISORY COMMITTEE:	**Professor Mahmut Parlar (Chairman)** **Professor Prakash L. Abad** **Associate Professor Manish Verma**
NUMBER OF PAGES:	**xi, 113**

Abstract

This thesis examines decision optimization of product recalls. Product recalls in recent years have shown unprecedented impact on both immediate economic and reputational damage to the company and long-lasting impact on the brand and industry. Admittedly, imperfect product quality makes recalls inevitable. Thus, we explore from three perspectives to elicit business insights regarding better management and risk control.

Chapter 1 introduces the topic of product recall management optimization and its real-world motivation.

Chapter 2 views the decision making of "when to initiate a product recall" as a dynamic process and takes the feedback of customer returns to update the product defect rate. Updating is simplified by the conjugate properties of beta distribution and Bernoulli trials. We develop the optimal stopping model to find the thresholds of total product returns above which initiating recall is optimal. We implement dynamic programming to solve the model optimally. For large-size problems, we propose a simulation method to balance computation time with solution quality.

Chapter 3 allows the company to control the recall risk by investing in quality. We adopt the one-stage stochastic newsvendor model and add quality-dependent recall risk. The resulting model is not concave in production quantity and quality levels. The parametric analysis reveals several interesting features such as the optimal ordering quantity and quality level have a conflicting relationship. We further extend our model from internal supply to external supply from multiple sources.

Chapter 4 examines managing product recalls from the closed-loop supply chain management and disruption management perspectives. We model the location and allocation decisions of both manufacturing plants and reprocessing facilities where facilities are built after the recalls. Numerical experiments show the costs of overlooking potential recalls vary greatly, indicating the necessity of considering recalls in initial designs and the importance of accurate recall probability prediction.

Chapter 5 summarizes.

Acknowledgements

Foremost, I acknowledge and thank my supervisor and mentor, Dr. Mahmut Parlar, who lit the way at its darkest time in my pursuit of the Ph.D. degree. Dr. Parlar has inspired me tremendously by his dedication to rigorous research and having fun with it, his philosophy of enjoying life to its fullest, and his kindness to others with his "always carrots" policy. His confidence and trust in my ability to finish the Ph.D. degree and be a good researcher lay the very foundation of this thesis.

I would like to extend my sincere thanks to Drs. Prakash Abad and Manish Verma for their careful review of the manuscripts and their many helpful suggestions. My thanks also go to my fellow Ph.D. students in management science and other areas for their constructive feedback. It is the warm and competitive environment created by all the faculty, staff and students of DeGroote that helped reduce my inertia and continuously pushed me forward.

This thesis is dedicated to my dear family for their endless love, encouragement and support throughout my Ph.D. studies.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

This thesis examines three important topics in product recall management optimization with a particular focus on product recall timing, the impact of investment in product quality on product recalls, and aftermath mitigation with reprocessing center construction. These topics are of interest to academics, operations management professionals, and policy makers.

General Motors's decade-long delay in initiating the Cobalt recall, XL Inc.'s disposing of six hundred tons of beef products in intact packaging at landfills, and Samsung's multi-billion-dollar loss for settling the market and societal costs related to its potentially explosive Galaxy Note 7 have captured wide interest in academia, practice and the public — poor management of product recalls can cause great harm to the public, the company and its customers (Chao et al. [8]). Our research has practical operations management implications about the need for integrating product recall concerns into the product quality control system and viewing the recall decision making as a dynamic process with timely updated information.

The first essay, "Product Recall Timing Optimization using Dynamic Programming", examines the decision-making of product recall as a dynamic process and searches for the optimal stage of initiating a product recall. This topic matters because the timing of a company's product recall initiation is critical in limiting the financial and social damage whereas early recalls result in unnecessary shock to the market and are associated with negative stockmarket returns. (Davidson and Worrell [14]). Accurate estimation of product quality and costs for recall and other alternatives is key for correct and timely decisions, yet we know little about how to effectively integrate valuable information such as customer feedback into the decision making.

Take the General Motors's Cobalt recall for example: the company was aware of the faulty ignition switch as early as 2005, yet persistently denied responsibility for unexpectedly high accident rates. GM did not initiate a voluntary recall until it was facing a class action law suit. One source of the company's stubbornness was a balance-analysis using ten-year old data (Maiorescu [26]). Since product quality is defined by customer satisfaction (Krajewski et al. [21]), we use customer feedback to update the estimation of product defect rate for a better understanding for recall costs.

A classic approach for modeling and optimizing a dynamic process is the dynamic programming (DP) method. We use beta distribution to estimate product defect rate and binomial distribution for the number of product returns of each period that provides the conjugate property to facilitate the calculation of posterior distribution and expectations' calculation. We focus on the following tasks: identify the probability distributions for modeling product defect rate and customer feedback process; model and solve the product recall optimization problem with DP method; explain the optimal solution found and use its structure to explore alternative methods other than DP to solve large-size problems.

The second essay, "Newsvendor Problem using Quality Investment for Product Recall Risk Control", studies the optimization problem that integrates product recall concerns with the production quality control by extending the classic newsvendor model. We capture the probability of product recalls using a decreasing function of the level of product quality.

The subject firm controls product quality by focused manufacturing investment — this impacts production cost. Our goal is to maximize total expected profits. We obtain an objective function of sales, operational costs and cost of recall risks. What's more, we have identified necessary conditions for the objective to be negative semi-definite.

Using parametric analysis of this extended newsvendor model we observe two interesting features. First, production quantity and quality levels seem to have conflicting effects — one waxes and the other wanes in optimal solutions for most instances of parameter changes. Second, increasing profitability discourages investment in quality — this result is counterintuitive to our initial expectations.

We further extend our model (which focuses on internal supply) by examining the case of external supply from multiple sources in which two external suppliers satisfy independent demands and cover each other's demand only when the other is having recalls. Our results suggest little impact from recall-covering interaction on optimal solutions.

The third essay, "Optimal Facility Location to Mitigate Product Recall Risks", aims to minimize the aftermath of major product recall events with a closed-loop network design optimization model. This chapter was written under the supervision of Dr. Kai Huang, McMaster University. Our motivation comes from the 2012 XL. Inc. beef products recall which became the largest meat recall in Canadian history. We found the management of this product recall to be shockingly poor. Hundreds of tons of beef products — usable for other purposes such as rendering — went directly to landfills.

In the moment of crisis, the company managing a major recall is usually unwilling or unable to properly coordinate the reverse flow of returned products. We propose the two-stage facility location model extended to include reprocessing and disposal facilities for management of both forward and reverse flows. Our research connects the two areas of disruption management and closed-loop network design.

Disruption management uses facility location models to ensure the consistent supply of products to satisfy customer demand and protects the system against unexpected and rare disasters such as earthquake and tsunami (Qi et al. [37]). Closed-loop network design aims to minimize the long-run average cost of forward and reverse flows by improving the use of recyclable production components (Fleischmann et al. [17]), decreasing the collection costs (Pishvaee et al. [34]), and increasing service rates and customer satisfaction (Min and Ko [31]).

The existing literature treats reverse flow for daily operations but does not serve the properties of major product recalls well. Our challenge is to design an optimal network that can accommodate product returns in the context of major product recalls. We propose a two-stage stochastic mixed integer programming model, in which we locate the manufacturing plants in the first stage and the reprocessing and disposal facilities in the second stage. We adopt a scenario-based approach to describe the uncertainty of major recall events that may happen in manufacturing plants as well as the availability of reprocessing facilities. Given the complexity induced by our nested facility location problem, we devise an algorithm based on Lagrangian relaxation to solve the uncapacitated case.

To summarize, my dissertation research makes the following primary contributions. In the first essay, we model and solve the optimization problem of product recalls by combining dynamic programming with the Bayesian conjugate property of beta distributions and Bernoulli processes. The flexibility of the beta distribution in modeling various shapes and ranges of probability distributions well serves the potential estimation of product defect rate. We can reasonably assume the process of customer discovery of defective products is composed of independent and identical events and follows the Bernoulli process. Combining a beta distribution and a Bernoulli process enables the Bayesian conjugate property that posterior distribution remains a beta distribution whose parameters follow a simple updating rule.

Moreover, we define the threshold curve as the connection of states of largest allowable customer returns — one more returned product switches the optimal decision from continuing to initiate a recall — of all stages in the optimal solution. The threshold curve for the optimal decision shows a non-decreasing trend that crosses the origin. This observation inspires our approximation of root functions and the application of using the simulation method for large sized problems. Overall, our experiments show the proposed DP model and simulation method could solve various sizes of problems in a satisfying balance of solution accuracy and computation time.

The main contribution in the second essay is that we model and solve the optimization problem of product recall management by incorporating quality control into the classic newsvendor problem. We first examine the internal supply case in which the core company has direct control of quality level. Concavity analysis shows the objective is neither concave nor convex. But our numerical experiments suggest that with the chosen function forms, namely exponential and Erlang, the objective function has one stationary point of local minimum and one stationary point of global maximum.

We extend our model to multiple external suppliers. Both suppliers satisfy their own demands independently and supply for each other's customers only when the other party has a recall. Our numerical experiments show that the prospect of covering the other party's demand in case of recall has very little impact on either supplier's optimal ordering quantity and quality level.

My third essay contributes to the literature by designing an optimal network that can accommodate product returns in the context of major product recalls. We design a two-stage stochastic mixed integer programming model, in which we locate the manufacturing plants in the first stage and the reprocessing and disposal facilities in the second stage. Our results of comparing total cost and computation time in the search of optimal modeling setting based on the minimax regrets method suggest our proposed model minimizes worst-case regrets. Moreover, considering potential product recalls reduces total costs in the long run — disregarding potential recalls could lead to selection of plant locations that initially seem to minimize costs, but that in hindsight are risky candidate sites with high expected costs to handle possible recalls.

Overall, my dissertation demonstrates that it is effective to use operations research methods to handle the unique features of product recalls. Our dynamic programming model suggests that utilizing the customer feedback information to estimate product defective rate and treating product recall decisions in a dynamic framework avoids the over reaction of early recall and the heavy expense of late recall. Our extended newsboy model introduces the dimension of product quality whose impact on recall probability offers the opportunity to reduce product recall risks by investing in manufacturing. Our closed-loop facility location model promotes the integration of recall risks in locating manufacturing plants thus reduce the worst-case regrets.

The rest of the thesis proceeds as follows. Chapter 2 models the timing decisions for initiating product recall in a dynamic process and applies the conjugate property of beta distribution and Bernoulli process to solve for optimal. Chapter 3 extends the newsvendor problem to incorporate quality induced product recall concerns and optimize decisions of ordering quantity and quality levels. Chapter 4 builds a two-stage stochastic location-allocation model to optimize long-run average cost of forward and reverse flows to properly manage the returned products from recalls. Chapter 5 summarizes the thesis and suggests directions for future research.

Chapter 2
Product Recall Timing Optimization using Dynamic Programming

In this chapter, we treat the product recall timing as a dynamic process and use the information of customer returns of each period to update the perceived product defect rate. Higher defect rate indicates a high risk for the future including high return maintenance and likely product recall costs, for which immediate recall can be a wiser choice than continuing. Based on the system state, we aim to find thresholds that above which initiating recall is the optimal decision. We first develop an optimal stopping model with fixed defect rate shared by all periods and solve the problem with dynamic programming (DP) technique. Then we extend the model using defect rate updated by the number of product returns of preceding period and solve it with DP method. We show computing complexity increases dramatically with the problem size, thus implementing DP method is unrealistic for practical problems with large data. We use simulation method of parametric optimization to select the best fitting function form and parameter for the threshold curve. Also, we will explore the possibility of using approximate dynamic programming to solve our proposed model.

2.1 Introduction

Making decisions on when to initiate product recalls should be a dynamic process that uses information from customer feedback. Traditionally, decision makers choose recall timing using prior estimation for potential damage caused by defective products. Because the prior estimation is based on historical data of similar products and manager's subjective opinions, it could result in wrong decisions. Estimation inaccuracy may be due to fallible subcomponents that significantly limit the effectiveness or safety of product usage, ignorance of negative side-effects from long term usage, or disregard for seemingly unlikely events that trigger critical social or economic damage. However, using information from customer feedback after a product's market release helps fix this inaccuracy.

When facing the issue of having defective product released in the marketplace, managers make the decision to initiate a product recall with the goal of optimizing both the company's short-term and long-term costs. Information leading to the recall may originate within the company or from customer feedback. Internal quality systems and external audits help firms identify design and production problems in a structural way.

Many firms have comprehensive quality systems that involve internal processes and external audits to ensure product safety. Depending on the nature of the product, quality standards may stipulate 100% inspection and test, or may measure adherence to specifications based on statistical sampling before shipping products to customers. For example, firms typically use sampling for batch processes or large unit volumes, whereas 100% testing is more appropriate when manufacturing specialized equipment, such as custom engineered equipment or premium products such as luxury cars.

Firms usually complement internal quality processes, either by choice or by law, with external quality audits. For example, companies may value the reputational benefit of having its manufacturing facilities certified by a third party company (e.g., International Standards Organization (ISO) 9001). Once a firm has achieved ISO 9001 certification at a facility, it would typically pay for an ISO inspector to audit that facility on a regular basis (e.g., every six months). The inspector would report back to the firm on any issues and work with the plant to ensure continued compliance with the ISO standards. The regular review process might be a source of information that reveals a potential recall situation. The second type of external inspection derives from laws to protect public safety. Countries have standards organizations (e.g., Canadian Standards Association (CSA) in Canada and Underwriters Laboratories (UL) in the United States). In Canada, for example, firms submit products for testing and CSA approval before being able to sell them in the marketplace. On an ongoing basis, CSA sends its inspectors without prior notice to check manufacturing process compliance and to select product samples for off-site testing. If they uncover problems, CSA inspectors can stop production and quarantine inventory.

Although reliable and effective in detecting foreseeable design and production issues, internal quality systems and external audits are not able to eliminate the product failures or the inaccuracy of prior estimations mentioned earlier.

There are three possible reasons that products can pass internal or external quality inspections but still result in serious safety concerns leading to product recalls after products reach the market.

Firstly, changes in the using environment could lead the product to fail or cause safety issues (i.e., items that meet specifications during the tests, but fail sometime after being sold). For instance, as described by Beauchamp and Littlefield [4], the 1998 recall initiated by Maple Leaf Inc. (MLF) was caused by Listeria contamination in one of its cutting machines. Internal testing procedures of MLF failed to identify this problem due to the light contamination and low safety standards at the time, yet the natural growth of the bacteria posed serious danger to customers' health upon consumption. Secondly, unanticipated problems do not have tests designed to detect them. For example, toys with small breakable parts may send children to the emergency room over choking hazards. Thirdly, components may fail in certain situations that are not taken seriously, or not considered during internal tests. For example, in the General Motors product recall of 2014 and 2015, the company overlooked the problem that the ignition switch under vibration may cause power steering to fail. In their internal tests, the car model operated normally for long durations under controlled situation. But in real use, if there is vibration such as that is caused by a heavy key chain moving with traffic, it could result in a fatal accident. Valukas [51] provided a detailed report on this incident. To sum up, it is justifiable to assume imperfect product in terms of quality.

Furthermore, some of the potential quality issues could lead to significant damage to society and to the company's profits and reputation. When the company believes that it has sold products that could pose a threat to public health or welfare or damage to its brand or reputation, it will initiate a product recall. Based on the manager's estimation of potential risks, he may choose to recall at any time after releasing products to the market. The decision periods could cover the entire lifetime of the product, or cover the warranty period offered by the company. Here we use the warranty time for the sake of simplicity.

Theoretically, there exists an optimal timing to initiate a product recall. Early recall actions result in unnecessary shock to the market and are associated with negative impacts on company revenue and stock markets performance. See studies by Jarrell and Peltzman [20] in drugs and automobile industries, Davidson and Worrell [14] for the other industries. Delayed recalls, on the other hand, may results in massive negative media coverage and liability costs, and could add serious pressure to the company and severe reputational damage.

Here, we treat the decision of recall timing as a dynamic process and use customer feedback to update estimation of product defect rate for a better picture of expected costs. With this approach, we aim to design a sequential analysis model to discover the structure of optimal policy for product recall timing problem.

The rest of the chapter is arranged as follows: Section 2.2 provides a brief review of related literature on product recall timing; Section 2.3 models the problem with optimal stopping problem and assumes defective rate as a constant, and provides solutions with dynamic programming and parameter analysis; Section 2.4 extends the model with updating defective rate based on product returns obtained from previous periods, and shows solution with dynamic programming and compares results of Section 2.2; and Section 2.5 proposes directions for future study.

2.2 Literature Review

This survey of papers from the past twenty years shows a gap in the literature on product recall timing optimization.

Daughety and Reinganum [13] are among the first to consider products with imperfect quality and unobservable safety status. They assume unsafe products can cause injury to customers and incur liability-related costs to the firm. They develop a monopoly model of production planning and product safety signaling, aiming to find the equilibrium that can balance the expected liability cost of unsafe products and high initial production cost of high safety standard products. The firm can observe its product risk and the probability that product use can cause injury or damage, but this information is unobservable to customers. However, the firm's pricing decision affects both the customers' perception of product safety and demand. If an injury happens, the company needs to cover liability costs including direct and indirect costs of lawsuits. The authors assume safer products can reduce marginal expected liability cost, but will increase manufacturing cost. They find that if marginal total cost largely depends on production cost, then the company signals via high price; and if total cost is largely decided by expected liability cost, then the company signals through volume.

Noticeably, the magnitude of capital market punishment on recall announcements for involuntary actions (compared to voluntary actions) is not consistent in relevant studies. Jarrell and Peltzman [20] use event study methodology to examine the impact of producing defective products on stakeholder wealth in the drug and automobile industries. They find that capital markets severely penalize companies with recalls and, thereby, create a considerable deterrence against producing faulty products. The authors also show that spillover effects impact other production lines of the recall company and the whole industry as well. Davidson and Worrell [14] examine abnormal returns caused by product recalls in industries other than drugs and automobiles. They show abnormal returns are significant upon recall announcements and are more negative when products are replaced than being checked and repaired. Their results do not provide statistically significant evidence that government-ordered recalls cause more abnormal returns than voluntary recalls. However, Thirumalai and Sinha [49] conclude that financial markets are indifferent to recall announcement. They analyze empirically the recall data of medical devices from 2002 to 2005 to find the financial consequences of defective devices and the firm characteristics that are determinants of recalls. Unlike findings in previous literature, they find that the financial market does not impose significant deterrence to producing defective products. They also discover that firms can learn from their previous recall experience as analysis indicates reduced recall likelihood.

In their review, Maruckeck et al. [29] identify three research opportunities in product recall management, including identifying a product recall problem, mitigating recall risks and learning from recall. The closest issue related to recall timing is timely communicating recall messages through the supply chain, which is much different than our product recall timing problem.

Research on product recall timing has emerged only recently. The following papers examine product recall timing using empirical or analytical approaches. Hora et al. [19] pursue the reasons for lengthy recalls in the US toy industry. They measured the time taken to initiate product recall by the difference between the dates of recall announcement and first product sale. Their empirical analysis shows recall time is impacted by recall strategy (preventative vs. reactive), recall reason (design flaws vs. manufacturing defects) and recall firm's position in the supply chain. With all other factors equal, companies using preventative strategies, recalling due to design flaws, and having an upper position in the supply chain take longer to recall due to operational difficulties and larger responsibilities.

With mechanism design approach, Chao et al. [8] adopt a threshold time-of-product recall-initiation for the design of recall cost-sharing contracts. They propose two new contractual agreements of recall cost sharing schemes to coordinate quality improvement efforts made by two parties of the supply chain (manufacturer vs. supplier). Both contracts are based on root cause analysis, a method that accurately reveals the party responsible for the recall. Analysis begins at the earliest expiration of sold products. Contract S (selective root cause analysis) uses root cause analysis only when a recall happens before a threshold time and allocates all costs to the responsible party; after the threshold time, recall cost is shared with a fixed rate. Contract P (partial cost allocation), on the other hand, uses root cause analysis regardless of the time of recall occurrence and always shares the cost between two parties with the responsible party incurring more recall cost. Their results show root-cause-analysis is not necessary when information is perfect and costs nothing. When the information cost is not negligible, they show Contract S is consistently better than both Contract P and fix-rate contracts in improving supply chain performance and product quality.

The most relevant study of recent years is done by Sezer and Haksöz [46] who treat product recall management as a continuation of quality control, and use an optimal stopping model to address when to initiate a recall for a dyad supply chain. During the seller's manufacturing, a production fault could happen randomly that affects product lifetimes. Upon expiration of each sold product, the seller decides whether to initiate a recall. If no recall action is taken, a public inspection takes place and detects the production fault with certain accuracy. If found at fault, the seller receives a fine that is more expensive than the cost of a voluntary recall. The authors capture the random factors with different approaches, including using state of nature for the state of manufacturing fault, using exponential distributed random variable for product expiration time, and using fixed rate probability for a successful detection rate of public inspection. They solve the problem with dynamic programming after applying smaller filtration and likelihood ratio process to the model. Besides the general model, they also examine optimal solutions for a static model of a single product, a general case of multiple but finite products, and a special case of infinite products. Finally, the authors explore two extensions of the general model. The first considers conditional public inspection that only takes place when the lifetime is shorter than expected. The second assumes observable manufacturing faults that the seller can detect.

We construct the mathematical model for recall timing problem as one of the optimal stopping. But unlike Sezer and Haksöz [46] who use techniques of sequential hypothesis testing of Poisson processes, we intend to use the conjugate property of the beta distribution and Bernoulli processes to integrate customer feedback information into recall decision making.

The literature shows recall costs are composed of various factors. Berman [6] summarizes lists of both direct and indirect recall costs incur for product recalls.

Direct recall costs are positively linked to recall size (i.e., the number of products to be collected) and whether the recall is voluntary. Berman [6] lists the direct costs of product recalls, including communication costs, product disposition costs, and overhead. Regulation costs can also be part of product recalls. For instance, Hooker et al. [18] identify food safety regulation costs, which vary by plant size, in the food sector. Sezer and Haksöz [46] use product price, P, as the variable cost for voluntary recall and authority inflicted a big fine, K, as the variable cost for involuntary recalls. In their numerical example, P equals 1.5, while K equals 100. Min [30] considers transportation costs as part of recall costs and measures the loss of customer goodwill by the time required to finish a recall.

Indirect costs include the lost sales or revenue and long lasting negative effects on demand (Marsh et al. [28]), potential damage to financial health (Marino [27] and Welling [52]), product liability risks (Thomsen and McKenzie [50] and Salin and Hooker [43]), and possible severe impact of litigation loss (Marucheck et al. [29]). Customers may switch to other brands or other product types if the company's products are perceived unsafe; for instance, Marsh et al. [28] find demand for meat products dropped during 1982-1988 after consumers responded to meat product recalls by switching to meat substitute products. Marino [27] and Welling [52] show the probability of both civil and criminal charges increase with the length of time that unsafe products stay in the market. Thomsen and McKenzie [50] and Salin and Hooker [43] find strong evidence that recalls are associated with significant decreases in share price. Jarrell and Peltzman [20] show indirect costs of recall are likely to be far more than direct costs.

Hora et al. [19] argue that prompt recall initiation reduces operational costs for three reasons. (1) Earlier defect detection allows the company to fix similar problems in unsold products; (2) Variable costs to collect defective products may be lower since products are more likely located at downstream intermediaries instead of end customers; (3) fewer liability risks since defective products stay in the market for a shorter time. However, their conclusion on direct recall costs works better with continuous production and release of products and, as such, does not fit for our problem setting.

Our recall cost setting has two parts, variable costs dependent on the number of products remaining in the market and a one-time fixed cost for initiating a recall. We can use variable costs to indicate direct recall costs that largely depend on the recall magnitude (i.e., the number of products to be collected). Fixed costs indicate the indirect recall costs, which are larger than the variable costs and increase with time and the number of products returned.

2.3 Model with Stationary Product Defect Rate Distribution

In this section, we explain the story, construct the model, and develop a numerical case for the product recall timing model, which later extends into a general form with dynamically updated return rate.

Consider a company that sells M products to the market[1]. The products, however, are not all of perfect quality. The firm assures its customers of its products' quality during the warranty time of T periods. Within warranty coverage, if a product is found defective, it will be returned and managed with certain compensation c_1 (per unit) paid to the customer. All returned products contribute to the loss of customer goodwill which is calculated at the end of the decision process with cost c_F per unit. If all products have been returned before the end time T, the process ends prematurely and the cost of goodwill is evaluated on all M products with same unit cost c_F.

To prevent the potentially large costs of customer goodwill loss, the responsible manager (he) can observe the total number of product returned (s_t) at the beginning of any period t, $t = 1, \ldots, T$, and decide whether to stop the process by initiating a product recall. During the recall, an immediate cost K will occur along with the costs of collecting products remaining in the market with c_0 per unit.

In order to facilitate the decision, he can estimate the return rate $\tilde{p}$, i.e., the probability of a randomly selected product being defective, based on historical data of similar products. Suppose the return rate $\tilde{p}$ follows a beta distribution of parameters k and n ($k \leq n$) estimated from historical data, i.e., $\tilde{p} \sim \text{Beta}(k, n)$, and the probability density function is

[1] Specifically, we assume these M units belong to one production lot of the same product type. If the firm produces other lots of this type of product or other product types, product recall decisions of those lots will follow similar and independent decision making processes.

Notations	**Definitions**
c_0	unit cost of recall products in market,"0" marks system halts
c_1	unit cost of managing product returns, "1" marks system continues
c_F	unit cost of goodwill loss, "F" marks the process finishes
K	immediate cost of recall action
$k_t,\ k$	parameter of beta distribution for product defect rate in period t
M	initial number of products in market
m_t	number of products remaining in the market at the beginning of period t
$n_t,\ n$	parameter of beta distribution for product defect rate in period t
$\tilde{p}_t$	random variable estimating product defect rate in period t
$\tilde{r}_t$	random variable estimating number of product returns in period t
s_t	number of products returned so far at the beginning of period t
T	total number of time periods

Table 1. Notations and meanings for parameters and variables in product recall timing optimization models.

$$f_{\tilde{p}}(p) = \frac{(n-1)!}{(k-1)!(n-k-1)!} p^{k-1}(1-p)^{n-k-1} \text{ for } 0 \leq p \leq 1. \tag{1}$$

Since the manager decides at each stage in a dynamically changing process that evolves through time, this decision process belongs to the category of sequential analysis. To be specific, our problem of focus is similar to the optimal stopping problem. Therefore we adopt a dynamic programming approach and use the principle of optimality to construct the model. We introduce notations in Table 1 to facilitate model formulations.

2.3.1 Model Building

In this section the manager's estimate of the product rate $\tilde{p}_t$ follows a beta distribution indifferent to time periods, i.e., $\tilde{p}_t \sim \text{Beta}(k, n)$, for which the expectation $E(\tilde{p}_t)$ is k/n. We use k and n to denote the beta distribution parameters instead of k_t and n_t since their values are constant through all time periods. The number of products remaining in market (m_t) and the total number of returned products (s_t) abide the relationship $m_t + s_t = M$. The number of products returned $\tilde{r}_t$ in period t follows the binomial distribution since the process is similar to conducting m_t identical

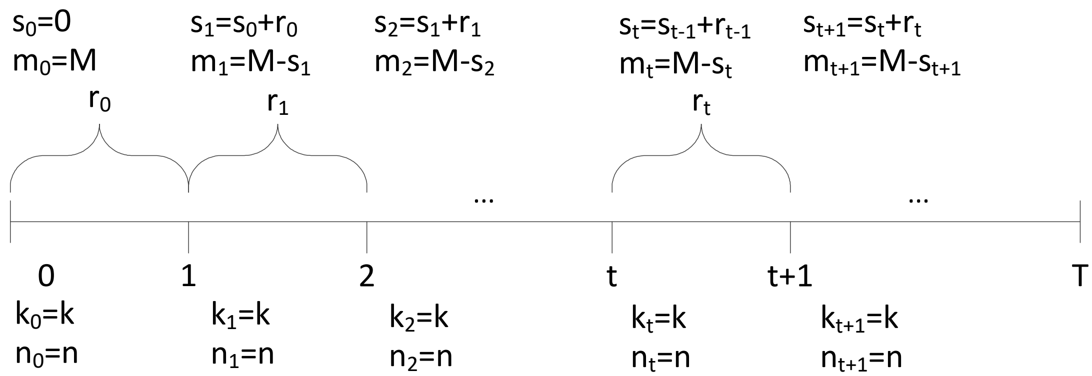

Figure 1. An illustration of the product return process in the constant defect rate model.

Bernoulli processes with success rate $\tilde{p}_t$, i.e., $\tilde{r}_t \sim \text{Bin}(m_t, \tilde{p}_t)$.

In this binomial distribution, we use a random variable as the estimation of event occurrence rate instead of a fixed probability because the true value is unknown. Instead, the manager can estimate the distribution of the defect rate from historical data. Therefore, we use this beta-distributed random variable $\tilde{p}_t$ for defect rate at the tth period with which we generate the estimation of the number of product return $\tilde{r}_t$ accordingly. In our extension of the dynamically updating defect rate model in the following section, the defect rate estimation from historical data is treated as prior information and updated each period using obtained value of product returns. An illustration of the process is shown in Figure 1.

Using the principle of optimality, we model this dynamic programming process with value function $V_t(s_t)$ for $t = 0, 1, \ldots, T-1$ and terminal value $V_T(s_T)$ in expressions (2-3).

$$V_T(s_T) = c_F s_T \tag{2}$$

$$V_t(s_t) = \begin{cases} \min \begin{cases} c_0(M - s_t) + K & \text{Recall} \\ c_1 E(\tilde{r}_t) + E[V_{t+1}(s_t + \tilde{r}_t)] & \text{Continue} \end{cases} & \text{If } s_t < M \\ c_F M \qquad \text{Stop} & \text{If } s_t = M \end{cases} \tag{3}$$

where $\tilde{r}_t \sim \text{Bin}(m_t, \tilde{p}_t)$ and $\tilde{p}_t \sim \text{Beta}(k, n)$. Terminal value $V_t(s_t)$ in (2) shows that the cost of goodwill loss is linear to the total number of product returns if the process finishes the entire warranty time. Depending on the number of returns at the beginning of period t, the value function $V_t(s_t)$ in (3) takes different forms. If all the products have been returned, i.e., s_t equals M, the process terminates with the cost of goodwill loss proportional to total returns M. Otherwise, the manager evaluates the choices between immediate recall costs and the expected cost to continue. If he decides to recall, a variant cost proportional to products remaining in the market takes place along with a fixed cost K. If he decides to continue, he expects to take the cost of managing returns of the tth period along with expected costs of the following periods $E[V_{t+1}(s_t + \tilde{r}_t)]$.

Using the property of conditional expectation which is well exemplified in Ross [39], $E(X) = E[E(X \mid Y)]$, we calculate the cost incurred to manage product returns at stage t, given system state s_t, in the following

$$\begin{aligned} c_1 E(\tilde{r}_t) &= c_1 E[E(\tilde{r}_t \mid \tilde{p}_t)] \\ &= c_1 \int_0^1 E(\tilde{r}_t \mid \tilde{p}_t = p) f_{\tilde{p}}(p)\ dp \\ &= c_1 \int_0^1 p m_t f_{\tilde{p}}(p)\ dp \\ &= c_1 m_t \int_0^1 p f_{\tilde{p}}(p)\ dp \\ &= c_1 m_t E(\tilde{p}_t) \\ &= c_1 (M - s_t) k/n \end{aligned} \tag{4}$$

where $f_{\tilde{p}}(p)$ is the probability density function of beta distribution for $\tilde{p}_t \sim \text{Beta}(k, n)$ as in (1), and other parameters are explained in Table 1.

By the law of total probability, the probability of getting r products returned in period t is

$$\begin{aligned} \Pr(\tilde{r}_t = r \mid s_t) &= \int_0^1 \Pr(r \mid \tilde{p}_t = p, s_t) f_{\tilde{p}}(p)\ dp \\ &= \int_0^1 \binom{M - s_t}{r} p^r (1-p)^{M - s_t - r} f_{\tilde{p}}(p)\ dp \end{aligned} \tag{5}$$

and the expectation of cost-to-go function starting from stage $t+1$ depends on s_t only, with the expression as follows

$$E[V_{t+1}(s_t + \tilde{r}_t)] = \sum_{r=0}^{M-s_t} V_{t+1}(s_t + r) \Pr(r \mid s_t). \tag{6}$$

Since value function for the process end T is given in (2), recursively calculating the expected costs of remaining periods $E[V_{t+1}(s_{t+1})]$ with (6) and value function $V_t(s_t)$ with (3) will determine the optimal decision for any possible state s_t at any time stage t.

2.3.2 Solution using Dynamic Programming

The classic dynamic programming (DP) procedure can solve the static product defect rate distribution model with ease and result in an optimal policy that works for any possible state of the system. This policy assists decisions of whether to initiate a product recall or continue with the process based on the states observed in the current stage and discovers a critical level of the system state at which the best decision switches from "CONTINUE" to "RECALL". We call this critical level the threshold, which is the largest system state that a "CONTINUE" decision remains optimal for the stage.

2.3.2.1 The Threshold Curve

Firstly, we examine the threshold from the last period θ_{T-1} to gain some insight of DP solving procedures.

Proposition 1 *Recall threshold for the last time period is determined by given parameters from Table 1 and beta parameters k and n with following equation:*

$$\theta_{T-1} = \left\lfloor \frac{K + [c_0 - (c_1 + c_F)\, k/n]\, M}{(1 - k/n)\, c_F + c_0 - c_1 k/n} \right\rfloor \tag{7}$$

Proof. Given the definition of recall threshold, θ_{T-1} is the largest number of return that "CONTINUE" remains optimal decision, i.e., the cost to continue is equal to or less than the cost to recall. From the model (2-3), certain states s_{T-1} satisfies the following inequality

$$c_0\,(M - s_{T-1}) + K \geq c_1 E\,(\tilde{r}_{T-1}) + E\,[V_T\,(s_{T-1} + \tilde{r}_{T-1})]$$

From analysis in (4), we can write $E\,(\tilde{r}_{T-1}) = k/n\,(M - s_{T-1})$, therefore

$$\begin{aligned} c_0\,(M - s_{T-1}) + K &\geq c_1 \frac{k}{n}\,(M - s_{T-1}) \\ &\quad + E\left[V_T\left(s_{T-1} + \frac{k}{n}\,(M - s_{T-1})\right)\right] \\ \left[\left(1 - \frac{k}{n}\right) c_F + c_0 - c_1 \frac{k}{n}\right] s_{T-1} &\leq K + \left[c_0 - (c_1 + c_F)\,\frac{k}{n}\right] M \end{aligned}$$

It is reasonable to assume that unit cost incurred during a recall is higher than that for managing returns in a "CONTINUE" decision, i.e., $c_0 > c_1$. Since the ratio k/n is equivalent to the expectation of defect probability $E(\tilde{p})$, $0 < k/n < 1$. Hence the coefficient of s_{T-1} is positive, and we can solve the above inequality for all eligible s_{T-1}.

$$s_{T-1} \leq \frac{K + [c_0 - (c_1 + c_F) k/n] M}{(1 - k/n) c_F + c_0 - c_1 k/n}$$

Because recall threshold corresponds to the largest eligible s_{T-1} solved, the result in (7) holds. ■

Now extend the above procedure to compute thresholds prior to the last periods. When $0 \leq s_t \leq M - 1$ for $t = 2, \ldots, T - 1$, let function $A(s_t) = c_0 (M - s_t)$ denotes recall costs, function $B(s_t) = c_1 E(\tilde{r}_t) = c_1 (M - s_t) k/n$ denotes return management costs, and function $w_t(s_t) = E[V_{t+1}(s_t + \tilde{r}_t) \mid s_t < M]$ denotes expected costs for all future periods since $t + 1$ if continue, we have

$$V_t(s_t \mid s_t < M) = \min \begin{cases} A(s_t) + K \\ B(s_t) + w_t(s_t) \end{cases}$$

Since cost-to-go for stage $t + 1$ is

$$V_t(s_t + \tilde{r}_t) = \begin{cases} \min \begin{cases} A(s_t + \tilde{r}_t) + K \\ B(s_t + \tilde{r}_t) + w_{t+1}(s_t + \tilde{r}_t) \end{cases} & \text{If } \tilde{r}_t < M - s_t \\ c_F M & \text{If } \tilde{r}_t = M - s_t \end{cases}$$

we have the following expression for expected cost to continue $w_t(s_t)$:

$$\begin{aligned} w_t(s_t) &= \sum_{r=0}^{M-s_t-1} \Pr(\tilde{r}_t = r \mid s_t) \min\{A(s_t + r) + K, B(s_t + r) + w_{t+1}(s_t + r)\} \\ &\quad + \Pr(\tilde{r}_t = M - s_t \mid s_t) c_F M \end{aligned}$$

The company initiates recall only if $A(s_t) < B(s_t) + w_t(s_t)$. Let function $G_t(s_t)$ refer to the difference of recall costs and return management costs, i.e.,

$$\begin{aligned} G_t(s_t) &= A(s_t) - B(s_t) \\ &= \left(c_1 \frac{k}{n} - c_0\right) s_t + \left(c_0 - c_1 \frac{k}{n}\right) M + K \\ &= \alpha s_t + \beta \end{aligned}$$

where $\alpha = c_1 k/n - c_0$ and $\beta = (c_0 - c_1 k/n) M + K$. Consequently, the firm initiates product recall only when

$$G_t(s_t) = \alpha s_t + \beta < w_t(s_t)$$

Calculating backwards, we obtain

$$w_t(s_t) = E[V_{t+1}(s_t + \tilde{r}_t)] = \sum_{r=0}^{M-s_t} V_{t+1}(s_t + r) \Pr(\tilde{r}_t = r \mid s_t)$$

from previous steps going back to the ending period T.

Conditional on the value of α, ranges of s_t to make product recall decisions are as follows

α value	Initiate recall if s_t satisfies
$\alpha > 0$	$s_t < [w_t(s_t) - \beta] / \alpha$
$\alpha = 0$	$\beta < w_t(s_t)$
$\alpha < 0$	$s_t > -[\beta - w_t(s_t)] / \alpha$

Table 2. Recall decision ranges of system states for different parameter setting.

Using the dynamic programming method, one obtains the value of $w_t(s_t)$ for each state s_t. Using Table 2, the largest state s_t that solves for a "CONTINUE" decision is the value for threshold θ_t at stage t. Thus we obtain the optimal decision.

2.3.2.2 Numerical Experiments

We now describe a numerical example to illustrate the optimal solution and threshold curve for our proposed model (2-3). In this numerical example, with parameters of k and n given in Table 3, the probability of getting $\tilde{r}_t = r$ products returned in period t is

$$P(r \mid s_t) = \frac{3(6 - s_t - r)(5 - s_t - r)}{(7 - s_t)(6 - s_t)(5 - s_t)},$$

given that $0 \leq r \leq M - s_t$ and $0 \leq s_t \leq M$.

With parameters in Table 3, results of value function $V_t(s_t)$ and threshold θ_t in optimal policy μ_t^* given any state s_t of any stage t are listed in Table 4. Note that during each time stage, the cost-to-go increases at first then decreases after switching decision. This happens because the function corresponding to "CONTINUE" decision $(c_1 E(\tilde{r}_t) + E[V_{t+1}(s_t + \tilde{r}_t)])$ increases over state s_t while the function representing "RECALL" decision $(c_0(M - s_t) + K)$ decreases linearly over state s_t. Optimal policy μ_t^* is to choose to "RECALL" when observed total returns s_t exceeds threshold θ_t, and choose to "CONTINUE" if observed total returns s_t is less or equal to the threshold θ_t.

c_0	c_1	c_F	K	n	k	M	T
2	1	3	5	4	1	4	3

Table 3. Initial parameter values for the numerical example.

For instance, at period t equals 1, the optimal action is "CONTINUE" when observed total returns s_t is less or equal to 2 and optimal action switches to "RECALL" when s_t is greater or equal to 3; thus the largest observed returns s_t before switching decision is 2, which is the threshold θ_2.

Figure 2 shows optimal decisions for all states in every stage. The states for which optimal decision changes from "CONTINUE" to "RECALL" form the thresholds. Thresholds curve that reflects the optimal policy is shown as dark round dots in Figure 2.

Changing the value of parameter K only while other parameters remain the same in Table 3, the recall threshold θ_t varies as in Table 5. Recall threshold increase when immediate recall cost increase, which suggest high immediate recall cost discourages the action of product recalls. In our case, when immediate cost is higher or equal to 10, recall threshold is equal to 3 which equals $M-1$; since recall action only take action when $s_{T-1} = \theta_{T-1} + 1$ for $s_{T-1} <= M-1$, this result suggests the decision maker will always choose to continue no matter how many products being returned.

Changing the value of parameter unit cost of goodwill c_F only while other parameters remain the same in Table 3, the threshold θ_t varies as in Table 6. When the unit cost of goodwill c_F decreases, recall threshold θ_t increases and the manager is less willing to recall. The unit cost of goodwill is the company's estimation of long-term impact of product returns to its reputation and customer loyalty. The more a company cares about its market sustainability and long-term profit, the more cautious action it will take and the more willing it is to take a recall action.

When c_F is within the range of $[6, 11]$, the optimal decision for the first period is "CONTINUE". In contrast, when c_F is within the range of [12,30], the optimal decision for the first period is "RECALL". Under this circumstance, the company should not release products in the first place.

2.3.2.3 Sensitivity Analysis

We experiment the impact of comparative magnitude of unit cost of managing return c_1 and unit cost of recall c_0 with four cases: (1) $c_0 \ll c_1$, for instance c_0 equals 2 while c_1 equals 20; (2) $c_0 < c_1$, for instance c_0 equals 2 while c_1 equals 5; (3) $c_0 > c_1$, for instance c_0 equals 5 while c_1 equals 2; (4) $c_0 \gg c_1$, for instance c_0 equals 20 while c_1 equals 2. The parameters settings are shown in Table 7.

t	s_t	$V_t(s_t)$	μ_t^*	θ_t	t	s_t	$V_t(s_t)$	μ_t^*	θ_t
0	0	8.54	CONTINUE	0		0	4.00	CONTINUE	
	0	6.74	CONTINUE			1	6.00	CONTINUE	
	1	7.80	CONTINUE		2	2	8.00	CONTINUE	2
1	2	8.60	CONTINUE	2		3	7.00	RECALL	
	3	7.00	RECALL			4	12.00	STOP	
	4	12.00	STOP						

Table 4. Numerical experiment results of static rate recall timing problem.

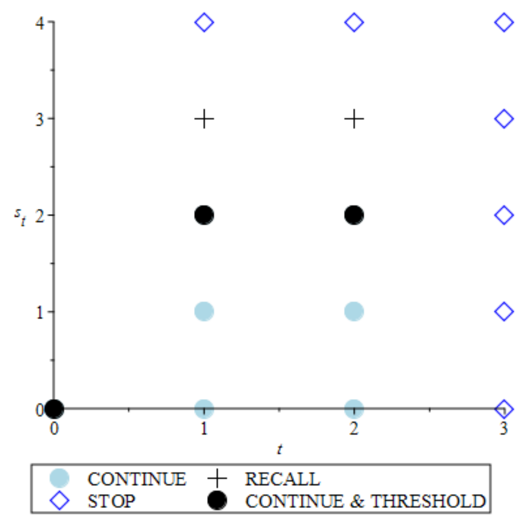

Figure 2. Optimal decisions for all possible states and the threshold curve of the numerical example of four products and three stages.

K	Corresponding threshold
$1,\ldots,4$	$\theta_t = 1$ for $t = 1,2,3$
$5,\ldots,8$	$\theta_t = 2$ for $t = 1,2,3$
9	$\theta_t = \begin{cases} 2 & t = 1 \\ 3 & t = 2,3 \end{cases}$
10	$\theta_t = 3$ for $t = 1,2,3$

Table 5. Recall threshold increases with fixed recall cost K increases.

c_F	Corresponding threshold
1	$\theta_t = 3$ for $t = 1,2,3$
2	$\theta_t = \begin{cases} 2 & t = 1,2 \\ 3 & t = 3 \end{cases}$
3	$\theta_t = 2$ for $t = 1,2,3$
$4,5$	$\theta_t = 1$ for $t = 1,2,3$
$6,\ldots,11$	$\theta_t = 0$ for $t = 1,2,3$
$12,\ldots,30$	$\theta_t = 0^*$ for $t = 1,2,3$

Table 6. Recall threshold decreases with unit cost of goodwill increases.

Common parameters						Case 1		Case 2		Case 3		Case 4	
c_F	K	n	k	M	T	c_0	c_1	c_0	c_1	c_0	c_1	c_0	c_1
3	15	100	1	10	12	2	20	2	5	5	2	20	2

Table 7. Other initial parameter values for numerical examples.

Case 1. Optimal decisions of all possible states are shown in Figure 3 where the threshold curve plotted show a increasing trend of thresholds along with time. This is because the expected cost to continue decreases when there is less time left in the warranty periods.

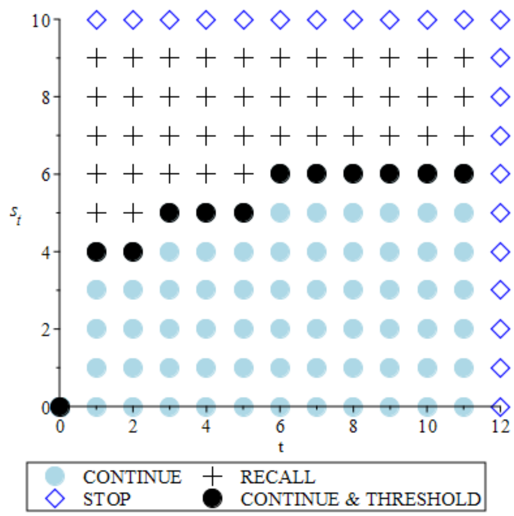

Figure 3. Optimal decisions of all possible states and the threshold curve in case 1: $c_0 \ll c_1$ $(c_0 = 2, c_1 = 20)$.

Case 2. Optimal decisions of all states are plotted in Figure 4 where the threshold curve also shows a increasing trend of thresholds with time. But the slope is much more gentle compared to case 1 and the threshold at time $t = 2$ is higher.

Case 3. Optimal decisions of all states are plotted in Figure 5 where the threshold curve shows a constant ratio of thresholds with time of $t = 1, \ldots, T-1$. Note the threshold curve of $t = 1, \ldots, T-1$ increases to $\theta_t = 8$ compared to $\theta_t = 6$ in Case 2.

Case 4. Optimal decisions plotted in Figure 6 where the threshold curve shows a constant ratio of thresholds with time $t = 1, \ldots, T-1$ which are higher than the results in previous three cases.

Our experiments in four cases show that thresholds are sensitive to the parameter settings of unit cost of managing return c_1 and unit cost of recall c_0. The smaller c_0 is relative to c_1, the steeper the threshold curve as time increases. Otherwise, the thresholds line is flat and higher on the graph when $c_0 - c_1$ gets larger. These results suggest when recall costs increase faster than the return managing cost for "CONTINUE" decisions, managers tend to wait-and-see and take higher recall risks.

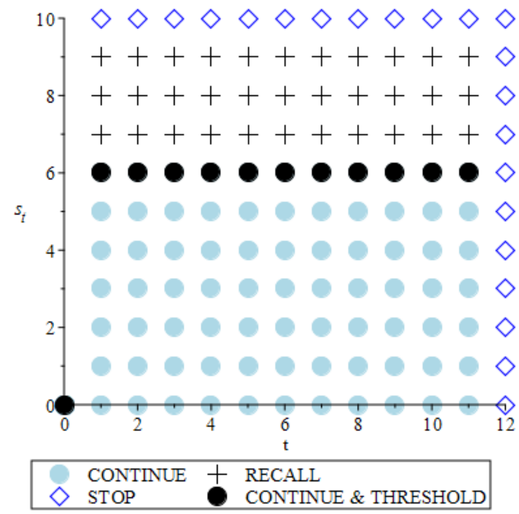

Figure 4. Optimal decisions of all possible states and the threshold curve in case 2: $c_0 < c_1$ $(c_0 = 2, c_1 = 5)$.

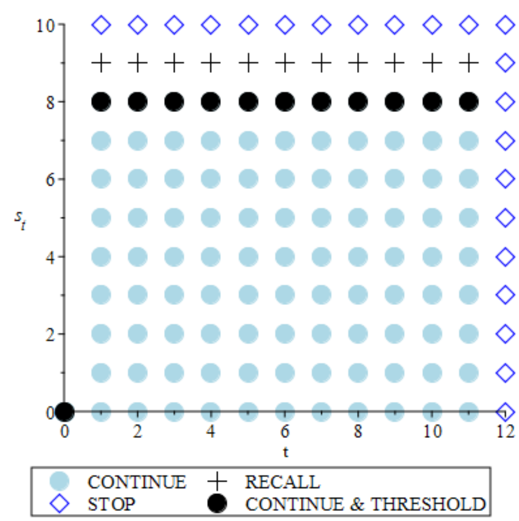

Figure 5. Optimal decisions of all possible states and the threshold curve in case 3: $c_0 > c_1$ $(c_0 = 5, c_1 = 2)$.

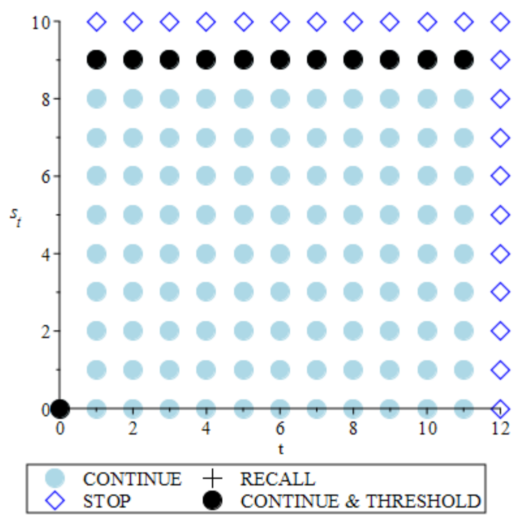

Figure 6. Optimal decisions of all possible states and the threshold curve in case 4: $c_0 \gg c_1$ $(c_0 = 20, c_1 = 2)$.

Adding Kt in Recall Costs Using the same set of parameters same as in the Case 2 sensitivity analysis (shown in Table 8), we compare impacts of using a linearly increasing cost with time (Kt) with a fixed cost (K) in recall costs computation.

c_0	c_1	c_F	K	n	k	M	T
2	5	3	15	100	1	10	12

Table 8. Parameters for numerical example testing impact of Kt versus K case.

Compared to the threshold curve plotted in Figure 4 which show a slowly increasing trend, the linearly increasing cost Kt results in a steeper increasing trend of threshold curve as shown in Figure 7.

The above comparison shows that using the linearly increasing costs Kt permits higher thresholds to recall in later periods than having a fixed value of K in the recall costs. This finding contradicts our assumption, which is inspired by previous studies such as Hora et al. [19], that if recall costs increase with time, managers tend to initiate recalls promptly. Our results show that managers react similarly for early stages given the same number of product returns on hand, but choose to continue the process with a much higher threshold with Kt in recall costs than with K in constant defect rate problems. This result suggests that increasingly higher recall cost adds more inertia and resistance for initiating product recalls.

2.4 Model with Dynamically Updated Product Defect Rate Distribution

In this case, we assume the firm's manager is aware that the number of product returns can help reveal the true defect rate. Based on this belief, he decides to use periodically-collected information on product returns to update his estimation of products defect rate. An illustration of product return process is shown in Figure 8.

2.4.1 Model Building

Defining state variables for stage t as the three dimensional vector (s_t, n_t, k_t) and using the principle of optimality, we model the dynamic programming process with value function $V_t(s_t, n_t, k_t)$ for $t = 0, 1, \ldots, T-1$ and terminal value function

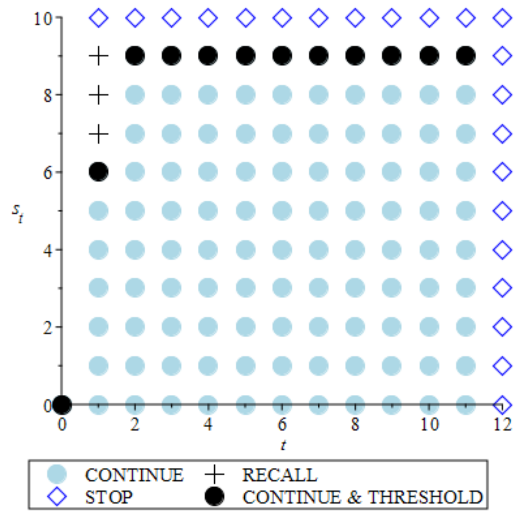

Figure 7. Optimal decisions of all possible states and the threshold curve considering recall costs linearly increasing with time.

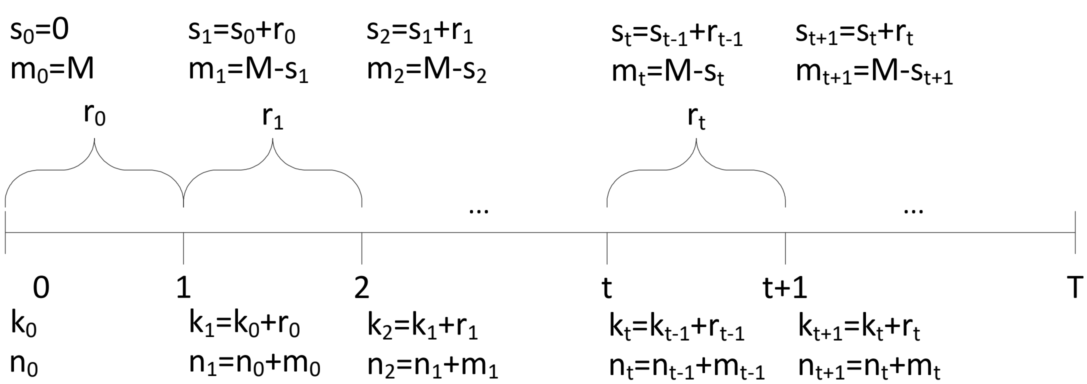

Figure 8. An illustration of product return process for dynamically updating product defect rate model.

$V_T(s_T, n_T, k_T)$ in expressions (8-9).

$$V_T(s_T, n_T, k_T) = c_F s_T \tag{8}$$

$$V_t(s_t, n_t, k_t) = \begin{cases} \min \begin{cases} c_0(M - s_t) + K & \text{Recall} \\ c_1 E(\tilde{r}_t \mid s_t, n_t, k_t) & \text{Continue} \\ \quad + E[V_{t+1}(s_{t+1}, n_{t+1}, k_{t+1})] & \end{cases} & \text{If } s_t < M \\ c_F M \quad\quad \text{Stop} & \text{If } s_t = M \end{cases} \tag{9}$$

where $\tilde{r}_t \sim \text{Bin}(m_t, \tilde{p}_t)$ and $\tilde{p}_t \sim \text{Beta}(k_t, n_t)$. Parameters are defined in Table 1. Definitions of value functions are similar to those of the constant defect rate model. The difference is the state space expands to three dimensional, i.e., (s_t, n_t, k_t), because knowledge of k_t and n_t are necessary to update the defect rate of stage t. As a consequence, expected costs for returns management and the following periods are conditioned on the entire state space (s_t, n_t, k_t).

By the law of total probability, the probability of getting $\tilde{r}_t = r$ products returned in period t is

$$\begin{aligned} \Pr(\tilde{r}_t = r \mid s_t, n_t, k_t) &= \int_0^1 \Pr(r \mid \tilde{p} = p, s_t) f_{\tilde{p}}(p \mid n_t, k_t)\, dp \\ &= \int_0^1 \binom{M - s_t}{r} p^r (1-p)^{M-s_t-r} f_{\tilde{p}}(p \mid n_t, k_t)\, dp \\ &= \frac{\Gamma(r + k_t)\,\Gamma(n_t - k_t + M - s_t - r)\,\Gamma(M - s_t + 1)\,\Gamma(n_t)}{\Gamma(M + n_t - s_t)\,\Gamma(1 + r)\,\Gamma(M - s_t - r + 1)\,\Gamma(k_t)\,\Gamma(n_t - k_t)} \end{aligned} \tag{10}$$

whose complexity contributes to the solving difficulty of the dynamically updated defect rate problem. Gamma function $\Gamma(n)$ is an extension of factorial functions where

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$$

and $\Gamma(n) = (n-1)!$ if n is an integer. Applying the law of total probability, similar to the computing steps in finding $c_1 E(\tilde{r}_t)$ for the constant rate model with formula (4), we have

$$c_1 E(\tilde{r}_t \mid s_t, n_t, k_t) = c_1 \frac{k_t}{n_t} (M - s_t)\,. \tag{11}$$

2.4.1.1 Conjugate property of beta distribution and Bernoulli trials

Suppose in period t the return rate $\tilde{p}_t$ follows a beta distribution with parameters k_t and n_t ($k_t \leq n_t$), i.e., $\tilde{p}_t \sim \text{Beta}(k_t, n_t)$ for which the expectation $E(\tilde{p}_t)$ is k_t/n_t. At the first stage ($t = 0$), prior estimation of return rate follows $\text{Beta}(k_0, n_0)$ for which parameters k_0 and n_0 are exogenously determined by historical data or the decision maker. The number of products remaining in market m_t and the total number of returned products s_t follows the relationship that $m_t + s_t = M$ as mentioned previously. The number of product returns in period t follows a binomial distribution since the process is similar to conducting m_t identical Bernoulli processes with success rate $\tilde{p}_t$, i.e., $\tilde{r}_t \sim \text{Bin}(m_t, \tilde{p}_t)$. The conjugate property of beta distributions and Bernoulli processes states that if the prior distribution follows a beta distribution $\text{Beta}(r, n)$ and the sampling process is Bernoulli with r' successes out of n' trials then the posterior distribution also follows a beta distribution $\text{Beta}(r'', n'')$ where $r'' = r + r', n'' = n + n'$. Winkler [53] provides excellent materials on the conjugate property for those interested, covering proofs and examples. In our case, with initial estimation following a beta distribution and all sampling processes being Bernoulli, we can prove by induction that the resulting posterior distributions are also beta distributions.

With the conjugate property of beta distribution and Bernoulli trials, the system states for the dynamically updated defect rate model update using the following recursive expressions for any $t = 1, 2, \ldots, T$.

$$k_{t+1} = k_t + \tilde{r}_t \tag{12}$$

$$n_{t+1} = n_t + (M - s_t) \tag{13}$$

Therefore given system state (s_t, n_t, k_t) at stage t, we can solve for the expected costs of stages starting from $t + 1$ using expression (14).

$$E[V_{t+1}(s_{t+1}, n_{t+1}, k_{t+1})] = \sum_{r=0}^{M-s_t} V_{t+1}\left(s_t + r, n_t + M - s_t, k_t + r\right) \Pr(\tilde{r}_t = r \mid s_t, n_t, k_t). \tag{14}$$

2.4.1.2 Applying Conjugate Property to States k_t and n_t

Since the total number of returned products at the beginning of period t is $s_t = \sum_{i=1}^{t-1} r_i$, for $t = 1, 2, \ldots, T-1$ and $s_1 = 0$, parameters k_t and n_t have expressions in the following.

$$k_t = \begin{cases} k_0 & t = 0 \\ k_0 + s_t & t = 1, 2, \ldots, T-1 \end{cases} \tag{15}$$

$$n_t = \begin{cases} n_0 & t = 0 \\ n_0 + tM - \sum_{i=0}^{t-1} s_i & t = 1, 2, \ldots, T-1 \end{cases} \tag{16}$$

$$= \begin{cases} n_0 & t = 0 \\ n_0 + M & t = 1 \\ n_0 + tM - \sum_{i=0}^{t-2} (t-1-i) r_i & t = 2, 3, \ldots, T-1 \end{cases} \tag{17}$$

We can prove formulas for k_t and n_t by induction.

Proof. Since k_0 is given, $k_1 = k_0 + r_0$ satisfies the form $k_{t+1} = k_0 + \sum_{i=0}^{t} r_i$. Suppose this formula holds for any ℓ that $1 \leq \ell \leq T-1$, i.e., $k_\ell = k_0 + \sum_{i=0}^{\ell-1} r_i$. Then

$$\begin{aligned} k_{\ell+1} &= k_\ell + r_\ell \\ &= k_0 + \sum_{i=0}^{\ell-1} r_i + r_\ell \\ &= k_0 + \sum_{i=0}^{\ell} r_i \end{aligned}$$

also satisfies the proposed formula.

For the formula of n_t with respect to system states, with n_0 given $n_1 = n_0 + m_0 = n_0 + M - s_0$ which satisfies the formula $n_t = n_0 + tM - \sum_{i=0}^{t-1} s_i$. Suppose this formula holds for any ℓ that $1 \leq \ell \leq T-1$, i.e., $n_\ell = n_0 + \ell M - \sum_{i=0}^{\ell-1} s_i$. Then

$$\begin{aligned} n_{\ell+1} &= n_\ell + m_\ell \\ &= n_0 + \ell M - \sum_{i=0}^{\ell-1} s_i + M - s_\ell \\ &= n_0 + (\ell+1)\, M - \sum_{i=0}^{\ell} s_i \end{aligned}$$

also satisfies the proposed formula.

For the formula of n_t with respect to returns in previous stages, with n_0 given, $n_1 = n_0 + m_0 = n_0 + M$, $n_2 = n_1 + m_1 = n_0 + M + (M - r_0) = n_0 + 2M - r_0$ which satisfies the formula $n_t = n_0 + tM - \sum_{i=0}^{t-2} (t-1-i) r_i$. Suppose this formula holds for any ℓ that $2 \leq \ell \leq T-1$, i.e., $n_\ell = n_0 + \ell M - \sum_{i=0}^{\ell-2}(\ell-1-i)r_i$. Then

$$\begin{aligned} n_{\ell+1} &= n_\ell + m_\ell \\ &= n_0 + \ell M - \sum_{i=0}^{\ell-1} s_i + M - s_\ell \\ &= n_0 + (\ell+1)\, M - \sum_{i=0}^{\ell} s_i \end{aligned}$$

also satisfies the proposed formula.

For the formula of n_t with respect to returns in previous stages, with n_0 given, $n_1 = n_0 + m_0 = n_0 + M$, $n_2 = n_1 + m_1 = n_0 + M + (M - r_0) = n_0 + 2M - r_0$ which satisfies the formula $n_t = n_1 + (t-1)M - \sum_{i=0}^{t-2}(t-1-i)r_i$. Suppose this formula holds for any ℓ that $3 \leq \ell \leq T$, i.e., $n_\ell = n_1 + (\ell-1)M - \sum_{i=0}^{\ell-2}(\ell-1-i)r_i$. Then

$$\begin{aligned} n_{\ell+1} &= n_\ell + m_\ell \\ &= n_0 + \ell M - \sum_{i=0}^{\ell-2}(\ell-1-i)r_i + (M - \sum_{i=0}^{\ell-1} r_i) \\ &= n_0 + \ell M - \sum_{i=0}^{\ell-2}(\ell-1-i)r_i + (M - \sum_{i=0}^{\ell-2} r_i - r_{\ell-1}) \\ &= n_0 + (\ell+1)\, M - \sum_{i=0}^{\ell-1}(\ell-i)r_i \end{aligned}$$

also satisfies the proposed formula. By induction, we show the formulas (15), (16) and (17) hold for $t = 2, 3, \ldots, T-1$. ■

2.4.1.3 Identifying State Space

The notion of "curse of dimensionality" was coined by Bellman [5]. This innate feature is as old as the technique of dynamic programming itself. Exponentially increasing state space makes the calculation for value functions very difficult, if not impossible, when the problem size increases. To deal with this curse, researchers such as Powell [35] apply approximation methods to constrain the exploding dimensions. Powell [35] categorizes three types of curse of dimensions, including state space, action space and decision space.

Fortunately, in our problem we can reduce the dimensionality by observing the connections of states s_t, n_t and k_t. Given the number of product returns until time t, i.e., s_t, beta distribution parameter k_t is decided by $k_0 + s_t$ for all $t = 0, 1, 2, \ldots, T-1$

since $s_0 = 0$, but n_t could be a range if time t is at stage 2 or after. Using $\underline{n_t}$ and $\overline{n_t}$ to denote the lower and upper bound of the range for n_t, then

$$\overline{n_t} = n_0 + tM, \quad \forall t \geq 2 \tag{18}$$

when $s_i = 0$ for all $i = 1, 2, \ldots, t-1$.

$$\underline{n_t} = n_0 + tM - (t-1)s_t, \quad \forall t \geq 2 \tag{19}$$

when $r_0 = s_t$ and $r_i = 0$ for all $i = 1, \ldots, t-2$, i.e., $s_i = s_t$, for all $i = 1, \ldots, t-1$.

$$\begin{aligned} \underline{n_t} \quad &= \quad \min \; n_0 + tM - \sum_{i=0}^{t-2}(t-1-i)r_i \\ \text{subject to} \quad & \sum_{i=0}^{t-2} r_i \leq s_t \\ & 0 \leq r_i \leq M - s_i, \quad \forall i = 0, \ldots, t-2 \\ & \sum_{i=0}^{k-1} r_i \leq s_t, \quad \forall k = 1, \ldots, t-1 \\ & s_0 = 0 \\ & r_i \in \mathbb{N}^+, \quad \forall i = 0, \ldots, t-2 \end{aligned} \tag{20}$$

To explain the proposed lower bound in expression (19), we present the system (20) that minimizes n_t defined in expression (17) as the objective and in its constraints confine product return r_i ($\forall i = 0, \ldots, t-2$) following the updating rules. Because function $\sum_{i=1}^{t-2}(t-1-i)r_i$ is a weighted sum of non-negative decision variables r_i, its minimum can be found by assigning the decision variable that has the largest weight, i.e., r_0, with its largest possible value s_t and the rest of r_i to be zero. If we change the minimum in the objective to maximum, then the revised version of system (20) will find the upper bound of n_t in expression (18). The maximum can be found when assigning r_i ($\forall i = 0, \ldots, t-2$) to be zero, in which case the number of product returns of the first $t-1$ periods is always zero.

Hence given time $t \geq 2$, the state space is bounded separately with $s_t \in [0, M]$, $k_t = k_1 + s_t$, and $n_t \in [n_1 + tM - (t-1)s_t, n_1 + tM]$. The range of s_t is $M + 1$, k_t depends on s_t, and the range of n_t is $(t-1)s_t + 1$. Thus given s_t, the number of products returned at stage $t = 2, 3, \ldots, T$, there are $(t-1)s_t + 1$ corresponding states (s_t, n_t, k_t) in the dynamically updated product defect rate distribution model.

2.4.2 Solution using Dynamic Programming

Similar to solving the fixed defect rate model, we apply dynamic programming to solve the dynamic updating defect rate model (8-9). Because the state space extends from s_t to (s_t, k_t, n_t), we calculate the value function based on the ranges of k_t from expression (15) and n_t from expression (16). In this backward induction process, we use the conjugate property of the beta distribution and Bernoulli trials for states updating and compute the transition probability with expression (10) to find the value of cost-to-go (14).

Comparing the results of taking "RECALL" action with immediate recall costs and taking "CONTINUE" action with the expected expenditure of cost-to-go, we obtain the optimal decisions for each possible state at all stages. Since the state s_t has much more impact than the states k_t and n_t, we use the concept of thresholds defined for the fixed rate model, i.e., threshold θ_t is the largest state s_t that the "CONTINUE" decision remains optimal at stage t. Thresholds of all stages comprise the optimal policy.

2.4.2.1 The threshold curve

Similar to our study approach for the constant defect rate model, we start by examining the threshold from the last period θ_{T-1} to gain some insight of DP solving procedures.

Proposition 2 *Product recall threshold for the last time period is determined by given parameters from Table 1 and beta parameters k and n with following equation:*

$$\theta_{T-1} = \left\lfloor \frac{K + [c_0 - (c_1 + c_F)\, k_{T-1}/n_{T-1}]\, M}{(1 - k_{T-1}/n_{T-1})\, c_F + c_0 - c_1 k/n} \right\rfloor \tag{21}$$

Proof. The definition of recall threshold, θ_{T-1} is the largest number of return that "CONTINUE" remains the optimal decision, i.e., the cost to continue is equal to or less than the cost to recall. From the model (2-3), certain states s_{T-1} satisfy the following inequality

$$c_0 (M - s_{T-1}) + K \geq c_1 E(\tilde{r}_{T-1}) + E[V_T (s_{T-1} + \tilde{r}_{T-1})]$$

From analysis in (4), we can write $E(\tilde{r}_{T-1}) = k_{T-1}/n_{T-1} (M - s_{T-1})$, therefore

$$\begin{aligned} c_0 (M - s_{T-1}) + K &\geq c_1 \frac{k_{T-1}}{n_{T-1}} (M - s_{T-1}) \\ &\quad + E\left[V_T\left(s_{T-1} + \frac{k_{T-1}}{n_{T-1}} (M - s_{T-1})\right)\right] \\ \left[\left(1 - \frac{k}{n}\right) c_F + c_0 - c_1 \frac{k_{T-1}}{n_{T-1}}\right] s_{T-1} &\leq K + K + \left[c_0 - (c_1 + c_F) \frac{k_{T-1}}{n_{T-1}}\right] M \end{aligned}$$

It is reasonable to assume that unit cost incurred during a recall is higher than that for managing returns in a "CONTINUE" decision, i.e., $c_0 > c_1$. The ratio k_{T-1}/n_{T-1} is equivalent to the expectation of defect probability $E(\tilde{p})$, $0 < k_{T-1}/n_{T-1} < 1$. Hence the coefficient of s_{T-1} is positive, and we can solve the above inequality for all eligible s_{T-1}.

$$s_{T-1} \leq \frac{K + [c_0 - (c_1 + c_F)\, k_{T-1}/n_{T-1}]\, M}{(1 - k_{T-1}/n_{T-1})\, c_F + c_0 - c_1 k_{T-1}/n_{T-1}}$$

Because recall threshold corresponds to the largest eligible s_{T-1} solved, the result in (21) holds. ■

Now extend the above procedure to compute thresholds prior to the last periods. When $0 \leq s_t \leq M-1$ for $t = 2, \ldots, T-1$, let function $A(s_t, n_t, k_t)$ denote recall costs $c_0 (M - s_t) + K$, function $w_t(s_t, n_t, k_t)$ denote expected costs for all future periods if continue $E[V_{t+1}(s_t + \tilde{r}_t, n_t + M - s_t, k_t + r_t)]$, and function $B(s_t, n_t, k_t)$ denote return management costs $c_1 E(\tilde{r}_t \mid s_t, n_t, k_t)$ which equals to $c_1 (M - s_t)\, k_t/n_t$, value function transforms to the following:

$$V_t(s_t, n_t, k_t) = \min \begin{cases} A(s_t, n_t, k_t) \\ B(s_t, n_t, k_t) + w_t(s_t, n_t, k_t) \end{cases}$$

The company initiates recall only if $A(s_t, n_t, k_t) < B(s_t, n_t, k_t) + w_t(s_t, n_t, k_t)$. Let function $G_t(s_t, n_t, k_t)$ refer to the difference between recall costs and return management costs, i.e.,

$$\begin{aligned} G_t(s_t, n_t, k_t) &= A(s_t, n_t, k_t) - B(s_t, n_t, k_t) \\ &= \left(c_1 \frac{k_t}{n_t} - c_0\right) s_t + \left(c_0 - c_1 \frac{k_t}{n_t}\right) M + K \\ &= \alpha_{k_t, n_t} s_t + \beta_{k_t, n_t} \end{aligned}$$

where $\alpha(k_t, n_t) = c_1 k_t/n_t - c_0$ and $\beta(k_t, n_t) = (c_0 - c_1 k_t/n_t)\, M + K$. Consequently, the firm initiates product recall only when

$$G_t(s_t) = \alpha(k_t, n_t)\, s_t + \beta(k_t, n_t) < w_t(s_t, n_t, k_t)$$

Calculating backwards, with expression (14) we obtain

$$\begin{aligned} w_t(s_t, n_t, k_t) &= E[V_{t+1}(s_t + \tilde{r}_t, n_t + M - s_t, k_t + r_t)] \\ &= \sum_{\tilde{r}_t = 0}^{M - s_t} V_{t+1}(s_t + \tilde{r}_t, n_t + M - s_t, k_t + r_t) \Pr(\tilde{r}_t \mid s_t, n_t, k_t) \end{aligned}$$

from previous steps going back to the ending period T.

Theoretically, given each system state (s_t, n_t, k_t), we can compute the corresponding $\alpha(k_t, n_t)$ and $\beta(k_t, n_t)$. Conditional on the value of $\alpha(k_t, n_t)$, ranges of s_t to make product recall decisions are as follows

$\alpha(k_t, n_t)$ value	Initiate recall if s_t satisfies
$\alpha(k_t, n_t) > 0$	$s_t < [w_t(s_t, n_t, k_t) - \beta(k_t, n_t)] / \alpha(k_t, n_t)$
$\alpha(k_t, n_t) = 0$	$\beta(k_t, n_t) < w_t(s_t, n_t, k_t)$
$\alpha(k_t, n_t) < 0$	$s_t > -[\beta(k_t, n_t) - w_t(s_t, n_t, k_t)] / \alpha(k_t, n_t)$

Table 9. Recall decision ranges of system state s_t for dynamically changing defect rate model.

2.4.2.2 Numerical experiments

We experiment with a numerical example using the parameters in Table 10 and find results shown in Figure 9. The solid circle dots indicate the optimal decisions to "CONTINUE", the cross dots indicate the decision to "RECALL", the empty diamond dots indicate the decision to "STOP" and solid square dots indicate that decisions depend on n_t state.

The generic model considers a larger state space in which the current state is determined not only by the number of products returned so far, i.e., s_t, but also how these products were returned through time, i.e., n_t. Hence, with the same products returned, the optimal decisions could be either "CONTINUE" or "RECALL" depending how the returned products were accumulated. The decision in this case is identified as "n_t-DEPENDENT", because the manager cannot make a decision based on the number of products returned alone, he also needs the return history which is reflected by n_t.

We increase the number of time stages T from 4 (Figure 9) to 8 (Figure 10), 12 (Figure 11) and 16 (Figure 12) to identify its impact on optimal decisions and threshold curve. With larger problem sizes, the situation of "n_t-DEPENDENT" appears more frequently at the borders between regions of "CONTINUE" and "RECALL" decisions. For example when T is 4, only one "n_t-DEPENDENT" dot shows for $t = 2$ and $s_t = 9$. Detailed results show when the state is in $(s_2 = 9, 21 \leq n_2 \leq 26, k_2 = 10)$, we choose "RECALL"; when the state is in $(s_2 = 9, 27 \leq n_2 \leq 30, k_2 = 10)$, we choose "CONTINUE". In contrast, we have three "n_t-DEPENDENT" states for $T = 8$, four and six "n_t-DEPENDENT" states for $T = 12$ and $T = 16$ respectively.

Comparing the four results in Figures 9-12, it appears that, with the same cost parameters, the threshold curve shifts up with more initial products in the market.

c_0	c_1	c_F	K	n_0	k_0	M	T
15	2	3	15	10	1	10	4

Table 10. Parameters for numerical example.

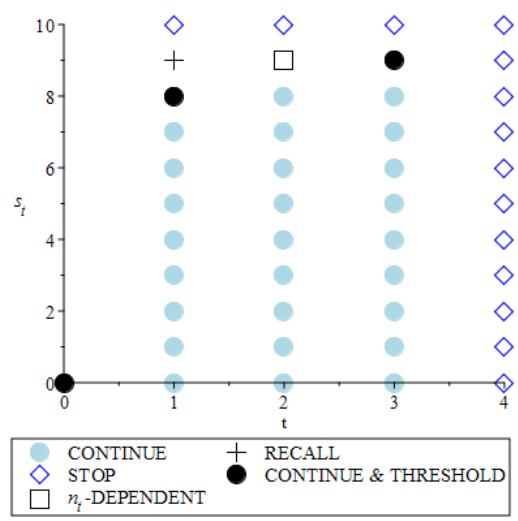

Figure 9. Optimal decisions of all possible states for dynamically updated defect rate distribution model with six products and four time periods. "n_t-DEPENDENT" suggest managers need to know state n_t to make decisions.

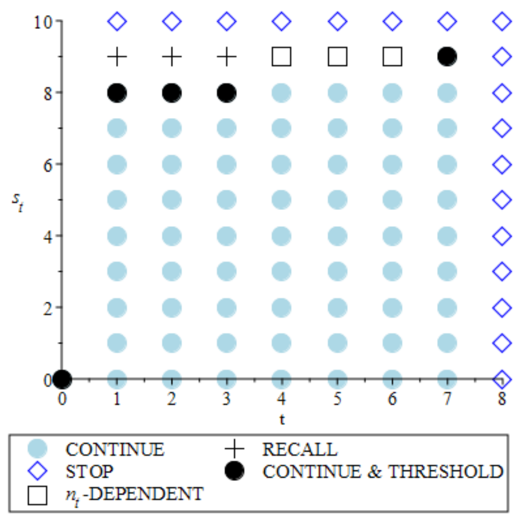

Figure 10. Optimal decisions of all possible states for dynamically updated defect rate distribution model with six products and eight time periods. "n_t-DEPENDENT" suggest managers need to know state n_t to make decisions.

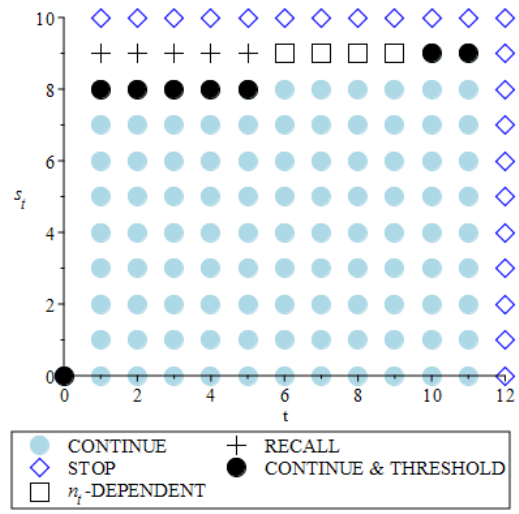

Figure 11. Optimal decisions of all possible states for dynamically updated defect rate distribution model with six products and 12 time periods. "n_t-DEPENDENT" suggest managers need to know state n_t to make decisions.

States with "n_t-DEPENDENT" optimal decisions appear as the intermediate states at the inflection points when threshold curve rises up. For these states, complete knowledge of the states (s_t, n_t, k_t) is required to make optimal decisions.

2.4.3 Size of State Space

Using results of identifying dimensionality, based on the range of n_t there are $(t-1)s_t + 1$ corresponding states (s_t, n_t, k_t) for any given s_t at stage $t = 2, 3, \ldots, T$. For any stage $t = 2, 3, \ldots, T$, the number of states is

$$\begin{aligned}
\sum_{s_t=0}^{M} [(t-1)s_t + 1] &= M + 1 + \sum_{s_t=0}^{M} (t-1)s_t \\
&= M + 1 + (t-1)\sum_{s_t=0}^{M} s_t \\
&= M + 1 + (t-1)\frac{M(M+1)}{2}.
\end{aligned}$$

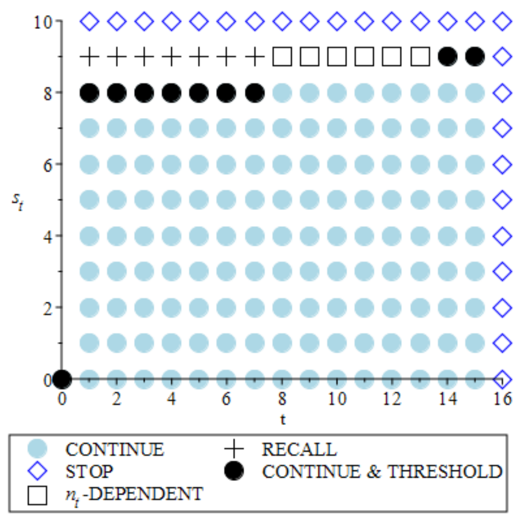

Figure 12. Optimal decisions of all possible states for dynamically updated defect rate distribution model with six products and 16 time periods. "n_t-DEPENDENT" suggest managers need to know state n_t to make decisions.

The number of states for all stages since the second stage is

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s_t=0}^{M}[(t-1)s_t+1] &= \sum_{t=2}^{T}\left[M+1+(t-1)\frac{M(M+1)}{2}\right] \\ &= (M+1)(T-1)+\frac{M(M+1)}{2}\sum_{t=2}^{T}(t-1) \\ &= (M+1)(T-1)+\frac{M(M+1)}{2}\frac{T(T-1)}{2} \\ &= (M+1)(T-1)\left(1+\frac{MT}{4}\right) \\ &= \frac{1}{4}\left(M^2T^2-M^2T+MT^2+3MT-4M+4T-4\right).\end{aligned}$$

Adding the number in the first stage, which is one, and the number in second stage, which is $M+1$, the total number for states a dynamically updated defect rate problem with given initial products on market M and warranty time span T, is

$$\sum_{t=2}^{T}\sum_{s_t=0}^{M}[(t-1)s_t+1]+1+M+1=\frac{1}{4}\left(M^2T^2-M^2T+MT^2+3MT+4T\right)+1.$$

This number largely depends on the value $M^2T^2/4$ when M and T get large. For instance, when M equals 100 and T equals 24, the number of possible states is 1,396,225 which is close to 1,440,000, the result of $M^2T^2/4$. Since the number of states rapidly increases when M and T get large, the computation for value functions becomes increasingly time consuming, which motivates us to develop alternatives to find the optimal solutions.

2.4.4 Simulation Method

To avoid the time-consuming computation of the value function $V_t(s_t, k_t, n_t)$, we use the simulation method to directly find the best fitting curve for product recall thresholds. Results from our previous numerical example indicate that the thresholds curve can be deemed to be a function of time stages. To be specific, the thresholds curve shows a nondecreasing trend as time stage t increases. A general function form of

$$\theta_t = a\sqrt[n]{t}$$

could be used for nonnegative parameters a and n. For simplicity, we use three functional forms: linear functions ($n = 1$), square root functions ($n = 2$) and cubic root functions ($n = 3$) to approximate this relationship between thresholds and time stages. Thus we transform the process of finding thresholds into optimizing parameters in the approximation functions.

Given any of the three functional forms, the largest slope is defined when the threshold of the second stage ($t = 1$) reaches $M - 1$ because the process terminates when system state reaches M. Thus feasible ranges for slope a is $[0, M - 1]$. We can limit the selection of parameter a to a range of equidifferent integers to show its impact on the objective.

After choosing a functional form, we use randomly generated states to evaluate the total cost when applying a certain value of the parameter a in the function. Comparing among all candidates, the best choice of a completes the approximation function and provides thresholds of all time stages.

A benefit of adopting this simulation method is that we can ignore the complexity of computing value functions for all possible states in each time stage. This method also works for the dynamically distributed defect rate model if we are willing to sacrifice some accuracy in system states.

Globally optimizing the simulation method requires the model to satisfy a certain level of convexity, for instance, the golden section method requires quasi-convexity. The trait of convexity is hard to prove for our models. Therefore, we will adopt an approaching method of using discretized values of a from its applicable range. For example, if we choose an equal interval of 0.5 and suppose M equals 11, then the candidates of a are $\{0, 0.5, 1, 1.5, \ldots, 9.5, 10\}$. For each candidate, we run the simulation for a sufficient amount of trials so that the sample mean is reliable.

In stage t of a simulation trial with given system states (s_t, k_t, n_t), we first judge whether the system should "RECALL" or "CONTINUE" by comparing s_t and θ_t. If s_t is greater than θ_t, which means "RECALL" is the better choice, then we compute the immediate recall costs and add it to the total costs and end its trial. Otherwise, if s_t is less than or equal to θ_t, we make a "CONTINUE" decision and update the system as follows. We start by generating the probability for any product in the market failing during this time period, i.e., product defect rate of this stage. This probability p_t follows the beta distribution. We use the uniformly distributed random number U generated by the computer as the cumulative probability from 0 to $\tilde{p}_t$ with the beta distribution.

$$\int_0^{p_t} \frac{(n_t - 1)!}{(k_t - 1)!(n_t - k_t - 1)!} p^{k_t - 1}(1 - p)^{n_t - k_t - 1} dp = U \tag{22}$$

Solving expression (22) for the probability p_t prepares the calculation for the number of returns r_t in this stage. Since r_t follows a binomial distribution of m_t products in the market, which equals $M - s_t$, and event occurrence rate of p_t, we use another

c_0	c_1	c_F	K	n_0	k_0	M	T
15	10	3	15	10	1	16	16

Table 11. Parameter settings for the simulation example.

random number as the accumulative probability following Binomial distribution and obtain r_t. Based on system states and r_t, we compute the return management cost of this stage and add it to the total costs. Following the states updating formulas (12-13), we now obtain the states $(s_t + r_t, k_t + r_t, n_t + (M - s_t))$ as $(s_{t+1}, k_{t+1}, n_{t+1})$ for stage $t + 1$ and are ready for the next stage. The process terminates when all products are returned or it reaches the ending stage, and a termination cost incurs for both cases.

Comparing the mean of total costs for all the candidates, we select the one with the smallest costs and use it for the thresholds curve. The simulation method requires much less computation because the number of parameter candidates increases with total products number M alone, compared to the complexity of using DP which increases with M^2T^2.

2.4.4.1 Exact Solution and Comparison

We can show the efficiency of using the simulation method with a numerical example. Using the parameters as shown in Table 11, we calculated the optimal solutions shown in Figure 13 with dynamic programming. The expected total cost (V_0) is 127.60 from our DP computation.

We select the value candidates for parameter a from the set $\{1, 3, 5, 7, 9\}$ and aim to test the efficiency of approximation, explore the best fitting function and find the best parameter choice for each functional form. First, we approach the threshold curve with linear functions

$$\theta_t = at.$$

With repetition number of 5000, our simulation results are shown in Table 12.

We use the percentage of error instead of efficiency since the former shows a better trend. The error is defined as the percentage difference between the simulation averages $(E\,(TC))$ and expected cost calculated using dynamic programming (V_0).

$$\text{Error}(\%)=\frac{V_0 - E\,(TC)}{V_0} \cdot 100\%$$

Our linear function approximation results show $a = 5$ as the best choice with the lease error of 4.73%.

We also use the 95% confidence interval (CI) to complement our point esti-

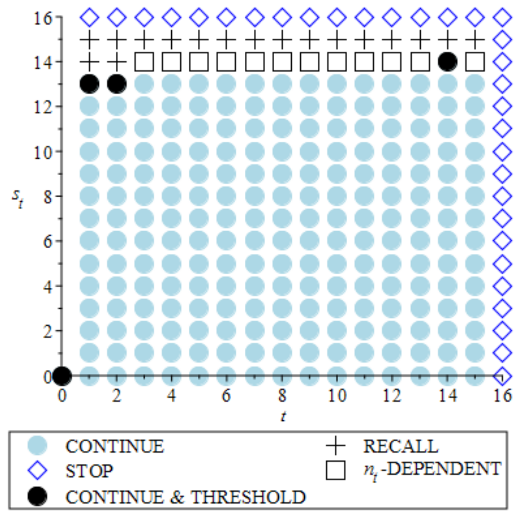

Figure 13. Optimal solutions calculated with dynamic programming method for a dynamically updating defect rate distribution model with parameters in Table 11.

a	9	7	5	3	1
$E\,(TC)$	135.41	134.54	133.64	138.23	167.04
Error	6.12%	5.43%	4.73%	8.33%	30.91%
95%CI-LB	133.59	132.75	131.81	136.29	164.60
95%CI-UB	137.22	136.32	135.47	140.17	169.48

Table 12. Simulation results with linear function approximations ($\theta_t = at$), where $a = 5$ is the best choice with the smallest error rate for problem parameters given in Table 11.

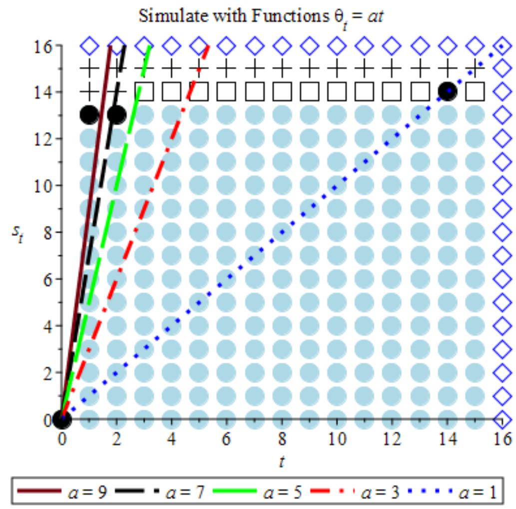

Figure 14. Combining DP solved optimal results with approximating linear functions ($\theta_t = at$), where $a = 5$ is the best choice with the smallest error rate for problem parameters given in Table 11.

mation of expected cost ($E(TC)$). The lower bound (LB) and upper bound (UB) of the confidence interval in Table 12 are calculated with the following formula

$$LB = \mu - t_{\alpha/2}\frac{s}{\sqrt{n}},\ UB = \mu + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where μ and s are the mean value and sample standard deviation of our experiment results respectively; since our experiment repetition n is 5000, we use $z_{\alpha/2}$ of 1.96 instead of the t distribution value.

Combining the linear function lines with the threshold curve and optimal solutions results in Figure 14. The number of states violating the DP optimal results is the least when $a = 5$ in linear approximations is perhaps the reason why it shows the least error percentage.

Next, we experiment the square root functions

$$\theta_t = a\sqrt{t}$$

with the same set of values $\{1, 3, 5, 7, 9\}$ as the candidates for parameter a. In Table 13, experiment results show the square root function performs the best when $a = 7$ with error equals 3.88%. Observing the combination of all five square root function curves and the DP solved optimal solutions in Figure 15, the curve using $a = 7$ has perhaps the least number of states violating the optimal decisions from DP solution.

a	9	7	5	3	1
$E(TC)$	134.46	132.55	133.00	151.11	210.04
Error	5.37%	3.88%	4.23%	18.42%	64.61%
95%CI-LB	132.67	130.74	131.16	148.92	207.97
95%CI-UB	136.25	134.36	134.84	153.29	212.21

Table 13. Simulation results with square root function approximations ($\theta_t = a\sqrt{t}$), where $a = 7$ is the best choice with the smallest error rate for problem parameters given in Table 11.

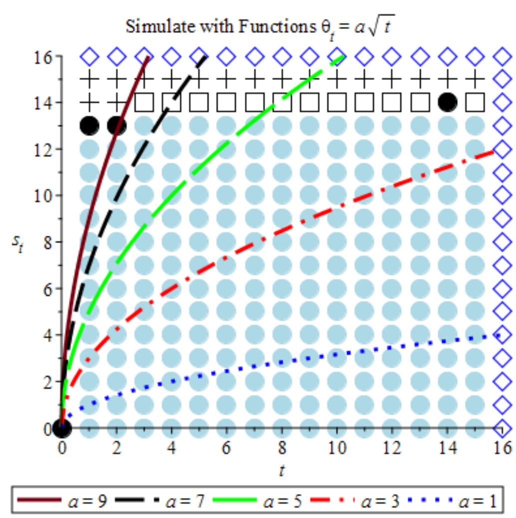

Figure 15. Combining DP solved optimal results with approximating square root functions ($\theta_t = a\sqrt{t}$), where $a = 7$ is the best choice with the smallest error rate for problem parameters given in Table 11.

a	9	7	5	3	1
$E(TC)$	133.41	132.92	138.42	170.77	216.71
Error	4.55%	4.16%	8.48%	33.83%	69.84%
95%CI-LB	131.59	131.16	136.43	168.44	214.79
95%CI-UB	135.22	134.67	140.41	173.10	218.64

Table 14. Simulation results with cubic root function approximations ($\theta_t = a\sqrt[3]{t}$), where $a = 7$ is the best choice with the smallest error rate for problem parameters given in Table 11.

Then we experiment with cubic root functions

$$\theta_t = a\sqrt[3]{t}$$

with the same value set for parameter a. The simulation results shown in Table 14 present $a = 7$ as the best choice for cubic root functions with error 4.16%. Also combining the cubic function curves and DP-obtained optimal results in Figure 16 confirms our observation that the best fitting curve has the least states violating optimal decisions.

Comparing all of the above results, the square root function using parameter $a = 7$ has the least error percentage. Therefore, we choose the square root function $\theta_t = a\sqrt{t}$ for simulations of large size problems.

2.4.4.2 Estimation of Threshold Curve for Large Scale Problems

We exemplify the efficiency of using simulation method with the setting that products number M is 100 and time stages T is 24. In this case the total number of states approaches 1.4 million as discussed in Section 2.4.3.

Since the number of products increases to 100, we choose the candidates for parameter a from the set $\{10, 30, 50, 70, 90\}$. Our computation results shows in Table 16 using 5000 repetitions for each parameter candidate. The results suggest that using $a = 50$ in the square root function has the lowest expected costs 944.90, which can be used as the upper bound estimation for expected costs. Our simulation with 5000 repetitions takes about 300 seconds (five minutes) to finish, yet applying the DP method takes over 361,200 seconds (above 100 hours) to finish.

Note that our simulation results provide the best choice within the given candidates. It is likely that in between these numbers there are better choices of the function parameter for more accurate approximation. The limitation of simulation comes from both the choice of functional form and the choice of parameter candidates set. The benefits of using simulation include a guaranteed solution and quick

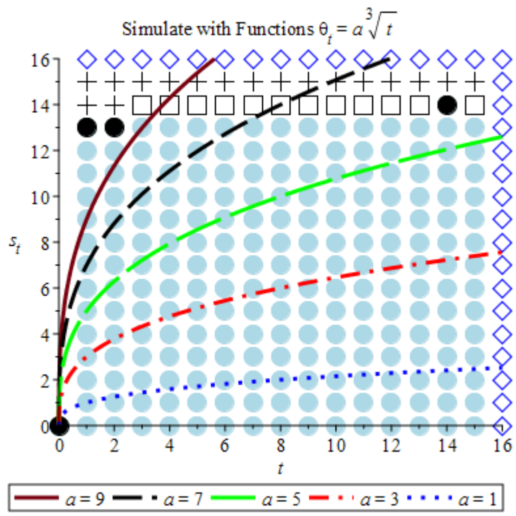

Figure 16. Combining DP solved optimal results with approximating cubic root functions ($\theta_t = a\sqrt[3]{t}$), where $a = 5$ is the best choice with the smallest error rate for problem parameters given in Table 11.

c_0	c_1	c_F	K	n_0	k_0	M	T
15	10	3	15	100	1	100	24

Table 15. Parameter settings for the large size problem ($M = 100, T = 24$) in assessing the simulation method.

a	90	70	50	30	10
$E\,(TC)$	951.93	950.87	944.90	945.70	1155.78
95%CI-LB	941.51	940.52	934.53	935.47	1143.08
95%CI-UB	962.34	961.21	955.27	955.92	1168.47
Time(sec.)	301	301	299	270	96
Time(min.:sec.)	5:01	5:01	4:59	4:30	1:36

Table 16. Simulation results for large size problem ($M = 100, T = 24$) with square root function approximations ($\theta_t = a\sqrt{t}$), where $a = 50$ is the best choice with the smallest expected costs for problem parameters given in Table 15.

response to the change of problem settings, which can provide the decision maker a reliable approach for examining different problem scenarios and a dependable upper bound estimation of expected costs.

2.5 Conclusions and Proposal for Future Research

In this chapter, we construct the model for making product recall decisions by using the number of product returns to update the estimation of product defect rate distribution. Our primary contribution to the literature is that we model and solve the optimization problem of product recalls by combining dynamic programming with the Bayesian conjugate property of beta distributions and Bernoulli processes. We build two models of different defect rate distribution estimation. The model for stationary defect rate distribution assumes the random variable for defect rate follows the same distribution during the entire time horizon, while the dynamically updating defect rate distribution uses the product returns of the proceeding stage to update the estimated distribution.

We define the threshold curve as the connection of states of largest number of total returns of all stages in the optimal solution. The threshold curve for the optimal decision shows a nondecreasing trend that crosses the origin. This observation inspires our approximation of the threshold curve by nonlinear functions, such as square root functions and cubic root functions.

Our proposed DP method is capable of solving moderate sized problems and obtaining optimal solutions of all possible states. Yet, the computation complexity is proportion to the squared product of the number of products in the market and total time stages. Therefore we apply the simulation method for problem-solving, which shows low-error rate performance when applied to moderate-size benchmark problems. Our experiments show the proposed simulation method could solve the large sized problems in a satisfying balance of solution accuracy and computation time.

In future work, we propose to explore the feasibility of using approximate dynamic programming methods to solve our model. Unlike the backward recursion used in dynamic programming, the Approximate Dynamic Programming (ADP) steps forward through time using value function approximation and sample paths in search of the optimal solution. Sample path is a sequence of exogenous information that describes the stage transitions of all defined stages. ADP implements sufficient amount of calculation iterations using sample paths to find the satisfactory approximation for value functions.

A commonly used method to approximate the value function is Q-learning. Application of Q-learning for optimal stopping problem has been discussed by Bertsekas and Tsitsiklis [7] where the transitional probability matrix is given and has a steady-state distribution. Based on their work, we may be able to explore the possibility of applying Q-learning to our dynamically updated product defect rate distribution model.

Chapter 3 Newsvendor Problem using Quality Investment for Product Recall Risk Control

This chapter extends the classic single-period stochastic newsvendor model with product recall risk management. The probability of product recalls is captured by a decreasing function of product quality level. The core company exerts control on product quality through manufacturing investment, which impacts unit production cost, to optimize total expected profits. Given the free choices of the cost function and recall probability function, the resulting objective function considering sale revenues, operational costs and recall risks is not necessarily negative semi-definite. Parametric analysis in the newsvendor model reveals several interesting features. One is that the production quantity and quality level seem to have conflicting effects as one waxes and the other wanes in optimal solutions during most cases of parameter changes. The only two exceptions include changing variable unit production cost and demand rate. Another feature is that increasing profitability discourages investment in quality. We further extend our model from internal supply to external supply from multiple sources. In the case of two external suppliers satisfying independent demands and covering each other's demand only when the other is having recalls, our results suggest little impact from recall-covering interaction on optimal solutions. Additional numerical results show consistency with internal supply case.

3.1 Introduction

This chapter extends the newsvendor problem with product quality control. In order to deal with risks in product recall, the core company decides its investment level on product quality, which impacts production cost and the probability of product recall. The classic newsvendor problem is a single-period stochastic inventory control problem. In their influential work, Arrow, Harris and Marschak [2] examine the newsvendor model and derive the critical fractile solution. We follow the same notation system in Parlar [32] which provides an excellent detailed review of the problem. Typically it considers seasonal products for which leftover products have little value — the salvage value is small compared to the sale price. Through deciding the optimal order quantity Q the newsvendor model considers the expected total profit that includes salvage values (v per unit), penalty costs (p per unit), sales revenue (s per unit) and purchasing costs (c per unit). Suppose the demand is denoted by random variable X with a probability density function $f(x)$, the random profit is denoted by Π and a realized value of profit π can be written as

$$\pi = \begin{cases} sx + v(Q-x) - cQ, & \text{if } x \leq Q \text{ (surplus)} \\ sQ - p\,(x-Q) - cQ, & \text{if } x > Q \text{ (shortage)} \end{cases}$$

Thus, the expected profit $E\,[\Pi] = P(Q)$ is a function of order quantity:

$$\begin{aligned} P(Q) &= \int_{x=0}^{Q} [sx + v(Q-x) - cQ]\, f\,(x)\, dx \\ &\quad + \int_{x=Q}^{\infty} [sQ - p\,(x-Q) - cQ]\, f\,(x)\, dx \\ &= (s-v)\,\mu - (c-v)Q - (s+p-v) \int_{x=Q}^{\infty} (x-Q)\, f\,(x)\, dx \end{aligned}$$

where $\mu = E\,[X]$ is the mean demand. Using Leibniz's rule of differentiation, we have

$$\begin{aligned} P'\,(Q) &= -(c-v) + (s+p-v) \int_{x=Q}^{\infty} f\,(x)\, dx \\ P''\,(Q) &= -\,(s+p-v)\, f\,(Q) \leq 0 \end{aligned}$$

and the negative second order derivative proves the objective function concave.

The concavity of the above objective function leads us to solve the problem with first and second order optimality conditions. Given the demand probability density function $f\,(x)$ and using $F\,(x)$ for the corresponding cumulative distribution function (CDF), one can obtain the optimal order quantity Q^* by solving the following equation

$$\bar{F}\,(Q^*) = \int_{x=Q}^{\infty} f\,(x)\, dx = \frac{c-v}{s+p-v}$$

where $\bar{F}\,(x) = 1 - F\,(x)$.

The newsvendor model assumes products have perfect quality, however, recent product recall events suggest otherwise. We consider possible product recalls, which occur after product sales, in response to low product quality. The core company controls product quality through investment in manufacturing; high investment ensures high product quality level. We use ℓ to denote product quality level and $R(\ell)$ to denote the probability of incurring a product recall because of quality concerns. Improving quality level ℓ decreases the product recall probability $R(\ell)$ ($0 \leq R(\ell) \leq 1$) and increases the unit cost of production, denoted by $c(\ell)$. Hence, first derivative functions satisfy the following conditions:

$$R'(\ell) = \frac{dR(\ell)}{d\ell} < 0,\, c'(\ell) = \frac{dc(\ell)}{d\ell} > 0$$

By choosing the right manufacturing quantity Q and quality level ℓ we can maximize the expect total profits, denoted by $E(\Pi)$, considering revenues, operational costs, and possible recall costs. Product recall will incur the cost of k per unit to manage the products in the market.

Calculating total profits in the case of no product recalls (π_1), given the realization of demand x, is the same as in the classic newsvendor problem, which is

$$\begin{aligned}
\pi_1 &= \begin{cases} sx + v(Q-x) - c(\ell)\,Q, & \text{if } x \le Q \\ sQ - p(x-Q) - c(\ell)\,Q, & \text{if } x > Q \end{cases} \\
&= s\min\{Q,x\} + v(Q - \min\{Q,x\}) \\
&\quad -p(x - \min\{Q,x\}) - c(\ell)\,Q \\
&= (s - v + p)\min\{Q,x\} + (v - c(\ell))\,Q - px
\end{aligned}$$

If there is a product recall, total profits (π_0) depend on whether there is a surplus or shortage of products compared to demand. If the seller overproduced, he needs to manage recall cost on top of sales revenue and production cost but will receive no salvage value for surplus production. If there is a shortage, the seller need not worry about shortage penalty since recall cost overshadows it in terms of the cost of loss of goodwill. Thus,

$$\begin{aligned}
\pi_0 &= \begin{cases} (s-k)\,x - c(\ell)\,Q, & \text{if } x \le Q \\ (s-k)\,Q - c(\ell)\,Q, & \text{if } x > Q \end{cases} \\
&= (s-k)\min\{Q,x\} - c(\ell)\,Q
\end{aligned}$$

Expected number of products sold, denoted by function $L(Q)$, is

$$\begin{aligned}
L(Q) &= E[\min\{Q,X\}] \\
&= Q\bar{F}(Q) + \int_{x=0}^{Q} xf(x)\,dx \\
&= \mu - \int_{x=Q}^{\infty} (x-Q)\,f(x)\,dx
\end{aligned}$$

Hence, the expected value for total profits $E(\Pi) = \mathcal{P}(Q,\ell)$ is calculated conditioning on product recall occurrence. Define function $\mathcal{P}_0(Q,\ell) = E[\Pi_0]$ and $\mathcal{P}_1(Q,\ell) =$

$E[\Pi_1]$. We have

$$\begin{aligned}\mathcal{P}(Q,\ell) &= E(\Pi \mid \text{no recall})\Pr(\text{no recall}) + E(\Pi \mid \text{recall})\Pr(\text{recall}) \\ &= \mathcal{P}_1(Q,\ell)\Pr(\text{no recall}) + \mathcal{P}_0(Q,\ell)\Pr(\text{recall}) \\ &= [(s+p-v)L(Q) + (v-c(\ell))Q - p\mu](1-R(\ell)) \\ &\quad + [(s-k)L(Q) - c(\ell)Q]R(\ell) \\ &= \left[\bar{s} - \bar{k}R(\ell)\right]L(Q) - [c(\ell) + vR(\ell) - v]Q \\ &\quad - p\mu[1-R(\ell)] \end{aligned} \tag{23}$$

where all parameters defined in Table 17 and for brevity using notation

$$\bar{k} = k + p - v,\ \bar{s} = s + p - v$$

To ensure our model is solvable and meaningful, we assume the following relationship hold for the four cost parameters:

$$k > s,\ p > v,\ s > c(\ell = 0) > v \tag{24}$$

Notations	Definition
k	unit cost to recall products in the market
p	penalty cost of per unit product at a shortage
Q	manufacturing quantity variable
s	sales income of products sold per unit
ℓ	product quality level variable
v	salvage value per unit of surplus product
X	demand random variable

Table 17. Notations and meanings for parameters and variables in newsvendor quality investment for product recall risk control models.

We explore the possible optimal solutions with first order conditions by solving the equations (25-26).

$$\frac{\partial \mathcal{P}(Q,\ell)}{\partial Q} = \left(\bar{s} - \bar{k}R(\ell)\right)\bar{F}(Q) - c(\ell) - vR(\ell) + v = 0 \tag{25}$$

$$\frac{\partial \mathcal{P}(Q,\ell)}{\partial \ell} = R'(\ell)\left(p\mu - \bar{k}L(Q) - vQ\right) - Qc'(\ell) = 0 \tag{26}$$

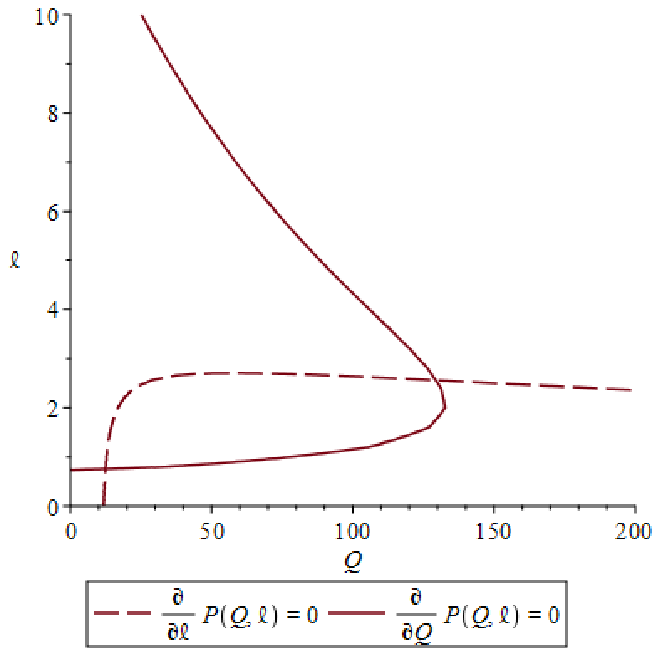

Figure 17. Implicit plot of first derivative functions shows two intersection points for potential solutions. The left point is $(Q = 12.44, \ell = 0.74)$ and the right point $(Q = 129.69, \ell = 2.55)$.

Equation (25) can reduce to the following

$$\bar{F}(Q) = \frac{c(\ell) + vR(\ell) - v}{\bar{s} - \bar{k}R(\ell)}$$

which is the exact optimal condition for the one-stage newsvendor problem when reducing quality factor ℓ, i.e., $c(\ell) = c$ and $R(\ell) = 0$. When we draw the implicit plot of equations (25-26), the intersection points are the candidates for our optimal solutions. For instance, using parameter values from Section 3.2.1, we find two candidate solutions as intersection points in the implicit plot as shown in Figure 17. Computing objective values from the candidate solutions leads us to the optimal solution. For example, the objective value for the left point in Figure 17 $\mathcal{P}(Q = 12.44, \ell = 0.74)$ is -339.94, and the right point has the objective value $\mathcal{P}(Q = 129.69, \ell = 2.55)$ of 310.96, which is the optimal solution.

3.2 Model Analysis

In this section, we explore the concavity of the proposed objective function. The Hessian matrix H of our objective (23) has the following entries

$$\begin{aligned} H_{11} &= \left(\bar{s} - \bar{k}R(\ell)\right) L''(Q) \\ H_{12} &= H_{21} = -\bar{k}R'(\ell) L'(Q) - c'(\ell) - vR'(\ell) \\ H_{22} &= -\left(\bar{k}L(Q) + vQ - p\mu\right) R''(\ell) - c''(\ell) Q \end{aligned}$$

where $c''(\ell)$ and $R''(\ell)$ are corresponding second derivative functions, respectively.

Since the two-by-two Hessian matrix H is symmetric, the sufficient and necessary conditions for H to be negative-semidefinite are $H_{11} \leq 0$, $H_{22} \leq 0$ and $H_{11}H_{22} - (H_{12})^2 \geq 0$ as provided by Bazaraa et al. [3]. These conditions for our objective (23) are as follows:

$$\left(\bar{s} - \bar{k}R(\ell)\right) L''(Q) \leq 0 \tag{27}$$

$$-\left(\bar{k}L(Q) + vQ - p\mu\right) R''(\ell) - c''(\ell) Q \leq 0 \tag{28}$$

$$\begin{aligned} -\left(\bar{s} - \bar{k}R(\ell)\right) L''(Q) \left\{ \left(\bar{k}L(Q) + vQ - p\mu\right) R''(\ell) + c''(\ell) Q \right\} & \\ - \left[\bar{k}R'(\ell) L'(Q) + c'(\ell) + vR'(\ell)\right]^2 &\geq 0 \end{aligned} \tag{29}$$

The above conditions hold if our model (23) is reduced to the simpler case of one-stage stochastic newsvendor problem by letting the production cost of highest quality level $\overline{\ell}$ equal the newsvendor's unit production price and assigning the corresponding recall probability as zero, i.e., $c\left(\overline{\ell}\right) = c$ and $R\left(\overline{\ell}\right) = 0$. Given this quality level $\overline{\ell}$, conditions (28-29) hold for derivatives of fixed values of functions $R\left(\overline{\ell}\right)$ and $c\left(\overline{\ell}\right)$ are all zeros and condition (27) reduces to the following

$$-\left(s + p - v\right) L''(Q) \geq 0. \tag{30}$$

Because $L''(Q) = -f(Q)$, the condition (30) is

$$(s + p - v) f(Q) \geq 0$$

and equivalent to the concavity condition $P''(Q) = -(s + p - v) f(Q) \leq 0$ for the aforementioned newsvendor problem.

Next, we explore some necessary conditions for desirable solutions (Q, ℓ) which will assist our concavity analysis. By desirability we mean gaining positive expected profit, i.e., $\mathcal{P}(Q, \ell) \geq 0$. Desirable Q and ℓ should satisfy the necessary condition of

$$\mathcal{P}_1(Q, \ell) \geq 0$$

since $\mathcal{P}_0(Q, \ell)$ is always negative. Finding desirable region for $\mathcal{P}(Q, \ell)$ can be very difficult, thus we explore necessary conditions of the desirable region by satisfying $\mathcal{P}_1(Q,\ell) \geq 0$ instead. We use $\mathcal{P}_1$-desirable to denote (Q,ℓ) that satisfies $\mathcal{P}_1(Q,\ell) \geq 0$.

We search for the $\mathcal{P}_1$-desirable region of quality level ℓ by identifying its upper and lower bounds separately. The highest quality level $\overline{\ell}$ requires the largest expected profits of no product recall condition to be positive, i.e.,

$$\overline{\ell} = \arg\max_{\ell} \mathcal{P}_1(Q^*,\ell) \geq 0$$

where the optimal quantity Q^* is obtained by

$$\bar{F}(Q^*) = \frac{c(\overline{\ell}) - v}{s+p-v}.$$

At the same time, with our assumption in (24) we have $c(\ell=0) > v$ to ensure practicality, therefore the lower bound of ℓ is zero. Together we obtain the $\mathcal{P}_1$-desirable region for ℓ as $[0,\overline{\ell})$.

The $\mathcal{P}_1$-desirable region of Q is a function of ℓ, i.e.,

$$Q \in \left\{Q \mid \mathcal{P}_1(Q,\ell) \geq 0,\ \forall\ \ell \in [0,\overline{\ell})\right\}.$$

Using parameters in Section 3.2.1, $\mathcal{P}_1$-desirable region of ℓ is suggested to be $[0,5.1)$ and $\mathcal{P}_1$-desirable region of Q shifts with the value of ℓ. Figure 18 shows $\mathcal{P}_1$-desirable region of Q gets larger and its center shifts to the right when ℓ decreases.

Proposition 3 *Necessary conditions for $\mathcal{P}\left(\bar{Q},\ell\right)$ to be concave on ℓ, given an arbitrary $\bar{Q}$ in its $\mathcal{P}_1$-desirable region, are $R''(\ell) \geq 0$ and $c''(\ell) \geq 0$ for any $\mathcal{P}_1$-desirable ℓ.*

Proof. Rewrite $\partial^2\mathcal{P}(Q,\ell)/\partial\ell^2 = H_{22}$ as follows

$$\partial^2\mathcal{P}(Q,\ell)/\partial\ell^2 = -c''(\ell)Q + R''(\ell)g(Q) \tag{31}$$

where

$$g(Q) = -\bar{k}L(Q) + p\mu - vQ.$$

Since both Q and ℓ are $\mathcal{P}_1$-desirable, they satisfy $\mathcal{P}_1$-desirable condition as follows

$$\begin{aligned} \mathcal{P}_1(Q,\ell) &= \overline{s}L(Q) + (v - c(\ell))Q - p\mu \geq 0 \\ \overline{s}L(Q) - c(\ell)Q &\geq p\mu - vQ \end{aligned}$$

Hence,

$$\begin{aligned} g(Q) &\leq -\bar{k}L(Q) + \overline{s}L(Q) - c(\ell)Q \\ &= (s-k)L(Q) - c(\ell)Q \end{aligned}$$

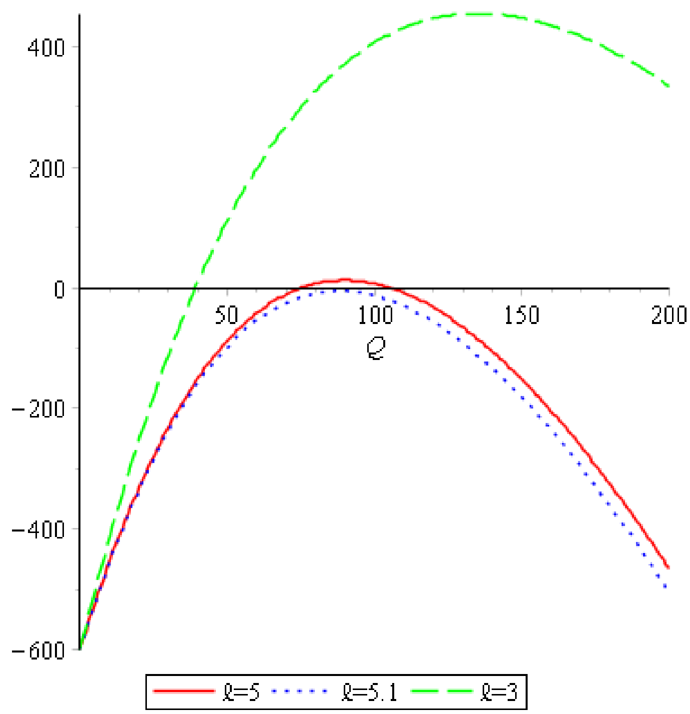

Figure 18. Illustration of $\mathcal{P}_1$-desirable region for ℓ is [0,5.1) since $\mathcal{P}_1(Q^*, \ell = 5.1)$ is negative (-5.5) and $\mathcal{P}_1(Q^*, \ell = 5.0)$ is a small positive number (12). Also, $\mathcal{P}_1$-desirable region of Q is determined by the value of ℓ. As ℓ increases, $\mathcal{P}_1$-desirable region of Q enlarges its size and its center shifts to the right ($Q^*(\ell = 5.1) \approx 88$, $Q^*(\ell = 5.0) \approx 90$, $Q^*(\ell = 3) \approx 135$).

With our assumption (24) that $s < k$, it is clear that $g(Q) \leq 0$.

To summarize, having both $R''(\ell)$ and $c''(\ell)$ being non-negative ensures $\partial^2 \mathcal{P}(Q,\ell)/\partial \ell^2 \leq 0$ given both variables are in their $\mathcal{P}_1$-desirable regions. These results suggest that when satisfying the conditions of $R''(\ell) \geq 0$ and $c''(\ell) \geq 0$ if the production quantity is fixed to a $\mathcal{P}_1$-desirable quantity $\bar{Q}$, the expected profit $\mathcal{P}(Q,\ell)$ is concave on $\mathcal{P}_1$-desirable quality level ℓ. ■

Proposition 4 *Sufficient and necessary condition for concavity condition (27), i.e., $H_{11} \leq 0$, is $R(\bar{\ell}) \leq \bar{s}/\bar{k}$ given $\mathcal{P}_1$-desirable (Q,ℓ).*

Proof. We have

$$H_{11} = \left(\bar{s} - \bar{k}R(\ell)\right) L''(Q) = -\left(\bar{s} - \bar{k}R(\ell)\right) f(Q) \leq 0$$

which is equivalent to $\left(\bar{s} - \bar{k}R(\ell)\right) f(Q) \geq 0$. Since $f(Q)$ is always positive, the condition transforms to

$$\bar{s} - \bar{k}R(\ell) \geq 0.$$

Because $R(\ell)$ is an increasing function of quality level ℓ, i.e., $R(\ell) \leq R(\bar{\ell})$ where $\bar{\ell}$ is the upper bound of the $\mathcal{P}_1$-desirable ℓ as defined previously. Thus the condition is equivalent to

$$\bar{s}/\bar{k} \geq R(\bar{\ell})$$

and complete our proof. ■

Proposition 5 *Sufficient and necessary conditions for concavity condition (28), i.e., $H_{22} \leq 0$, are $R''(\ell) \geq 0$ and $c''(\ell) \geq 0$ given $\mathcal{P}_1$-desirable (Q,ℓ).*

Proof. The condition of $H_{22} \leq 0$ is equivalent to

$$H_{22} = g(Q) R''(\ell) - c''(\ell) Q \leq 0$$

where $g(Q) = -\bar{k}L(Q) + p\mu - vQ$ as defined previously. We have proved $g(Q) \leq 0$ of any $\mathcal{P}_1$-desirable Q also quantity Q is always non-negative, thus the sufficient and necessary conditions are $R''(\ell) \geq 0$ and $c''(\ell) \geq 0$. ■

Proposition 6 *Necessary conditions for the concavity condition (29), $H_{11}H_{22} - (H_{12})^2 \geq 0$, include $\bar{s}/\bar{k} > R(\bar{\ell})$, $R''(\ell) \geq 0$, $c''(\ell) \geq 0$ and $R''(\ell) + c''(\ell) > 0$.*

Proof. This concavity condition requires $H_{11}H_{22} \geq 0$ if $H_{12} = 0$ or $H_{11}H_{22} > 0$ if $H_{12} \neq 0$. Since

$$\begin{aligned} H_{12} &= -\bar{k}R'(\ell) L'(Q) - c'(\ell) - vR'(\ell) \\ &= -\left[c'(\ell) + R'(\ell)\left(\bar{k}\bar{F}(Q) + v\right)\right] \end{aligned}$$

is not always zero given arbitrary values of (Q, ℓ), we need $H_{11}H_{22} > 0$.

$$\begin{aligned} H_{11}H_{22} &= -\left(\bar{s} - \bar{k}R(\ell)\right) L''(Q) \left\{ \left(\bar{k}L(Q) + vQ - p\mu\right) R''(\ell) + c''(\ell) Q \right\} \\ &= f(Q) \left(\bar{s} - \bar{k}R(\ell)\right) \left[-g(Q) R''(\ell) + c''(\ell) Q\right] \end{aligned}$$

Given any $\mathcal{P}_1$-desirable (Q, ℓ), we have proved $g(Q) \leq 0$. Because $R(\ell)$ ranges in $(0, 1)$ and $\bar{s} < \bar{k}$, we need

$$\bar{s} - \bar{k}R(\ell) > 0,\ R''(\ell) \geq 0,\ c''(\ell) \geq 0 \text{ and } R''(\ell) + c''(\ell) > 0.$$

Since function $R(\ell)$ is increasing of ℓ and $\bar{\ell}$ is the upper bound of $\mathcal{P}_1$-desirable quality level, above necessary conditions of expression (29) transform to the following

$$\bar{s}/\bar{k} > R\left(\bar{\ell}\right),\ R''(\ell) \geq 0,\ c''(\ell) \geq 0 \text{ and } R''(\ell) + c''(\ell) > 0.$$

and complete our proof. ■

Now we explore the convexity of recall function $\mathcal{P}_0(Q, \ell)$ because when quality level ℓ is small, $\mathcal{P}_0(Q, \ell)$ dominates the result of the total expected profit. The Hessian matrix for $\mathcal{P}_0(Q, \ell)$ is shown as follows

$$\begin{bmatrix} (k - s) f(Q) & -c'(\ell) \\ -c'(\ell) & -Qc''(\ell) \end{bmatrix}$$

Since k is greater than s, function $\mathcal{P}_0(Q, \ell)$ is convex on Q given a fixed quality level $\hat{\ell}$ because the second order partial derivative is non-negative, i.e.,

$$\frac{\partial^2 \mathcal{P}_0(Q, \ell)}{\partial Q^2} = (k - s) f(Q) \geq 0 \tag{32}$$

Satisfying conditions (27-29) will require certain constraints on parameters and structures of functions $R(\ell)$ and $c(\ell)$. Our numerical study in Section 3.2.1 suggests that an arbitrary choice of parameters and function structures will result in an objective function neither concave nor convex as shown in Figure 19. The existence of local minima can be explained by expression (32), i.e., the value of $\mathcal{P}_0(Q, \ell)$ plays a bigger role when the quality level is very small and given any fixed quality level $\hat{\ell}$, function $\mathcal{P}_0(Q, \ell)$ is convex on production quantity Q.

3.2.1 Base Case for Numerical Study

We have chosen the initial parameter values and functional forms to establish a base case. This allows us to examine the impact of parametric changes on the optimal solutions. We choose exponential functions for the demand density function $f(x)$ and recall probability function $R(\ell)$ and a linear unit production cost function

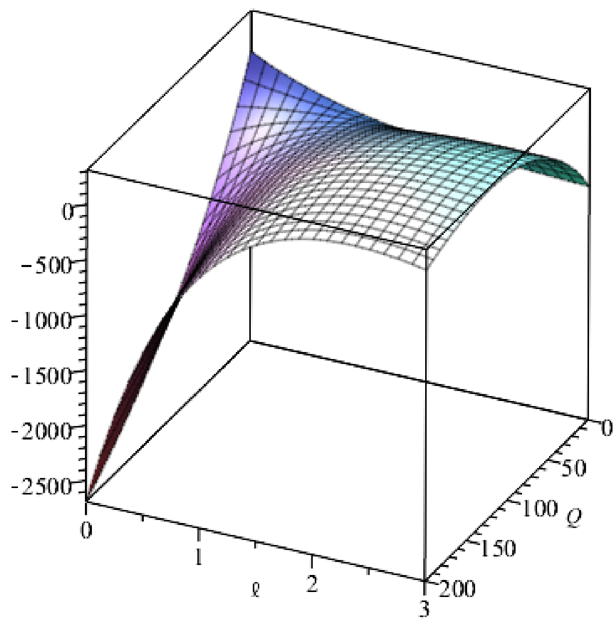

Figure 19. Illustration of objective being neither convex and concave with arbitrary choice of functions.

$c(\ell)$, as

$$f(x) = \lambda e^{-\lambda x}, \quad R(\ell) = \alpha e^{-\beta \ell}, \quad c(\ell) = \gamma + \theta \ell$$

with parameters from Table 18.

α	β	γ	θ	λ	k	p	s	v
0.9	1	5	2	1/100	50	6	25	4

Table 18. Parameter values for the base case study.

The solution for this special case is Q^* approximately equals to 130 and ℓ^* approximately equals to 2.55 resulting in an expected profit of 310.95. The surface of the base case illustrates that the model can be concave as shown in Figure 20. Conditions that allow concave models will be one the goals of our next step.

3.2.2 Parametric Analysis

In this section, we evaluate impacts of various parameters on our proposed model (23) when they deviate from base case values. We present optimal decisions of production quantity Q^* and quality level ℓ^* as well as corresponding expected profits $\mathcal{P}^*(Q^*, \ell^*)$ that results from changing parameters from their base case values. The base case parameter values are in bold in the Tables (19-25).

We first explore the impacts of the magnitude of recall probability, i.e., the impact of parameter α in $R(\ell) = \alpha e^{-\beta \ell}$ on the optimal solutions and expected profits. Experiments show that as the probability of recall decreases, the manager will produce more and invest less on quality level and obtain higher profits, as indicated by results in Table 19. This suggests that less pressure from potential product recalls leads the manager to focus on higher production and less on improving product quality. Expected profits also increase because of a lower product recall risk.

We also want to measure the impacts of changing unit production cost, i.e., impacts of parameters γ and θ in $c(\ell) = \gamma + \theta \ell$. Results in Table 20 suggest that increasing unit production cost will lead to lower production and higher investment on product quality, but increasing variable production cost will result in lower production quantity and also lower quality level. Increasing fixed part γ or varied part θ impacts differently on the optimal solution. Expected profit is quite sensitive to the changes in unit production cost. This suggests that when dealing with manufacturing becomes more expensive, managers are willing to invest more on product quality and produce less.

Changes in expected demand affect the production quantity alone in the optimal solutions, as shown by results in Table 21. This suggests the demand rate will

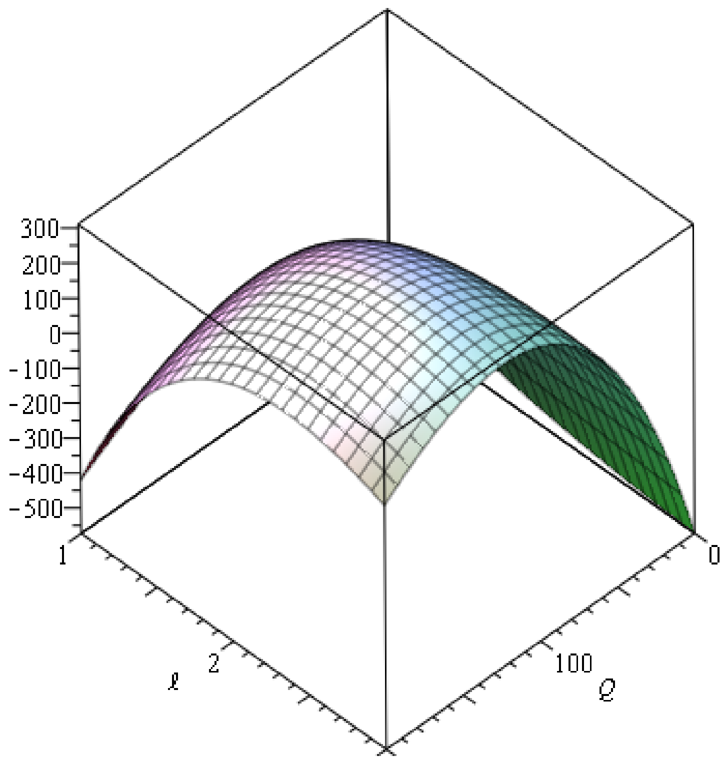

Figure 20. Function surface of the base case shows the model's concavity in local area.

α	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
0.6	144	2.11	422
0.7	138	2.28	378
0.8	134	2.42	342
0.9	130	2.55	311
1	126	2.67	284

Table 19. Impacts of recall probability parameter α on optimal solutions.

γ	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$	θ	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
4	148	2.50	449	1	176	3.11	735
5	130	2.55	311	1.5	150	2.78	496
6	115	2.59	189	**2**	130	2.55	311
7	102	2.63	81	2.5	114	2.37	162
8	91	2.65	-16	3	100	2.22	39

Table 20. Impacts of fixed production cost (left) and variable production cost (right) on optimal solutions.

$1/\lambda$	Q^*	ℓ^*	$\mathcal{P}^*(Q^*,\ell^*)$	$1/\lambda$	Q^*	ℓ^*	$\mathcal{P}^*(Q^*,\ell^*)$
80	104	2.55	249	**100**	130	2.55	311
90	117	2.55	280	200	259	2.55	622
100	130	2.55	311	300	389	2.55	933
110	143	2.55	342	400	519	2.55	1244
120	156	2.55	373	500	648	2.55	1555

Table 21. Impacts of expected demand rates on optimal solutions.

only impact the production quantity but not the quality level. Results of the right-side table concord with our observation that optimal quality level is not affected by changes in demand rates. Moreover, the optimal production quantity and expected profits increase linearly with the demand rates.

Increasing the unit cost of recall management will decrease the production quantity and increase quality investment, similar to that of increasing production cost, but the impact is much milder. In the right side of Table 22, when modifying unit recalls cost with the same ratio of changing unit production cost, its impact on expected profits remains much milder compared to that of unit production cost.

Varying the value of unit penalty cost shows a mild impact on both optimal solutions as shown in Table 23. Increasing unit penalty of shortage stimulates more production but lowers the quality level slightly at the same time.

Lowering the sales price too much will force the company to produce zero units and invest nothing in quality accordingly because there is no potential profit for production. In this situation, a cost of the penalty of unsatisfied demand will incur. But when the profit margin is high, the investment in quality will decrease which indicates the core company cares less about product quality when products become more lucrative. The right side of Table 24 shows a sharp increase of both production quantity and quality level when the sale price becomes high enough to make a profit.

Increasing the salvage value of product encourages more production because the risk of producing too much is dampened by an increasingly satisfactory compensation. Surprisingly, the safe ticket of high salvage value also disincentivizes investment in quality. Note when the salvage value is large enough, in our case when v equals 11, the company will produce as many as it can because the salvage value is larger than the total risk of product recall and low demand. This is why we set a lower limit on the fixed part of the unit cost and prevent the unrealistic situation of infinite production.

In the majority of the above parametric analysis, the optimal production quantity and optimal quality level change in opposite directions. For instance, when unit

k	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$	k	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
40	136	2.32	369	25	148	1.84	496
45	133	2.44	338	37.5	137	2.25	386
50	130	2.55	311	**50**	130	2.55	311
55	127	2.65	287	62.5	124	2.79	255
60	125	2.75	265	75	119	2.99	211

Table 22. Impacts of unit recall cost on optimal solutions.

p	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
4	121	2.59	364
5	125	2.57	337
6	130	2.55	311
7	134	2.53	286
8	138	2.52	262

Table 23. Impacts of penalty cost on optimal solutions.

s	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$	s	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
15	0	0	-60	17	0	0	-60
25	130	2.55	311	19	0	0	-60
35	167	2.45	1085	21	110	2.60	31
45	195	2.37	1922	23	120	2.58	168
55	217	2.31	2794	**25**	130	2.55	311

Table 24. Impacts of sales price on optimal solutions.

v	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$	v	Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
2	112	2.57	217	7	176	2.48	513
3	120	2.56	261	8	204	2.44	609
4	130	2.55	311	9	248	2.38	733
5	142	2.54	368	10	366	2.23	914
6	156	2.51	434	11	20000	1.62	16722

Table 25. Impacts of salvage value on optimal solutions.

		Q^*	ℓ^*	$\mathcal{P}^*(Q^*, \ell^*)$
$R(\ell) = \alpha e^{-\beta\ell}$	$\alpha \nearrow$	$\searrow$	$\nearrow$	$\searrow$
$C(\ell) = \gamma + \theta\ell$	$\gamma \nearrow$	$\searrow$	$\nearrow$	$\searrow$
$C(\ell) = \gamma + \theta\ell$	$\theta \nearrow$	$\searrow$	$\searrow$	$\searrow$
demand rate	$1/\lambda \nearrow$	$\nearrow$	$\longrightarrow$	$\nearrow$
unit recall cost	$k \nearrow$	$\searrow$	$\nearrow$	$\searrow$
penalty cost	$p \nearrow$	$\nearrow$	$\searrow$	$\searrow$
sales price	$s \nearrow$	$\nearrow$	$\nearrow\searrow$	$\nearrow$
salvage value	$v \nearrow$	$\nearrow$	$\searrow$	$\nearrow$

Table 26. Summary of increasing parameters' impacts on optimal solutions.

recalls cost increases, the optimal production quantity decreases while optimal quality level increases. This suggests with safer and higher quality products which have less profit margin, the core company should produce less of this type of product. One explanation may be that as the quality level increases, the punishment of overproduction overshadows the potential benefits of satisfying occasionally high demand. We observe two exceptions such that optimal solutions of quantity and quality level do not vary in divergent directions. One is changing the varied unit production cost, for which both production quantity and quality level decreases when varied production cost is high. The other exception is changing the demand rate, which affects optimal production quantity alone. We summarize the impacts of increasing parameter values on the optimal solutions in Table 26.

3.2.3 Modeling Demand with Erlang Distribution

To depict the various forms that demand distribution might take, we adopt the Erlang distribution for the demand random variable X. The probability density function is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where $n = 1, 2, \cdots$ with mean n/λ and variance n/λ^2. When parameter n equals one, the Erlang distribution reduces to the exponential distribution. Figure 21 shows that using an Erlang distributed random variable for demand brings flexibility in modeling.

We use the same parameters in Table 18 and experiment with Erlang parame-

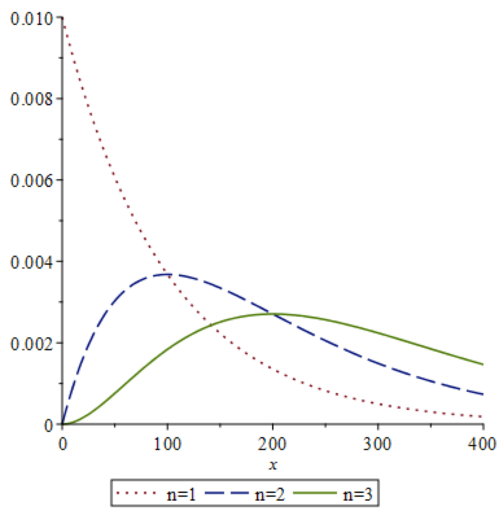

Figure 21. Erlang distributed demand X with $n = 1, 2, 3$ and $\lambda = 0.01$.

ter n increasing from 1 to 5. Table 27 shows the numerical results that both optimal ordering quantity Q^* and quality level ℓ^* increases as n increases. Our parametric analysis in Section 3.2.2 exemplifies that increasing average demand rate $1/\lambda$ in exponential distribution leads to higher optimal ordering quantity but has no impact on optimal quality level. Increasing Erlang parameter n also results in a larger average demand rate n/λ and increases both Q^* and ℓ^* concurrently. This difference may be attributed to the influence of n on the shape of Erlang distribution — when n grows, the probability density function curve is pressed down and shifted to the right with less skewness.

n	$\mathcal{P}^*(Q^*, \ell^*)$	Q^*	ℓ^*
1	311	130	2.55
2	1102	254	2.69
3	1992	373	2.76
4	2931	488	2.80
5	3899	601	2.83

Table 27. With Erlang distributed demand, both optimal ordering quantity and quality level increases when Erlang parameter n increases.

When applying first order optimality condition to the objective $\mathcal{P}(Q, \ell)$, we observe increasing Erlang parameter n has more notable impacts on the partial differential function $\partial\mathcal{P}(Q, \ell)/\partial Q$. Figure 22 shows that increasing Erlang parameter n does not significantly impact the shape of two implicit differential function curves comparing to using exponential distribution, but pushes the curves up and to the right, which explains the increase of both ordering quantity and quality level in optimal solution.

3.3 Extending the Model with External Suppliers

In practice, companies do not always own the manufacturing of its products or viewing production as its core competence. For instance, a typical warehouse retailer does not produce any of the products on its shelves; instead, it depends on multiple suppliers for high quality products supply. Therefore, we want to examine the impacts of using multiple external suppliers as the production sources. In this case, the product quality level is no longer within the direct control of the core company (the seller). However, the company can indirectly control the product quality level by offering suppliers higher purchase price on meeting higher quality standards.

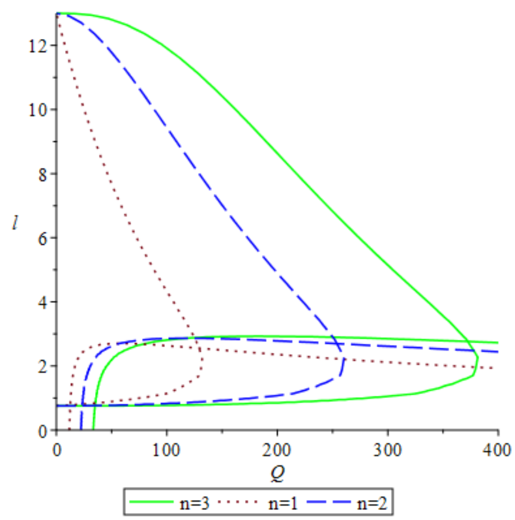

Figure 22. Implicit plots using first order optimality conditions shows increasing Erlang parameter n does not change the shape of two function curves significantly but pushes the curves up and to the right, which explains the increasements of both decision variables in optimal solution.

Supplier	S_1	S_2
Product quality level	ℓ_1	ℓ_2
Unit production cost	$c_1(\ell_1)$	$c_2(\ell_2)$
Recall probability	$R_1(\ell_1)$	$R_2(\ell_2)$
Order quantity	Q_1	Q_2
Demand random variable	$X_1 \sim f_1(x_1)$	$X_2 \sim f_2(x_2)$

Table 28. Variables and functions for the extension of two suppliers.

Here is a real life example from our neighborhood. Fortinos is a chained retailer owned by Loblaw which offers daily grocery for its customers. Loblaw sells products of its own brand Presidential Choice (PC) on a variety of convenient vegetable packages, organic produce, processed foods (cakes and ice cream) and etc. Of a common produce, spinach, Fortinos offers two types of products from two separate suppliers. One is the cooking spinach which costs $2.55 per 500 grams; the other is the organic baby spinach which costs $5.99 per 500 grams. The organic spinach would have higher quality standards on freshness, appearance and bacteria control along with other production requirement of being organic. While including two or more suppliers in its supply chain, Fortinos would seek the best ordering quantities and quality standards when managing its suppliers to optimize its expected profit.

3.3.1 Modeling with Two Suppliers

Suppose the seller collaborates with two suppliers (S_1 and S_2) that specialize in manufacturing. We use random variable X_1 with PDF $f_1(x_1)$ to denote demand satisfied by S_1, random variable X_2 with PDF $f_2(x_2)$ for demand satisfied by S_2, and X_1 and X_2 are independent. Suppose the seller is aware of the cost-quality relations of both suppliers from historical data. Based on the seller's required quality level, unit manufacturing cost from suppliers $(i = 1, 2)$ follows the function $c_i(\ell_i)$ where ℓ_i is the required product quality level. The seller also determines purchase quality of each supplier (Q_i) to maximize his expected total profits considering the risks of a potential product recall. The product recall probability of suppliers $(i = 1, 2)$ follows the function $R_i(\ell_i)$. Table 28 shows variables and functions for the two suppliers.

The focus of this proposed study is to examine impacts of the cost functions and recall probability functions on the seller's optimal order quantities, product quality levels, and the expected profits. There are four cases of possible recall situations, for which we use '0' to denote recall and '1' to denote good products with no recall and represent the four cases in Table 29.

Four events happen in the single period of decision making, including ordering

Supplier	S_1	S_2	Probability	Profit function
Both are good	1	1	$(1 - R_1(\ell_1))(1 - R_2(\ell_2))$	Π_{11}
Only S2 recalls	1	0	$(1 - R_1(\ell_1)) R_2(\ell_2)$	Π_{10}
Only S1 recalls	0	1	$R_1(\ell_1)(1 - R_2(\ell_2))$	Π_{01}
Both recall	0	0	$R_1(\ell_1) R_2(\ell_2)$	Π_{00}

Table 29. Four cases of suppliers recall situations.

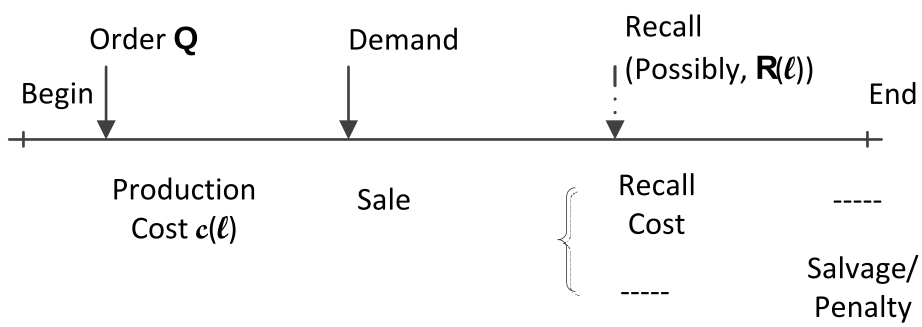

Figure 23. Four events in single period decision-making process.

products with quantity Q_i and quality level ℓ_i from supplier i ($i = 1, 2$), satisfying demand ($X_1 = x_1$ and $X_2 = x_2$) and obtain sale income, incurring recall costs if recall happens, or incurring penalty costs and collecting salvage values if recall does not happen. Figure 23 illustrates the sequence and outcome of the above four events.

Hence, the total expected profit ($E(\Pi)$) of one-stage stochastic newsvendor quality control model with two suppliers is

$$E(\Pi) = \sum_{i=0}^{1} \sum_{j=0}^{1} E(\Pi_{ij} \mid S_1 = i, S_2 = j) \Pr(S_1 = i, S_2 = j) \tag{33}$$

where

$$\begin{aligned}
\Pi_{11} &= s[\min(X_1, Q_1) + \min(X_2, Q_2)] - [c_1(\ell_1) Q_1 + c_2(\ell_2) Q_2] \\
&\quad + v[(Q_1 - \min(X_1, Q_1)) + (Q_2 - \min(X_2, Q_2))] \\
&\quad - p[(X_1 - \min(X_1, Q_1)) + (X_2 - \min(X_2, Q_2))] \\
\Pi_{10} &= s[\min(X_1, Q_1) + \min(X_2, Q_2)] - [c_1(\ell_1) Q_1 + c_2(\ell_2) Q_2] \\
&\quad + v[Q_1 - \min(X_1, Q_1)] - p[X_1 - \min(X_1, Q_1)] - k \min(X_2, Q_2) \\
\Pi_{01} &= s[\min(X_1, Q_1) + \min(X_2, Q_2)] - [c_1(\ell_1) Q_1 + c_2(\ell_2) Q_2] \\
&\quad + v[Q_2 - \min(x_2, Q_2)] - p[X_2 - \min(X_2, Q_2)] - k \min(X_1, Q_1) \\
\Pi_{00} &= s[\min(X_1, Q_1) + \min(X_2, Q_2)] - [c_1(\ell_1) Q_1 + c_2(\ell_2) Q_2] \\
&\quad - k[\min(X_1, Q_1) + \min(X_2, Q_2)]
\end{aligned}$$

and probabilities $\Pr(S_1 = i, S_2 = j)$ are shown in Table 29. This model may be extended for multiple suppliers by adding the set of decision variables (Q_i, ℓ_i) and functions $(c_i(\ell_i), R_i(\ell_i))$ for each supplier i ($i = 1, 2, \ldots, n$). We will explore the solution methods and discover insights for the proposed model of two or more suppliers.

3.3.2 Numerical Studies

We explore the impacts to the optimal solution of four cases: (1) the second supplier takes additional demand; (2) the two suppliers share demand evenly; (3) demand-sharing while decreasing production cost of the second supplier; (4) demand-sharing and two suppliers have different prices. Supplier One's parameters are the same as in base case for internal supply (Section 3.2.1).

3.3.2.1 Case 1. $S2$ takes additional demand.

α_1	β_1	γ_1	θ_1	λ_1	α_2	β_2	γ_2	θ_2	λ_2	k	p	s	v
0.9	1	5	2	$1/100$	0.96	1	6	2.5	$1/50$	50	6	25	4
Q_1^*		ℓ_1^*			Q_2^*		ℓ_2^*			$\mathcal{P}^*(Q_1^*, \ell_1^*, Q_2^*, \ell_2^*)$			
129.69		2.55			49.26		2.47			330.28			

Table 30. Parameter setting and optimal solutions for case 1 the second supplier takes additional demand.

The results suggest adding the second supplier does not affect the optimal solution of supplier one ($S1$) even that $S1$ may serve customers of $S2$ in case that $S2$ initiates a recall. This might attribute to the small probability setting of both suppliers, especially of the newly added supplier $S2$.

3.3.2.2 Case 2. $S1$ and $S2$ share the demand evenly.

When both suppliers now share the original setting of 100 demand evenly, the results show each supplier optimize its process separately as if they are supplying alone with the half demand. Reducing demand alone does not impact the optimal quality level, but reduces the optimal order quantity the same way demand alters, which is consistent with our numerical findings of internal supply case. We notice that supplier two ($S2$) has slightly lower quality level but much less order quantity in optimal solutions comparing to supplier one, the reason is its high production cost. Thus in case 3 we aim to examine the effects of improving production cost.

3.3.2.3 Case 3. Demand-sharing while decreasing production cost of $S2$.

By improving production cost of supplier two, the solutions show that both its quality level and ordering quality improve, which echo with results in our previous numerical analysis in Section 3.2.1.

3.3.2.4 Case 4. Demand-sharing and two suppliers have different prices.

α_1	β_1	γ_1	θ_1	λ_1	α_2	β_2	γ_2	θ_2	λ_2	k	p	s	v
0.9	1	5	2	**1/50**	0.96	1	6	2.5	1/50	50	6	25	4
Q_1^*		ℓ_1^*			Q_2^*		ℓ_2^*			$\mathcal{P}^*(Q_1^*, \ell_1^*, Q_2^*, \ell_2^*)$			
64.84		2.55			49.26		2.47			170.80			

Table 31. Parameter setting and optimal solutions for case 2 demand-sharing evenly between two suppliers.

α_1	β_1	γ_1	θ_1	λ_1	α_2	β_2	γ_2	θ_2	λ_2	k	p	s	v
0.9	1	5	2	1/50	0.96	1	**4**	**1.5**	1/50	50	6	25	4
Q_1^*		ℓ_1^*			Q_2^*		ℓ_2^*			$\mathcal{P}^*(Q_1^*, \ell_1^*, Q_2^*, \ell_2^*)$			
64.84		2.55			84.41		2.79			475.49			

Table 32. Parameter setting and optimal solutions for case 3 demand-sharing and cost-improved second supplier.

α_1	β_1	γ_1	θ_1	λ_1	α_2	β_2	γ_2	θ_2	λ_2	k	p	$\mathbf{s}_1$	$\mathbf{s}_2$	v
0.9	1	5	2	1/50	0.96	1	4	1.5	1/50	50	6	**30**	**20**	4
Q_1^*		ℓ_1^*			Q_2^*		ℓ_2^*			$\mathcal{P}^*(Q_1^*, \ell_1^*, Q_2^*, \ell_2^*)$				
74.93		2.50			72.26		2.86			465.86				

Table 33. Parameter setting and optimal solutions for case 4 demand-sharing and different prices.

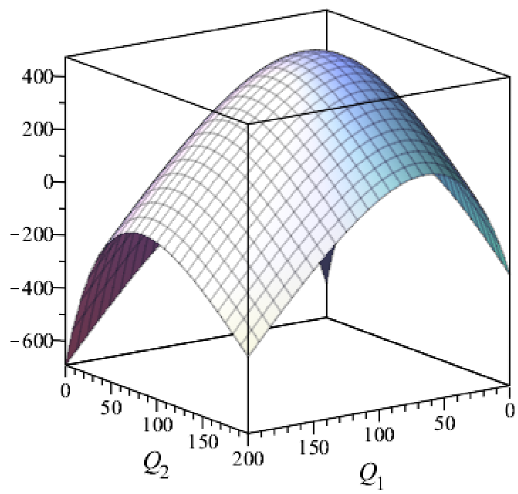

Figure 24. Expected profit $(E(\Pi))$ is concave on ordering quantities when fixing optimal quality levels ℓ_1^* and ℓ_2^*.

Case 4 is modified based on case 3 that both suppliers share the demand evenly but in this case supplier $S1$ has higher price (\$30) and supplier $S2$ has lower price (\$20) compared to the unified price \$25 in case 3. With increased price, the optimal ordering quantity increases by 10.09 for $S1$ but the quality level decreases by 0.05. Similarly for $S2$ as the price decreases the optimal ordering quantity decreases by 12.15 and the quality level increases by 0.07. These results are consistent with our analysis of price changes for internal supply in Section 3.2.1.

One surprising result shown in case 4 is that the more expensive products of $S1$ have lower optimal quality level than the less expensive products of $S2$ comparing their cost structure. This counter intuitive result suggests a high cost and high selling price product could have lower optimal quality level because higher profit margin does not incent quality improvement. On the contrary it may accommodate the recall risks and allow lower quality.

3.3.3 Concavity Results

We explore the concavity of the proposed model (33) by isolating the ordering quantities/quality levels while setting the remaining decision variables with their optimal solution values. The results show the expected profit function is concave on both ordering quantities and quality levels as shown in Figure 24 and Figure 25 respectively.

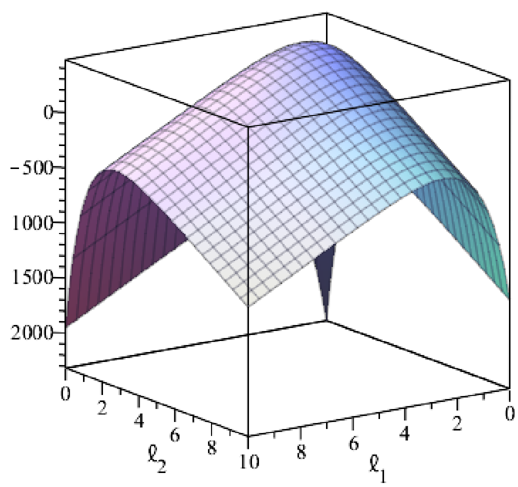

Figure 25. Expected profit ($E(\Pi)$) is concave on the quality levels when fixing optimal order quantities Q_1^* and Q_2^*.

3.4 Conclusion

In this chapter we extend the newsvendor problem for product recall risk management by optimizing both quality level and ordering quantity. Our main contribution is that we model and solve the optimization problem of product recall management by incorporating quality control in the classic newsvendor problem. We add quality-dependent functions for recall probability and production cost so that recall risks can be well quantified in the objective function.

Our first discussion centers on the internal supply case in which the core company has direct control of quality level. Concavity analysis shows the objective is neither concave nor convex in the feasible region, but applying the first and second order optimality conditions can effectively find the optimal solution. Our numerical analysis shows in most cases an inverse relationship between optimal ordering quantity and quality level as summarized in Table 26. One counter-intuitive finding is that higher selling price does not incentivize investment on quality levels, instead it promotes lower quality level since higher profit margin can accommodate higher potential costs of recall.

We extend our model to external multiple suppliers, especially the case of two suppliers. Both suppliers satisfy their own demands independently and supply for each other's customers only when the other party has a recall. Our results show that the prospect of covering the other party's demand in case of recall has very little impact on either supplier's optimal ordering quantity and quality level. In other words, the optimal solutions are determined by parameter and function settings of the individual supplier. Our concavity analysis suggests that the objective is concave for quality levels (ordering quantities) of the two suppliers when ordering quantities (quality levels) are set to be optimal.

Chapter 4
Optimal Facility Location to Mitigate Product Recall Risks

This chapter is motivated by major food product recall events in recent years, especially how the timely and effective response using post-recall management can make a difference. We consider the rare but very influential major product recalls as disruptions to the supply chain and incorporate locating reprocessing centers for the returned products to mitigate expected operational costs. We adopt the closed loop network design framework and assume the location decisions for reprocessing center take place after the product recall events. Our scenario based analysis shows the approach is effective in both absolute and relative measures.

4.1 Introduction

In September 2012, the rejection of ground beef imports by the U.S. custom and the later outbreak of the E.coli disease forced XL Inc. (XL) to start a series of beef recalls, which turned out to be the largest meat recall event in Canada history. The recall of 1,800 products impacted over 33 retail chains across Canada. Over 4,000 tons of meat and meat products were sent back to plants for disposal. This unprecedented amount of recalled products overwhelmed the capacities of any existing disposal methods of XL, the largest domestically owned meat processor in Canada at the time, resulting in 600 tons of frozen beef being sent directly to landfill. The failure to adequately process recalled products raised the public concern over XL's capability to maintain food safety. This incident eventually led to a transfer of XL's ownership for the Calgary plant. Charlebois et al. [9] discussed this event and its impacts in detail.

In comparison, when facing the 2008 Listeria outbreak, Maple Leaf Food Inc. (MLF) recalled all potentially affected products promptly and dealt with the disposal of 1,300 tons of beef and beef products. MLF's actions bought time to discover the contamination source and recover the brand. Nevertheless, there was a substantial direct cost of $19 million related to recall activities (e.g., collection and destruction, shutdown and sanitation of facility, media, and customer response call center). In total, MLF suffered approximately $200 millions loss in this incident.

Food safety is generally referred to as the prevention of illness resulting from the consumption of contaminated food as discussed by Akkerman [1]. This topic has attracted more attention recently because of the growing rigorous government standards as well as the large social and financial impacts of major food safety failure. Effective control of food safety along the supply chain is important but very complicated because food is vulnerable to contamination and food supply chains are sophisticated. For example, Desmarchelier et al. [15] provide a summary of food safety management in the red meat industry of Australia. Risk mitigation strategies are applied in the entire food supply chain.

However, food borne disease is a critical inherent risk factor in food manufacturing and distribution, due to the indigenous existence of microbial contaminations in raw food materials, natural growth of pathogens, inevitable mistakes in manual operations, contamination and other factors. In response to food safety incidents, business and society adopt food recalls to correct the situation and mitigate monetary and social costs. Manufacturing companies manage food recalls by collecting products from their distribution channels and adopting best methods to recondition and dispose recalled products.

Proper preparation could help companies manage food recalls more efficiently and effectively, especially in strategic planning. For example, firms could use location-allocation decisions for both manufacturing plants and reprocessing facilities for recalled products. Linking to optimization studies, Akkerman et al. [1] provide a review of improving food supply chain management with network planning models. Food quality, food safety and sustainability are considered as key objectives. Three levels of network planning models are considered, namely strategic network design (e.g., facility location-allocations), tactical network planning (e.g., production and distribution) and operational transportation planning (e.g., routing). They suggest that strategic network design is critical in food safety control, impacting how long food products travel and how widely the products spread geographically, both of which determine the size of potential product recalls.

This chapter addresses the supply chain safety control issue by designing the supply chain network to incorporate the negative effects of product recalls. Although our model is motivated by food recall, it can be applied to any supply chain with significant impacts of recall events. Extending from the closed-loop network design concept, this work focuses on managing the reverse flow (recalled products) in a cost-efficient manner. However, rather than maintaining a closed-loop supply chain on a daily basis as in most studies (e.g., repair and post-sale service systems), we study the efficient way of managing random and rare major product recalls. The features of rareness and randomness of major products recalls lead us to disruption management studies in which researchers focus on how to consistently satisfy customer demands given that some suppliers may fail. Our focus is different in that we consider how to quickly build a reprocessing network to dispose recalled products.

We study the location-allocation problem with random occurrence of product recalls and treat the recall incidences as disruptions to the supply chain. In our setting, the company first makes decisions to locate manufacturing plants and allocate demands. After the product recall occurs, we make decisions to locate the reprocessing center(s) from internal (self-owned recall facility) or external (third-party business) sources and allocate recalled products for reprocessing or local disposal. Three features distinguish our problem from other location-allocation problems. Firstly, facility location and allocation decisions occur in two stages. Secondly, the second stage location-allocation happens under uncertainty. Thirdly, reverse logistic flows exist in the second stage.

We design a two-stage stochastic mixed integer programming model, in which we locate the manufacturing plants in the first stage and the reprocessing/disposal facilities in the second stage. We adopt a scenario-based approach to describe the uncertainty of major recall events that may happen in manufacturing plants as well as of availability of reprocessing facilities. Given the complexity induced by our nested facility location problem, we devise an algorithm based on Lagrangian relaxation to solve the uncapacitated case.

This chapter will be organized as follows. Section 4.2 provides a brief review of literature. Section 4.3 introduces the mathematical model, an analysis based on facility capacities, and a Branch-and-Bound algorithm incorporating Lagrangian relaxation for the uncapacitated case. Section 4.5 presents the computational results and managerial insights from experiments. Section 4.6 summarizes contributions and discusses further research directions.

4.2 Literature Review

There are two research streams closely related to our research, i.e., reliable supply chain network design, and location-allocation with bidirectional logistical flows.

There is a well-developed literature on modeling supply chain disruption management. Snyder et al. [48] provide a review of optimization models in supply chain network planning with disruption management. They categorize by network status, underlying mathematical models and risk measures. They show various models extending the classical P-median and Uncapacitated Facility Location Problem (UFLP), Capacitated Facility Location Problem (CFLP) models with reliability features (i.e., consistent satisfaction of demands when some facilities fail in random disruptions). Qi et al. [37] consider a fortification model with disruptions. They manage locations, allocations and inventory. Both suppliers and retailers can experience random disruptions. Qi et al. [36] examine different sourcing and replenishment decisions with two suppliers. Dada et al. [11] develop a newsvendor procurement model selecting from multiple unreliable suppliers. Their results suggest that newsvendor, customers and retailers perceive different service level changes when disruption occurs. Compared to reliability concerns, cost is the most determinant factor. This body of work focuses on satisfying customer demand – there is no focus on reverse flow.

On the other hand, in closed-loop supply chain models, the emphasis is to minimize the long-run average cost of forward and reverse flows, which does not model the random occurrence of product recalls. Uncertainty in closed-loop supply chain network design is not always considered. Some studies use a fixed return rate based on historical data (e.g., Min and Ko [31], Salema et al. [41], Lu and Bostel [25], and Lee and Dong [22]). Savaskan et al. [45] describe product return rate as a function of investment used to promote product return.

Hitherto, uncertain factors in closed-loop supply chain include return rate, demand for re-manufactured products (secondary market), quality of returned products, and variable costs for collection, processing and transportation. Pishvaee et al. [34] describe uncertain demands, returns and transportation costs given by a robust network design model. Salema et al. [42] address uncertain demand and return with scenario dependent uniform random numbers to minimize the total cost of the reverse logistics network. Ramezani et al. [38] consider uncertainty in demand and return ratio as well as various variable costs in their multi-objective forward/reverse network design. Listeş [24] uses scenario-dependent parameters to describe uncertain demand and returned product quantity for their supply and product-return networks. Few researchers assume the scenario of randomly failing manufacturing plants which results in major recall events.

The literature has taken different approaches to deciding the quantity to dispose during returned products collection and reprocessing. Early literature tends to not consider disposal cost. For instance, Savaskan et al. [45] do not take disposal as a cost factor when comparing different re-manufacturing channels. With increasing focus on environment-friendly and efficient supply chains, disposal costs are reflected in later studies. For example, Min and Ko [31] alter the repair facility capacities to accommodate returned quantities at each time period so no disposal will occur.

Decisions regarding disposal cost can be categorized in three types: fixed ratios, market driven and cost driven. Fixed disposal ratios are deduced from historical data and adopted to simplify the model (e.g., Pishvaee et al. [34], Lu and Bostel [25], and Lee and Dong [22]). Cost driven decisions aim to minimize total cost of collection, reprocessing and disposal. Salema et al. [42] minimize total supply chain costs by using fraction of customer demand used for disposal or recovery. Ramezani et al. [38] use disposed quantity to maximize the total profit within the capacity of opened disposal centers. Market driven decisions select the best efforts to satisfy demands of secondary market. Pishvaee et al. [34] model the disposal quantity as decision variable so that secondary market can be satisfied in the most efficient way. Listeş [24] take the perspective that returned products can be disposed in two decisions: before collection and before reprocessing, both of which aim to maximize the total profit while satisfying market demands.

In summary, the existing literature treats reverse flow on day-to-day basis. This modeling approach does not serve major product recalls well. Our challenge is to design an optimal network that can accommodate product returns in the context of major product recalls.

4.3 Facility Location to Mitigate Recall Risks

4.3.1 Model Development

In this section, we use a two-stage stochastic programming approach described by Ruszczynski and Shapiro [40] to model the problem. In the first stage, we make facility location and transportation decisions. In the second stage, under each disruption scenario, we make recall decisions including recall facility locations (e.g., centers for reconditioning, reprocessing, and rendering) and recall allocation decisions (e.g., use local disposal or recall center). The objective is to minimize the sum of facility location costs, transportation costs and recall costs.

Define $\mathcal{I}$ as the set of candidate locations for manufacturing facilities, and use i as the index. Define $\mathcal{K}$ as the set of candidate locations for recall centers and use k as the index. Note that recall centers could reuse the manufacturing facilities, or use third party processing facilities. Thus we could have $\mathcal{I} \subset \mathcal{K}$ or $\mathcal{I} \bigcap \mathcal{K} = \emptyset$. We use FF and RF to indicate forward product flow and reverse product flow (initiated by recall events) respectively. Clearly, facilities built for manufacturing and recall processing have different fixed costs, defined as f_i^{FF} and f_k^{RF}. Facility capacity is denoted by M_i^{FF} or M_k^{RF}.

Define $\mathcal{J}$ as the set of retailers and use j as the index. Each retailer has a demand D_j, the cost of shipping one unit of product demand from facility i to retailer j is c_{ij}^{FF}, while the reverse flow costs c_{jk}^{RF} per unit. For recalled products, two recall modes are available, i.e., local disposal and central processing. Local disposal incurs a retailer location related cost c_j^{LD}, and central processing incurs a recall center related cost c_i^{CP}.

To describe the uncertainty of facility disruption, we use scenario set $\mathcal{S}$ and index s. We used p_s to denote the probability of scenario s. Manufacturing facilities failing in scenario s is denoted by set $\mathcal{I}_s$. Accordingly, recall centers available in scenario s is denoted by set $\mathcal{K}_s$. The choice of $\mathcal{K}_s$ can be decided extraneously.

For instance, recalled products could be prohibited from returning to their original manufacturing facility due to safety concerns, or be sent to third party facilities due to economic considerations.

Decision variables used in this model are facility location variables ($\boldsymbol{X}$ for manufacturing facilities and $\boldsymbol{Z}$ for recall facilities), transportation variables ($\boldsymbol{Y}$), and recall assignment variables ($\boldsymbol{V}$ for local disposal and $\boldsymbol{W}$ for central processing):

$$X_i = \begin{cases} 1 & \text{manufacturing facility at location } i \text{ is open} \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_{ks} = \begin{cases} 1 & \text{recall facility is open at location } k \text{ under scenario } s \\ 0 & \text{otherwise.} \end{cases}$$

Y_{ij} : quantity transported from facility i to retailer j

V_{js} : quantity for local disposal under scenario s at retailer j

W_{jks} : quantity from retailer j to plant k for central processing under s

With these notations, the two-stage stochastic program for the Facility Location with Recall Problem ($\mathcal{FLRP}$) is formulated as follows:

$$
\begin{aligned}
(\mathcal{FLRP})\ \min \quad & \sum_{i\in\mathcal{I}} f_i^{FF} X_i + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} c_{ij}^{FF} Y_{ij} \\
& + \sum_{s\in\mathcal{S}} p_s \left[\sum_{k\in\mathcal{K}_s} f_k^{RF} Z_{ks} + \sum_{k\in\mathcal{K}_s}\sum_{j\in\mathcal{J}} \left(c_{jk}^{RF} + c_k^{CP}\right) W_{jks} + \sum_{j\in\mathcal{J}} c_j^{LD} V_{js} \right] && (34) \\
\text{s.t.} \quad & \sum_{i\in I} Y_{ij} = D_j \quad \forall j \in \mathcal{J} && (35) \\
& \sum_{j\in J} Y_{ij} \le M_i^{FF} X_i \quad \forall i \in \mathcal{I} && (36) \\
& \sum_{j\in J} W_{jks} \le M_k^{RF} Z_{ks} \quad \forall s \in S, k \in \mathcal{K}_s && (37) \\
& \sum_{k\in\mathcal{K}_s} W_{jks} + V_{js} = \sum_{i\in I_s} Y_{ij} \quad \forall j \in \mathcal{J}, s \in \mathcal{S} && (38) \\
& Y_{ij} \ge 0 \quad \forall i \in \mathcal{I}, j \in \mathcal{J} && (39) \\
& W_{jks}, V_{js} \ge 0 \quad \forall j \in J, s \in \mathcal{S}, k \in \mathcal{K}_s && (40) \\
& X_i \in \{0,1\} \quad \forall i \in \mathcal{I} && (41) \\
& Z_{ks} \in \{0,1\} \quad \forall s \in \mathcal{S}, k \in \mathcal{K}_s && (42)
\end{aligned}
$$

The objective (34) is the expected total cost of facility location-allocation decisions from both stages. In the first stage, the manufacturing facility location and demand distribution are determined, while in second stage, recall center location and recall product distribution are determined. Constraint (35) ensures demand of each retailer is satisfied. Constraint (36) and (37) guarantees forward and recalled products are processed in an open facility within the capacities. Constraint (38) requires recalled products either be sent back to recall center for reprocessing or disposed locally. Other constraints are for non-negativity and for binary variables.

It is noteworthy that the proposed formulation is very general and can include different recall situations. Firstly, the model can cover cases of capacitated and uncapacitated facilities or any hybrid types by changing the capacity parameter M_i^{FF} and M_k^{RF} to be infinite or finite numbers. Secondly, the definition of set $\mathcal{K}_s$ provides a lot of flexibility in modeling. For instance, facilities in $\mathcal{K}_s$ can be the exact locations where the food safety incidents happen, suggesting that recalled products must return to their original manufacturing plants; a type of corrective action in this situation is to fix problems such as mislabeling. We can also require facilities in $\mathcal{K}_s$ be locations other than the original manufacturing facilities to model the situation in which the original facilities are unsuitable for processing (e.g., safety overhaul). Moreover, $\mathcal{K}_s$

could be third party facility locations to model the outsourcing of processing recalled products. Thirdly, for further extension, recall facility location decisions Z_{ks}'s could be moved to the first stage. The difference between this extension and $(\mathcal{FLRP})$ will show the difference between prepared recall facility location decisions and reactive facility location decisions.

In addition, the model could indicate different traceability capability in forward flow. Currently, we assume full traceability in forward flow, i.e., each retailer could distinguish the products from failed plants from others. We can also model incomplete traceability such that retailer can't identify the source of the recalled products, by altering constraint (38) to $\sum_{k\in\mathcal{K}_s} W_{jks} + V_{js} = \sum_{i\in\mathcal{I}} Y_{ij}$ for all $j \in \mathcal{J}$ and $s \in \mathcal{S}$.

4.3.2 Analysis

If the demand of a retailer is satisfied by more than one facility, it is called "demand splitting", otherwise, it is called "no demand splitting". Note that $(\mathcal{FLRP})$ contains two types of demands, corresponding to the two stages (i.e., customer demand and recall demand). To be concise, we use "plants" to denote manufacturing centers located for forward flows, and "facilities" to denote processing centers located for reverse flows. We have the following result in the optimal demand splitting schemes of $(\mathcal{FLRP})$:

Theorem 7 *In the optimal solutions of $(\mathcal{FLRP})$, demand splitting schemes of forward and reverse flows depend on the capacity constraints of both plants and facilities. There are four schemes:*

(a) If plants and facilities are both uncapacitated, then there exists no demand splitting for both forward and reverse flows;
(b) If plants are capacitated but facilities are uncapacitated, then there exists demand splitting for forward flow and no demand splitting for reverse flow;
(c) If plants are uncapacitated but facilities are capacitated, then there could exist demand splitting for both flows;
(d) If plants and facilities are both capacitated, then there could exist demand splitting for both flows.

Proof. Consider an arbitrary retailer $u \in \mathcal{J}$. Assume that except the allocation decisions for u, all other decisions are fixed to their optimal values. We use superscript $*$ to represent the optimal values. These decisions include:

$X_i^*(i \in \mathcal{I})$ locating plants for forward flow,

$Z_{ks}^*(k \in \mathcal{K}_s,\ s \in \mathcal{S})$ locating facilities for reverse flow,

$Y^*_{ij}(i \in I,\ j \in \mathcal{J}\backslash\{u\})$ allocating demands of retailers other than u in forward flow,

$V^*_{js}, W^*_{jks}\,(j \in \mathcal{J}\backslash\{u\},\ k \in \mathcal{K}_s,\ s \in \mathcal{S})$ allocating recall demands of retailers other than u for central reprocessing and local disposal respectively in reverse flow.

In the following, we discuss the demand splitting schemes in two steps. We isolate demand splitting decisions in the reverse flow in the first step, and apply the results in the second step to integrate both forward and revers flows. Note we use $*$ to denote optimal solutions for $(\mathcal{FLRP})$ in this proof, while later in algorithm design we use $*$ to denote best solutions found in Lagrangian relaxation.

Step 1. Demand allocations in forward flows, i.e., Y^*_{iu} $(i \in \mathcal{I})$, are known and optimal.

We use $\mathcal{K}^*_s$ to denote the set of facilities to open for reverse flow in scenario $s \in \mathcal{S}$ in the optimal solution. Let β_{ks} be the proportion of demand for retailer u allocated to facility $k \in \mathcal{K}^*_s$ in scenario s, β'_s be the proportion of demand allocated to local disposal in scenario s. We have $\sum_{k\in K^*_s} \beta_{ks} + \beta'_s = 1$ in scenario s.

Let D^*_{us} be the optimal recall demand from u in scenario s. Note that when forward flow is traceable, we have $D^*_{us} = \sum_{i\in\mathcal{I}_s} Y^*_{iu}$; when forward flow is untraceable, we have $D^*_{us} = \sum_{i\in\mathcal{I}} Y^*_{iu}$ if $Y^*_{iu} > 0$ for some $i \in \mathcal{I}_s$, and $D^*_{us} = 0$ otherwise.

Using Ω to denote the total cost of known location and allocation decisions, $(\mathcal{FLRP})$ can be simplified as follows:

$$
\begin{aligned}
\min \quad & \Omega + \sum_{s\in\mathcal{S}} p_s \left[\sum_{k\in\mathcal{K}^*_s} \left(c^{RF}_{uk} + c^{CP}_k\right) \beta_{ks} D_{us} + c^{LD}_u \beta'_s D_{us} \right] \\
\text{s.t.} \quad & \sum_{k\in\mathcal{K}^*_s} \beta_{ks} + \beta'_s = 1 \\
& \beta_{ks} D_{us} + \sum_{j\in\mathcal{J}\backslash\{u\}} W^*_{jks} \le M^{RF}_k \qquad \forall k \in \mathcal{K}^*_s, s \in \mathcal{S} \\
& \beta_{ks}, \beta'_s \ge 0 \qquad \forall k \in \mathcal{K}^*_s, s \in \mathcal{S}.
\end{aligned}
$$

Note that local disposal at retailer u always has unlimited capacity. This setting can be altered by adding constraint $\beta'_s D_{us} \le M^{RF}_u$ in the analysis, where M^{RF}_u is the capacity for local disposal of u.

If the facilities are uncapacitated, the optimal solution is either 0 or 1 because the objective function is a linear function of β_{ks} and β'_s defined on interval $[0, 1]$. In other words, there is no demand splitting for any scenario $s \in \mathcal{S}$. Denote

$$
\gamma^*_s = \min\left\{ c^{LD}_u, \min_{k\in\mathcal{K}^*_s} \left\{ c^{RF}_{uk} + c^{CP}_k \right\} \right\},
$$

and optimal recall cost of u as $\psi^*_s(D^*_{us}) = \gamma^*_s D^*_{us}$, then the optimal objective value is

$\Omega + \sum_{s\in\mathcal{S}} p_s \psi_s^*(D_{us}^*)$.

If the facilities are capacitated, in order to minimize the objective function, we sort the coefficients $(c_{uk}^{RF} + c_k^{CP})$ $(k \in \mathcal{K}_s^*)$ and c_u^{LD} in increasing order, denoted as $c_1, c_2, \cdots, c_{n_s+1}$ where $n_s = |\mathcal{K}_s^*|$. Then the optimal solution is to assign demand according to this order given the capacity constraints. We redefine reverse flow cost $\psi_s^*(D_{us}^*)$ as follows:

$$\psi_s^*(D_{us}^*) = \begin{cases} c_1 D_{us}^* & D_{us}^* \leq M_1^{RF} \\ c_{k+1}\left(D_{us}^* - \sum\limits_{i=1}^{k} M_i^{RF}\right) + \sum\limits_{i=1}^{k} c_i M_i^{RF} & \sum\limits_{i=1}^{k} M_i^{RF} \leq D_{us}^* \leq \sum\limits_{i=1}^{k+1} M_i^{RF} \\ & \forall k = 1, \cdots, n_s \end{cases}$$

where $c_1 \leq c_2 \leq \cdots \leq c_{n_s} \leq c_{n_s+1}$. Note that if the unit cost of local disposal is ranked before that of facility j, then calculations for facility j and its followers are unnecessary because local disposal is uncapacitated.

We have the same form of optimal objective value $\Omega + \sum_{s\in\mathcal{S}} p_s \psi_s^*(D_{us}^*)$. Note that reverse flow cost $\psi_s^*(D_{us}^*)$ is a continuous and nondecreasing piece-wise linear function of D_{us}^* when facility capacities are limited. Clearly $\psi_s^*(\cdot)$ is a convex function.

Step 2. In the optimal solution, denote $\mathcal{I}^*$ as the set of plants open in forward flow, $\mathcal{I}_s^*$ as the set of plants that are open and get disrupted in scenario $s \in \mathcal{S}$. Let α_i be the proportion of demand allocated to plant $i \in \mathcal{I}^*$ for forward flows, $\psi_s^*(D_{us}^*)$ be the optimal cost for reverse flow as defined in Step 1, where $D_{us}^* = \sum_{i\in\mathcal{I}_s^*} \alpha_i D_u$ (or in the case with untraceable demand, if $Y_{iu}^* > 0$ for some $i \in \mathcal{I}_s^*$, then $D_{us}^* = D_u$). Let Ψ be the cost of all known optimal decisions. $(\mathcal{FLRP})$ can be simplified as follows:

$$\begin{aligned} \min \quad & \Psi + \sum_{i\in\mathcal{I}^*} c_{iu}^{FF} \alpha_i D_u + \sum_{s\in\mathcal{S}} p_s \psi_s^* \sum_{i\in\mathcal{I}_s^*} (\alpha_i D_u) \\ \text{s.t.} \quad & \sum_{i\in\mathcal{I}^*} \alpha_i = 1 \\ & \alpha_i D_u + \sum_{j\in\mathcal{J}\setminus\{u\}} Y_{ij}^* \leq M_i^{FF} && \forall i \in \mathcal{I}^* \\ & \alpha_i \geq 0 && \forall i \in \mathcal{I}^* \end{aligned}$$

If facilities are uncapacitated, then $\psi_s^*(D_{us}^*) = \gamma_s^* D_{us}^* = \gamma_s^* \sum_{i\in\mathcal{I}_s^*} \alpha_i D_u$. Clearly, the objective function is linear in α_i's. It holds that when plants are uncapacitated, the plant incurring the lowest cost will be chosen, thus the optimal solution does not have demand splitting for neither forward nor reverse flows, which completes the proof for Theorem 7 Scheme (7). On the other hand, with capacitated plants the optimal demand splitting scheme for reverse flow remains to be no demand splitting; while capacity limits of plants in forward flow require demand splitting to satisfy all demands at the lowest cost, which completes the proof for Theorem 7 Scheme (7).

In the case of capacitated facilities and uncapacitated plants, apparently demand splitting in reverse flow exists because of facility capacity limits, would the demand splitting in reverse flow lead to demand splitting in forward flow even when there is no restriction on plant capacity?

Suppose, on the contrary, there is no demand splitting in the optimal decision. Consider the retailer u has demand D_u, served by two plants $\mathcal{I}^* = \{i_1, i_2\}$ and two facilities $\mathcal{K}_s^* = \{k_1, k_2\}$ that are open for all scenarios. Plants fail independently. Let us compare the following two cases.

In the first case, D_u is completely served by plant $i_1 \in \mathcal{I}^*$. Denote the probability of failure of i_1 as $q_{i_1} = \sum_{s:i_1 \in \mathcal{I}_s^*} p_s$, called the failure rate of i_1. The allocation cost TC_1 for i_1 can be represented as $TC_1 = c_{i_1u}^{FF} D_u + q_{i_1}\psi(D_u)$. Note we use $\psi(\cdot)$ instead of $\psi^*(\cdot)$ because it is not optimal solution. In the second case, D_u is satisfied by two plants i_1 and i_2. Let α_{i_1} and α_{i_2} denote the proportions of demand D_u satisfied by i_1 and i_2 respectively, where $\alpha_{i_1} + \alpha_{i_2} = 1$. Let the failure rates of plants i_1 and i_2 be q_{i_1} and q_{i_2} respectively. Then the total allocation cost TC_2 would be $TC_2 = \alpha_{i_1} c_{i_1u}^{FF} D_u + \alpha_{i_2} c_{i_2u}^{FF} D_u + q_{i_1}(1-q_{i_2})\psi(\alpha_{i_1}D_u) + (1-q_{i_1})q_{i_2}\psi(\alpha_{i_2}D_u) + q_{i_1}q_{i_2}\psi(D_u)$. Comparing TC_1 and TC_2:

$$\begin{aligned} TC_1 - TC_2 =& (1-\alpha_{i_1})c_{i_1u}^{FF} D_u - \alpha_{i_2} c_{i_2u}^{FF} D_u - q_{i_1}(1-q_{i_2})\psi(\alpha_{i_1}D_u) \\ & - (1-q_{i_1})q_{i_2}\psi(\alpha_{i_2}D_u) + q_{i_1}(1-q_{i_2})\psi(D_u) \\ =& \alpha_{i_2}(c_{i_1u}^{FF} - c_{i_2u}^{FF})D_u + (q_{i_1} - q_{i_2})\psi(\alpha_{i_2}D_u) \\ & + q_{i_1}(1-q_{i_2})\left[\psi(D_u) - \psi(\alpha_{i_1}D_u) - \psi(\alpha_{i_2}D_u)\right]. \end{aligned}$$

Because $\psi(\cdot)$ is nondecreasing and convex, we have $\psi(D_u) - \psi(\alpha_{i_1}D_u) - \psi(\alpha_{i_2}D_u) \geq 0$. If $c_{i_1u}^{FF} - c_{i_2u}^{FF} = q_{i_1} - q_{i_2} = 0$, then $TC_1 - TC_2 \geq 0$. Otherwise, by setting external parameters, we can always construct a problem where in the optimal solution, demand splitting is better than no demand splitting.

In more general case with multiple (more than two) plants and facilities available to serve the retailer, we can also construct a problem which has an optimal solution with demand splitting, which completes the proof for Scheme (7). In addition, a numerical example is given in Example 1. Finally, when plants and facilities are both capacitated, from the same analysis, the claim in Scheme (7) holds. ■

The most interesting result in Theorem 7 is the existence of demand splitting in forward flow when the (reverse flow) facilities are capacitated, as shown in Scheme (7). We use a numerical example to illustrate Theorem 7.

Example 1 *Suppose we have all the information of the optimal solution except demand splitting scheme for retailer u. With demand $D_u^* = 10$, retailer u has two plants*

$\mathcal{I}^* = \{1, 2\}$ available in forward flow, and two facilities $\mathcal{K}_s^* = \{3, 4\}$ for any scenario $s \in \mathcal{S}$ available in reverse flow. We assume the manufacturing facilities fail independently with probability $q_1 = 0.1$ and $q_2 = 0.9$ respectively. We have four scenarios (with probability of $p_1 = 0.81,\ p_2 = p_3 = 0.09,\ p_4 = 0.01$) for all possible outcomes.

Transportation unit costs in forward flow are $c_{1u}^{FF} = 1, c_{2u}^{FF} = 30$, unit cost of transportation and central processing for reverse flow are $c_{u3}^{RF} + c_{u3}^{CP} = 2, c_{u4}^{RF} + c_{u4}^{CP} = 60$, local disposal unit cost is $c_u^{LD} = 100$. Suppose plants and local disposal location have unlimited capacity, and capacities of central processing centers are $M_3^{RF} = 5, M_4^{RF} = 60$.

We use TC to represent the expected total allocation cost of u. Note facility 3 and 4 are capable of processing any amount of returned product from u, thus the local disposal at u need not to be considered (more expensive), which leaves TC to be solely dependent on the demand splitting factor α_1, i.e., the proportion of D_u satisfied by plant 1 in forward flow. With no demand splitting, customer u is solely served by plant 1 with $TC = 289$.

$$\begin{aligned}
\alpha_1 &= 1 \\
TC &= \sum_{i=1}^{2} c_{iu}^{FF} \alpha_i D_u + + \sum_{s \in \mathcal{S}} p_s \psi_s \left(\sum_{i \in \mathcal{I}_s^*} \alpha_i D_u \right) \\
&= 1 \cdot 10 + 0.81 \cdot (5 \cdot 2 + 5 \cdot 60) + 0.09 \cdot (5 \cdot 2 + 5 \cdot 60) \\
&= 289
\end{aligned}$$

Similarly when u is solely supplied by plant 2 we have $TC = 355.8$. However, if the demand is shared by the two plants evenly we have $TC = 191.9$. The results show demand splitting scheme with $\alpha_1 = 0.5$ gives lower cost than no demand splitting ($\alpha_1 = 1$ or $\alpha_1 = 0$). Moreover, solving the problem with MAPLE shows that $\alpha_1 = 0.5$ is the optimal solution, which complies with our statement in Theorem 7 Scheme (7).

4.4 Lagrangian relaxation

To find the optimal solution for model ($\mathcal{FLRP}$), we could use readily available commercial solvers such as CPLEX. However, preliminary computational studies show that this problem is hard to solve even for medium size instances. This motivates us to develop a more efficient method based on Lagrangian relaxation.

Major difficulties for solving ($\mathcal{FLRP}$) come from two aspects: 1) the demand balance constraint (35) impedes us from developing an analytical algorithm similar to that of Snyder and Daskin [47]; 2) constraint (38) increases the complexity further by connecting the forward flow and the reverse flow. We could relax constraints (35) and (38) of ($\mathcal{FLRP}$) by introducing Lagrangian multipliers $\boldsymbol{\lambda} = (\lambda_j)$ and $\boldsymbol{\mu} = (\mu_{js})$ to obtain the following Lagrangian relaxation problem:

$$
\begin{aligned}
&(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu})\\
&\min \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu})\\
&= \sum_{i\in\mathcal{I}} f_i^{FF} X_i + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} c_{ij}^{FF} Y_{ij} + \sum_{j\in\mathcal{J}} \lambda_j \left(D_j - \sum_{i\in\mathcal{I}} Y_{ij} \right)\\
&\quad + \sum_{s\in\mathcal{S}} p_s \left[\sum_{k\in\mathcal{K}_s} f_k^{RF} Z_{ks} + \sum_{k\in\mathcal{K}_s}\sum_{j\in\mathcal{J}} \left(c_{jk}^{RF} + c_k^{CP} \right) W_{jks} + \sum_{j\in\mathcal{J}} c_j^{LD} V_{js} \right]\\
&\quad + \sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{J}} \mu_{js} \left(\sum_{i\in\mathcal{I}_s} Y_{ij} - \sum_{k\in\mathcal{K}_s} W_{jks} - V_{js} \right)\\
&= \sum_{i\in\mathcal{I}} f_i^{FF} X_i + \sum_{j\in\mathcal{J}} \left[\sum_{i\in\mathcal{I}} (c_{ij}^{FF} - \lambda_j) + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \mu_{js} \right] Y_{ij} + \sum_{j\in\mathcal{J}} \lambda_j D_j\\
&\quad + \sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}_s} p_s f_k^{RF} Z_{ks} + \sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}_s}\sum_{j\in\mathcal{J}} \left[p_s \left(c_{jk}^{RF} + c_k^{CP} \right) - \mu_{js} \right] W_{jks}\\
&\quad + \sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{J}} \left(p_s c_j^{LD} - \mu_{js} \right) V_{js} \qquad (43)
\end{aligned}
$$

s.t. (36), (37), (41), (42)

$$0 \le Y_{ij} \le D_j \qquad \forall i \in \mathcal{I}, j \in \mathcal{J} \qquad (44)$$

$$0 \le W_{jks} \le D_j \qquad \forall j \in \mathcal{J}, s \in \mathcal{S}, k \in \mathcal{K}_s \qquad (45)$$

$$0 \le V_{js} \le D_j \qquad \forall j \in \mathcal{J}, s \in \mathcal{S}, k \in \mathcal{K}_s. \qquad (46)$$

Constraint (44) is modified from (39) to confine that the product quantity delivered to retailer j is no more than the retailer's demand. Similarly, constraints (45) and (46) are modified from constraint (40) to confine local disposal and central processing quantities.

4.4.1 Lower Bound

Relaxing constraint (38) breaks up the connection between forward and reverse flows, therefore we can separate ($\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu}$) into two optimization problems, i.e., ($\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{F}$) for forward flow and ($\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{R}$) for reverse flow, which can be solved separately. The optimization problem for the forward flow is defined as:

$$
\begin{aligned}
&(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{F}) \\
&\min \quad \sum_{i\in\mathcal{I}} f_i^{FF} X_i + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \left(c_{ij}^{FF} - \lambda_j\right) Y_{ij} + \sum_{j\in\mathcal{J}} \lambda_j D_j + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\sum_{j\in\mathcal{J}} \mu_{js} Y_{ij} \qquad (47)\\
&\text{s.t.} \quad (36), (41), (44).
\end{aligned}
$$

We use binary parameter δ_{is} equals one to denote the case plant i has triggered recall event in scenario s, and zero for otherwise. Therefore objective function of $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{F})$ is equivalent to $\sum_{i\in\mathcal{I}} f_i^{FF} X_i + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \left(c_{ij}^{FF} - \lambda_j + \sum_{s\in\mathcal{S}} \mu_{js}\delta_{is}\right) Y_{ij} + \sum_{j\in\mathcal{J}} \lambda_j D_j$.

Note that $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{F})$ can be solved by an efficient algorithm. We will open a plant at site i if and only if this decision decrease objective value. Objective value changes due to opening a plant at i (denoted by ϕ_i) can be determined by solving the following optimization problem:

$$
\begin{aligned}
\phi_i = \quad \min \quad & f_i^{FF} + \sum_{j\in\mathcal{J}} \left(c_{ij}^{FF} - \lambda_j + \sum_{s\in\mathcal{S}} \mu_{js}\delta_{is} \right) Y_{ij} \\
\text{s.t.} \quad & \sum_{j\in\mathcal{J}} Y_{ij} \le M_i^{FF}, \\
& (44).
\end{aligned}
$$

The calculation of ϕ_i depends on whether plant i has capacity limits.

In uncapacitated case $\left(M_i^{FF} = \infty\right)$, we have:

$$
Y_{ij}^* = \begin{cases} D_j, & c_{ij}^{FF} - \lambda_j + \sum_{s\in\mathcal{S}} \mu_{js}\delta_{is} < 0 \\ 0, & \text{otherwise.} \end{cases}
$$

In capacitated case $\left(M_i^{FF} < \infty\right)$, the problem is a continuous knapsack problem and we can find the optimal solution greedily as the method presented in [12]. For plant i, sort customer j in increasing order of $\left(c_{ij}^{FF} - \lambda_j + \sum_{s\in\mathcal{S}} \mu_{js}\delta_{is} \right)$. Let j' be the new ranking, and $\mathcal{J}^-$ be the set of customers with negative coefficients. Then:

$$
Y_{ij'}^* = \begin{cases} \min\left\{ D_{j'}, M_i^{FF} - \sum_{m=1}^{j'-1} D_m \right\}, & j' \in \mathcal{J}^- \\ 0, & \text{otherwise.} \end{cases}
$$

Next, the optimization problem for the reverse flow is defined as:

$$(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{R})$$
$$\min \sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}_s} p_s f_k^{RF} Z_{ks} + \sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}_s}\sum_{j\in\mathcal{J}} \widehat{c}_{jks} W_{jks} + \sum_{j\in\mathcal{J}}\sum_{s\in\mathcal{S}} \left(p_s c_j^{LD} - \mu_{js}\right) V_{js} \quad (48)$$
$$\text{s.t. } (37), (42), (45), (46),$$

where $\widehat{c}_{jks} = p_s\left(c_{jk}^{RF} + c_k^{CP}\right) - \mu_{js}$ for all $j \in \mathcal{J},\ s \in \mathcal{S},\ k \in \mathcal{K}_s$. Clearly, $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{R})$ is separable for $s \in \mathcal{S},\ k \in \mathcal{K}_s$. Also notice that local disposal decisions $\boldsymbol{V} = (V_{js})$ can be decided independently of $\boldsymbol{Z} = (Z_{ks})$ and $\boldsymbol{W} = (W_{jks})$ as follows:

$$V_{js}^* = \begin{cases} D_j, & p_s c_j^{LD} - \mu_{js} < 0 \\ 0, & \text{otherwise.} \end{cases}$$

Given a scenario s, if we locate a facility at candidate site $k \in \mathcal{K}_s$, changes of objective value (denoted by ψ_{ks}) are determined by the following optimization problem:

$$\begin{aligned} \psi_{ks} = \quad \min \quad & p_s f_k^{RF} + \sum_{j\in J} \widehat{c}_{jks} W_{jks} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{J}} W_{jks} \le M_k^{RF}. \end{aligned} \quad (49)$$

To solve $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu} - \mathcal{R})$, we open a facility k under scenario s if and only if $\psi_{ks} < 0$. The solution of (49) depends on whether plants have capacity limits.

In uncapacitated case ($M_k^{RF} = \infty$), we have:

$$W_{jks}^* = \begin{cases} D_j, & \widehat{c}_{jks} < 0 \\ 0, & \text{otherwise.} \end{cases}$$

In capacitated case ($M_k^{RF} < \infty$), sort customer j in increasing order of $\widehat{c}_{jks}$. Let j' be the new ranking, and $\mathcal{J}^-$ be the set of customers with negative $\widehat{c}_{jks}$. Then:

$$W_{jks}^* = \begin{cases} \min\left\{D_{j'}, M_k^{RF} - \sum_{m=1}^{j'-1} D_m\right\}, & j' \in \mathcal{J}^- \\ 0, & \text{otherwise.} \end{cases}$$

4.4.2 Upper Bound

We solve $(\mathcal{FLRP})$ using location decisions obtained from $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu})$ for an upper bound. With location decisions, $(\mathcal{FLRP})$ is reduced to a linear programming problem with allocation decisions only. However, in cases of capacitated plants, initial location decisions may be infeasible to obtain an upper bound problem due to insufficient plants' capacity. To restore feasibility, we design a greedy heuristic procedure by assigning unfixed closed plants to open in increasing order of contribution $\phi(i)$ (shown in the following algorithm).

Algorithm 1 Restore feasibility.

if $\rho <= 0$ **then**
Stop{open facilities have sufficient capacity}
else
if $\rho > \sum_{i \in \mathcal{I}'} M_i^{FF}$ **then**
Stop{No feasible upper bound can be found}
else
Sort $i \in \mathcal{I}'$ by the increasing order of ϕ_i, denoted by i^+
while $\rho > 0$ **do**
$X_{i^+}^* = 1$
$\rho = \rho - M_{i^+}^{FF}$
$i^+ = i^+ + 1$
end while
end if
end if

Let $\mathcal{I}'$ be the set of plants that are closed in the optimal solution of relaxed problem $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu})$ but not fixed to closure (e.g. in later mentioned branching process of branch and bound), and denote demands not fulfilled by total capacities of open plants with

$$\rho = \sum_{j \in \mathcal{J}} D_j - \sum_{i \in \mathcal{I}'} M_i^{FF}.$$

Note that we don't need to adjust location solutions of reverse flows, because local disposal has unlimited capacity and any returned products beyond capacities of central reprocessing can be handled with local disposal method.

Let $\hat{\mathcal{I}}$ be the set of all open plants and set $\hat{\mathcal{K}}_s$ denotes all open facilities in scenario s in the optimal solution of relaxed problem $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu})$. The problem to obtain upper bound is formulated as follows:

$$\min \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} c_{ij}^{FF} Y_{ij} + \sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{J}} p_s \left[\sum_{k\in\mathcal{K}_s} (c_{jk}^{RF} + c_k^{CP}) W_{jks} + c_j^{LD} V_{js}\right] \tag{50}$$

$$\text{s.t. } (35), (38), (39), (40)$$

$$\sum_{j\in J} Y_{ij} \le M_i^{FF} X_i^* = \begin{cases} M_i^{FF}, & i \in \hat{\mathcal{I}} \\ 0, & i \in I - \hat{\mathcal{I}} \end{cases} \tag{51}$$

$$\sum_{j\in J} W_{jks} \le M_k^{RF} Z_{ks}^* = \begin{cases} M_k^{RF}, & k \in \hat{\mathcal{K}}_s, s \in \mathcal{S} \\ 0, & k \in \mathcal{K} - \hat{\mathcal{K}}_s, s \in \mathcal{S} \end{cases} \tag{52}$$

With constraint (38) maintaining flow balance at different scenarios, above problem cannot be categorized as a transportation problem. With Theorem 7 for Schemes (a) and (b) where facilities' capacities are uncapacitated, costs incurred in stage two can migrate to stage one and constraint (38) is omitted.

In Schemes (a) and (b) with uncapacitated facilities, returned products of affected customer are fully processed in one recall mode (either local disposal or central processing). Optimal recall mode is selected based on comparative economic attractiveness. Let set $\mathcal{I}_{s+} = \cup_{s\in\mathcal{S}} \mathcal{I}_s$ denote plants that fail in at least one scenario. Let the unit cost for reverse flow at each customer j in scenario s in the upper bound solution be $\tilde{c}_{js}$, where $\tilde{c}_{js} = \min\left\{\min_{k\in\hat{\mathcal{K}}_s}\left\{c_{jk}^{RF} + c_k^{CP}\right\}, c_j^{LD}\right\}$.

Allocation decisions in reverse flows can be calculated directly:

$$W_{jks}^* = \begin{cases} \sum_{i\in\hat{\mathcal{I}}\cap\mathcal{I}_{s+}} Y_{ij}, & c_{jk}^{RF} + c_k^{CP} = \tilde{c}_{js} \\ 0, & \text{otherwise} \end{cases} \qquad V_{js}^* = 0;$$

OR

$$V_{js}^* = \begin{cases} \sum_{i\in\hat{\mathcal{I}}\cap\mathcal{I}_{s+}} Y_{ij}, & c_j^{LD} = \tilde{c}_{js} \\ 0, & \text{otherwise} \end{cases} \qquad W_{jks}^* = 0$$

Then the transportation problem can be transformed into:

$$\min_{(35),(39),(51)} \sum_{i\in\hat{\mathcal{I}}}\sum_{j\in\mathcal{J}} c_{ij}^{FF} Y_{ij} + \sum_{s\in\mathcal{S}}\sum_{j\in J} p_s \tilde{c}_{js} \sum_{i\in\hat{\mathcal{I}}} (\delta_{is} Y_{ij})$$
$$= \sum_{i\in\hat{\mathcal{I}}}\sum_{j\in J} \left(c_{ij}^{FF} + \sum_{s\in\mathcal{S}} p_s \tilde{c}_{js} \delta_{is} \right) Y_{ij}. \tag{53}$$

In Scheme (a) where manufacturing plants are uncapacitated, there is no demand splitting in both forward and reverse flows according to Theorem 7. Thus allocation decisions are either demand D_j or 0. We can calculate the optimal upper bound given solutions of the Lagrangian relaxation problem $(\mathcal{FLRP} - \mathcal{LR}_{\lambda,\mu})$ with the following algorithm.

Algorithm 2 Find upper bound in Scheme (a).

Define set I_1 for plants that are open and disrupted in at least one scenario, i.e., $I_1 = \hat{\mathcal{I}} \cap \mathcal{I}_{s^+}$.
while $j \in \mathcal{J}$ **do**
$\alpha_j = \min_{i\in\hat{\mathcal{I}}-I_1} \left\{c_{ij}^{FF}\right\}$, $\beta_{ij} = c_{ij}^{FF} + \sum_{s\in\mathcal{S}} p_s \tilde{c}_{js}\delta_{is}$
$\beta_{ij}^* = \min_{i\in I_1} \{\beta_{ij}\}$
if $\alpha_j \le \beta_j$ **then**
if $c_{i_j^* j}^{FF} = \alpha_j$ **then**
$Y_{i_j^* j}^* = D_j$
end if
else if $\beta_{i_j^* j} = \beta_{ij}^*$ **then**
$Y_{i_j^* j}^* = D_j$
Define set $S_{i_j^* j}$ for scenarios in which the chosen plant i_j^* fails
if $c_j^{LD} = \tilde{c}_{js}$ where $s \in S_{i_j^* j}$ **then**
$V_{js}^* = D_j$
else if $c_{k_j^*}^{CP} + c_{k_j^*}^{RF} = \tilde{c}_{js}$ where $s \in S_{i_j^* j}$ **then**
$W_{i_j^* j}^* = D_j$
end if
end if
$j = j + 1$
end while

In Scheme (b), problem (53) fit the classic form of transportation problem and can be solved by network simplex method. We refer to the work of [10] that provides a good guidance for network simplex method.

We close plants and facilities that serve no customers in the found upper bound solution.

In cases Scheme (c) and (d), we use Cplex LP solver.

4.4.3 Lagrangian Multipliers

Each vector pair of $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ forms a lower bound $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ to the optimal solution of $(\mathcal{FLRP})$. To obtain the optimal solution, we need to solve

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

We use the subgradient method to update the Lagrangian multipliers as in [16]. In the nth iteration of Lagrangian relaxation algorithm, denote the lower bound with $\mathcal{L}^n$, the best upper bound found so far with BUB, and the Lagrangian multipliers for the next iteration with λ^{n+1}, μ^{n+1} where:

$$\lambda_j^{n+1} \leftarrow \lambda_j^n + t^n \left(D_j - \sum_{i \in \mathcal{I}} Y_{ij} \right), \ \ \mu_{js}^{n+1} \leftarrow \mu_{js}^n + t^n \left(\sum_{i \in \mathcal{I}_s} Y_{ij} - \sum_{k \in \mathcal{K}} W_{jks} - V_{js} \right)$$

The step size is determined by

$$t^n = \frac{\beta^n (BUB - \mathcal{L}^n)}{\sum_{j \in \mathcal{J}} \left(D_j - \sum_{i \in \mathcal{I}} Y_{ij} \right)^2 + \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{J}} \left(\sum_{i \in \mathcal{I}_s} Y_{ij} - \sum_{k \in \mathcal{K}_s} W_{jks} - V_{js} \right)^2}$$

Note that β^n is a predetermined constant for the nth iteration, whose value will be halved if three consecutive iterations fail to make improvements.

The process of closing the gap between upper bound and lower bound is terminated if any of the following three criteria is satisfied:

- $\dfrac{BUB - \mathcal{L}^n}{\mathcal{L}^n} < \varepsilon$ where ε is a predetermined error tolerance
- $\beta^n < \beta^{min}$, where β^{min} is the minimal step size allowed
- $n > n^{max}$, where n^{max} is the maximal step number allowed.

4.4.4 Branch and Bound

We incorporate the Lagrangian relaxation into a branch and bound algorithm to ensure the optimality gap is closed at ε-level. Since plants location decisions are more impactive than facility location decisions, we branch on the former (i.e. X_i) only. At each node, branching plant selected is the unfixed open location with greatest contribution $\phi(i)$. The variable X_j is forced to be fixed value of zero and then one. Width-first search manner is applied in branching. A branch is fathomed if the lower bound of the parent node is greater than the best found upper bound. The tree is fathomed if obtained lower bound is within ε times the best found upper bound, for which the latter is the ε-optimal solution for the original problem. In each node, final Lagrangian multipliers are inherited to its children nodes and used as initial multipliers.

4.5 Computational Results

In the first experiment, we compare total cost and computation time of three models. The first model locates both manufacturing plants and recall processing facilities, and allocates customer demands all in the first stage. In the second stage, the first model allocates returned products given possible recall scenarios in located plants. We use RNM (recovery network design model) as studied from Fleischmann et al. [17] to represent the first model. The second model is our model ($\mathcal{FLRP}$), which makes location-allocation decisions for plants and customers in the first stage, and locates facilities and allocates returned products in the second stage, considering the availability of reprocessing centers in various recall scenarios. The third model designs the most cost-effective network in a non-fail situation and takes the best available reprocessing centers when failure scenarios happen. We use reactive facility location model (RFL) to represent the third model setting.

The first model RNM ignores the possibility that some facilities may not be available in some scenarios, and makes decisions of locating both plants and facilities in the first stage. On the contrary the neglected uncertainty is considered in ($\mathcal{FLRP}$), and we expect numerical results show the benefits of our consideration.

Our numerical example in Table 34 shows that the forward flow network of the first two models may not necessarily be the same. The second model shows better total cost (1%-5% less) and but a lot more complex to solve (using about five times more calculation time). Integration models (i.e. the first two models) performs better in total costs than reactive model (the third model). Location decisions are mostly different between integration models and reactive model.

In this section we use numerical experiments to test the necessity and significance of product recall modeling as well as the impacts of parameters. The model is coded with General Algebraic Modeling System (GAMS) language and tested on a GAMS server which runs on Intel Xeon dual CPU 2.00GHz 2.00 GHz processor with 8.0 GB RAM under Windows 64-bite operating system. We consider three variations in both environment settings and modeling settings in our experiments. Modeling settings describe the strategy of network design that a manager adopts, and environment settings simulate the true events for which the adopted strategic locations are used to satisfy demands and manage reverse flows. Depending on how many plants issue recalls in each second stage scenario, we consider three variations, namely no-recall, single-recall and dual-recall. No-recall does not consider the possibility of recall and

	RNM	$\mathcal{FLRP}$	RFL	RNM	$\mathcal{FLRP}$	RFL	RNM	$\mathcal{FLRP}$	RFL
Cost($000s)	6.92	6.36	7.54	1.68	1.62	1.81	2.85	2.78	2.96
Time(sec.)	0.61	0.24	0.50	0.92	0.98	0.70	6.64	24.58	2.56
Model size	$\mid \mathcal{I} \mid$	4		$\mid \mathcal{I} \mid$	8		$\mid \mathcal{I} \mid$	16	
	$\mid \mathcal{K} \mid$	3		$\mid \mathcal{K} \mid$	6		$\mid \mathcal{K} \mid$	12	
	$\mid \mathcal{S} \mid$	10		$\mid \mathcal{S} \mid$	36		$\mid \mathcal{S} \mid$	136	
	$\mid \mathcal{J} \mid$	4		$\mid \mathcal{J} \mid$	10		$\mid \mathcal{J} \mid$	20	
Cost(/$\mathcal{FLRP}$)	1.09	1.00	1.19	1.04	1.00	1.12	1.03	1.00	107
Time(/$\mathcal{FLRP}$)	2.56	1.00	2.11	0.94	1.00	0.71	0.27	1.00	0.10

Table 34. Compare costs and computation time of three models

its impact on the network. Single-recall assumes exactly one plant issues a recall in each scenario. Dual-recall considers the occurrence of recall and assumes at most two plants issues recalls per scenario, i.e., situations of exactly one plant and exactly two plants issuing recalls. Table 35 illustrates the relationship of no-recall, single-recall and dual-recall in environment and modeling settings.

	Number of plants issue recall per scenario	Environment notation	Modeling notation
No-recall	0	$S0$	$M0$
Single-recall	1	$S1$	$M1$
Dual-recall	1 OR 2	$S2$	$M2$

Table 35. Relationship of no-recall, single-recall and dual-recall in environment and modeling settings.

Combination of three modeling settings and three environment settings gives us nine optimal total costs for a specific set of parameters. For instance, the expected total cost with (M0, S1) reflects the expenditure of both forward and reverse flows when there are recalls in reality, in fact one recall per scenario, while the manager assumes no recall in strategic network designs. Table 36 shows one set of typical experiment results, i.e., expected total costs of three model settings under three environments.

Notice diagonal numbers are the costs when model settings match environment settings, which we can call perfect information decisions (PIDs). Comparing with PIDs, total costs naturally increase when failing probability is either over-estimated (e.g., model considers at most two facilities have recalls per scenario while only one facility has recall in reality), or under-estimated (e.g., model considers no recall while in reality exactly one plant issues recall per scenario). We use regrets to evaluate effects of wrong estimation on total costs.

In the search of optimal modeling setting in three scenarios, we use Savage's minimax regrets method, a widely applied approach that conservatively chooses the option of least worst case costs [44]. We define regrets in two ways: relative measure and absolute measure. Relative measure uses the percent cost increase when modeling settings mismatch environment settings compared to PIDs. Absolute measure uses value increase instead of percent increase. Table 37 shows results of applying relative and absolute measures of data in Table 36.

The reason that we adopt two measures to describe regrets is that PIDs under different environment settings can differ in significant ways. Therefore the best modeling option reflected by relative measure could be different from the choice of ab-

	$S0$	$S1$	$S2$
$M0$	50,389.44	142,210.15	186,006.79
$M1$	53,657.82	138,413.01	179,521.94
$M2$	53,657.82	138,413.01	179,521.94

Table 36. Numerical example of optimal costs for three modeling settings under three environment settings.

solute measure because the PID reference may change. Data in Table 37, for example, shows that, with relative measure, no-recall modeling (M0) is superior to single-recall (M1) and dual-recall modeling (M2) because, with M0, the worst cost increase is 3.61% more than PIDs compared to 6.49% increase for both M1 and M2. However, with absolute measure, the opposite is true (i.e., with M0, the worst cost increase is $6485 compared to $3268 for both M1 and M2). To avoid bias generated by choosing only one measure, we use both measures.

	$S0$	$S1$	$S2$	$S0$	$S1$	$S2$
$M0$	1	102.74%	103.61%	-	3,797.14	6,484.84
$M1$	106.49%	1	1	3,268.38	-	
$M2$	106.49%	1	1	3,268.38		-

Table 37. Apply relative measure (left) and absolute measure (right).

4.5.1 Parameter Settings

Our experiments test four sets of parameters settings: recall probability, capacity abundance, costly reverse flows and facility availability.

Based on the literature, the probability of a first stage plant issuing a recall in the second stage, recall probability, is within the range of [0.01, 0.08]. Due to physical, social and financial differences, recall probabilities of candidate plant locations may be quite different from each other. We use a uniform distribution with the range [0.01, 0.08] to generate the probability of each plant incurring recalls.

Capacity abundance is how abundant one plant's capacity is compared to the total demands, i.e., the proportion of total demands that one plant can satisfy with its capacity. Four degrees of capacity abundance are considered as in Table 38.

Costly reverse flows measure how expensive reverse flows are compare to forward flows. Three ratios are considered as in Table 39.

Facility availability measures the proportion of available facilities to open in the second stage compared to total number of possibly usable facilities. Four degrees of facility availability are considered as in Table 40.

4.5.2 Results

4.5.2.1 Impacts of Facility Availability

In order to find the impacts of facility availability on the choice of best modeling settings, we test four different levels of facility availability. With each level, we

Capacity abundance	Notation	Proportion of total demands that a facility can satisfy
Tight	Cap1	25%
Fair	Cap2	50%
Medium	Cap3	75%
Plenty	Cap4	100%

Table 38. Capacity abundance settings.

Costly reverse flows	Notation	Ratio of unit cost in reverse flows compare to that of forward flows
Inexpensive	costRF1	10
Medium	costRF2	50
Expensive	costRF3	100

Table 39. Costly reverse flows settings.

Facility availability	Notation	Percentage of facilities available to open in the second stage
Scarce	prKs1	20%
Somewhat	prKs2	40%
Adequate	prKs3	60%
Sufficient	prKs4	80%

Table 40. Facility availability settings.

experiment with four variations of capacity abundance and three variations of costly reverse flows, applying randomly generated recall probabilities following a uniform distribution. Results are shown in Table 41:

		prKs1	prKs2	prKs3	prKs4	Average
Relative	$M0$	12.20%	10.70%	11.52%	12.15%	11.64%
measure	$M1$	45.14%	44.95%	46.23%	45.46%	45.45%
	$M2$	42.66%	44.35%	42.25%	42.39%	42.91%
Absolute	$M0$	4.08%	3.86%	4.62%	4.52%	4.27%
measure	$M1$	48.62%	49.20%	49.86%	49.46%	49.29%
	$M2$	47.30%	46.94%	45.52%	46.02%	46.44%

Table 41. Proportion of optimal model setting under various facility availabilities.

When interpreting this table, note that the table columns for both relative and absolute measures sum to one. Each table entry represents the proportion of the time that the respective modeling setting is optimal. For example, using the relative measure, when facility availability is 20% (pKs1), the proportion of M1 being the optimal model is 45.14%. We see that facility availability in the second stage does not significantly impact the choice of the best model setting using either relative or absolute measures (i.e., the values in each row do not deviate much from the row average). For example, in the relative measure M1 row, proportions range from 44.95% to 46.23% with an average of 45.45%. Our prior is that decreasing facility availability could increase the proportion of modeling with recall considerations as optimal settings because higher facility availability indicates lower uncertainty in managing reverse flows and thus lower expected costs. Our results, on the contrary, show the optimal model setting is indifferent to facility availability. One explanation might be that our aggregation of data nullified the impact, or the impact is overshadowed by other more influential factors such as capacity abundance.

In all four cases, models considering recalls ($M1UM2$) performs significantly better than models without ($M0$) (e.g., for both relative and absolute measures, the M1 and M2 row values are much larger than the M0 ones). On average, with relative measure optimal models considering recalls, M1 and M2 account for 88.36% (45.45+42.91) of optimal model settings. With absolute measure, optimal models considering recalls account for 95.73% (49.29+46.44). This dominance proves that considering potential product recalls reduces total costs in the long run. Disregarding potential recalls could lead to selection of plant locations that initially seem to minimize costs, but that in hindsight are risky candidate sites with high expected costs to handle possible recalls. Our results with both relative and absolute measures support the assumption that designing with recall considerations minimizes the worst-case regrets.

The results also suggest that it is not obviously better to consider dual recall over single recall modeling. On average, of optimal models with recall consideration, using relative measure, single-recall models (M1) have proportions of 45.45% compared to 42.91% for dual-recall models; using absolute measure, single-recall models have 49.29% versus 46.44% for dual-recall models.

Dual recalls consider the following cases: a) two plants incurring recalls at the same, and b) only one plant incurring a recall at a time. Since plants incur recalls independently with a small probability, between 1% and 8%, the chances of two recalls happening at the same time appears to be too small to impact the network design in any noticeable scale. However, dual recalls modeling requires much more computation resources compared to single recalls since the scenario size increases exponentially. Balancing the above considerations, single recall modeling sufficiently serves our purpose of planning for potential recall risks and rationalizing computation power.

4.5.2.2 Impact of Capacity Abundance

Risks of not considering recall in network design decrease when plants' capacity abundance increases. The reason may be that insufficient plants' capacity leads to various recall scenarios and thus increases the expected costs of managing recalls. This suggests abundant plants' capacity allows more space for risk control for managers in designing networks without recall concerns. Results concerning capacity abundance are shown in Table 42. When capacity is tight (Cap1), it is almost never good to ignore the possibility of recalls (i.e., for relative measure, the M0 table entry is 2% and for absolute measure, the M0 table entry is 0%).

		Cap1	Cap2	Cap3	Cap4
Relative	$M0$	2.00%	5.48%	12.33%	41.05%
measure	$M1$	50.22%	47.95%	44.57%	32.31%
	$M2$	47.78%	46.57%	43.10%	26.64%
Absolute	$M0$	0.00%	5.00%	0.48%	20.71%
measure	$M1$	50.76%	47.48%	50.00%	45.33%
	$M2$	49.24%	47.52%	49.52%	33.95%

Table 42. Proportion of optimal model setting under various capacity availability.

The only time that it might be acceptable for decision makers to ignore recall costs is when plant capacity is large compared to demand (i.e., when there is plenty of capacity (Cap4), the M0 table entry is 41.05% for relative measure and the M0 table entry is 20.71% for absolute measure). However, if the business is growing, current excess capacity will ultimately disappear. Therefore planning with potential recall serves the long term goal of building reliable and cost-effective networks.

4.5.2.3 Impact of Costly Reverse Flows

Results concerning capacity abundance are shown in Table 43. Dominance of $M1andM2$ may be attributable to the distinguishable recall probabilities of candidate plant locations. To optimize the network design for first stage only, $M0$ may choose to open plants at relatively lower expense despite of their high chance of issuing recalls in the second stage, which results in premium payments when recalls do occur.

		cRF1	cRF2	cRF3
Relative measure	$M0$	7.36%	15.46%	14.67%
	$M1$	49.17%	43.26%	42.81%
	$M2$	43.47%	41.29%	42.52%
Absolute measure	$M0$	2.86%	4.58%	5.57%
	$M1$	51.70%	48.01%	48.27%
	$M2$	45.44%	47.41%	46.16%

Table 43. Proportion of optimal model setting under various costly degrees of reverse flows.

4.5.2.4 Cost Increase as a Result of Neglecting Recalls in Network Design

We apply the same set of recall probability and fix facility availability at 70%. Numerical results show the dominance of $M1andM2$ over $M0$ with minimax regrets similar to the results of Experiment 1. Results concerning capacity abundance are shown in Table 44.

We also notice $M1$ and $M2$ have very close or equal total costs in various capacity abundance and costly reverse flow settings. Both perform dominantly better than $M0$ when considering recalls.

We show the impact of capacity availability/costly reverse flow on the cost of $M0$ compared to $M1$ and $M2$ with relative measure, which is the cost of neglecting recalls in network design.

Max regret		Cap1 tight	Cap2 fair	Cap3 medium	Cap4 plentiful
costRF1	Inexpensive	70%	50%	40%	40%
costRF2	Medium	90%	70%	60%	60%
costRF3	Expensive	100%	80%	70%	70%

Table 44. Max regrets for choosing $M0$ with relative measures.

The costs of overlooking potential recalls vary largely from our randomly generated data sets, which indicates not only considering recalls in initial designs is necessary but also accurately predicting product recall probability can be crucial to effectively design the network.

4.6 Conclusions

This chapter addresses the supply chain safety control issue by designing the supply chain network to incorporate the negative effects of product recalls. This work focuses on managing the reverse flow (recalled products) in a cost-efficient manner. We study the efficient way of managing random and rare major product recalls and consider how to quickly build a reprocessing network to dispose recalled products. We study the location-allocation problem with random occurrence of product recalls and treat the recall incidences as disruptions to the supply chain. Three features distinguish our problem from other location-allocation problems. Firstly, facility location and allocation decisions occur in two stages. Secondly, the second stage location-allocation happens under uncertainty. Thirdly, reverse logistic flows exist in the second stage.

We design a two-stage stochastic mixed integer programming model, in which we locate the manufacturing plants in the first stage and the reprocessing/disposal facilities in the second stage. We adopt a scenario-based approach to describe the uncertainty of major recall events that may happen in manufacturing plants as well as of availability of reprocessing facilities. Given the complexity induced by our nested facility location problem, we devise an algorithm based on Lagrangian relaxation to solve the uncapacitated case.

The existing literature treats reverse flow on day-to-day basis. This modeling approach does not serve major product recalls well. We fill the gap by designing an optimal network that can accommodate product returns in the context of major product recalls.

We compare total cost and computation time in the search of optimal modeling setting in three scenarios based on the minimax regrets method using both relative and absolute measures. Our experiments test four sets of parameters settings: recall probability, capacity abundance, costly reverse flows and facility availability. We find that facility availability in the second stage does not significantly impact the choice of the best model setting. However, we find that designing with recall considerations minimizes worst-case regrets. Moreover, considering potential product recalls reduces total costs in the long run – disregarding potential recalls could lead to selection of plant locations that initially seem to minimize costs, but that in hindsight are risky candidate sites with high expected costs to handle possible recalls. Risks of not considering recall in network design decrease when plants' capacity abundance increases.

Chapter 5
Thesis Summary and Concluding Remarks

In this thesis, we examine three important topics of product recall management optimization with a particular focus on recall timing, recall risk control with quality investment, and post-recall damage mitigation with reprocessing center location. Real-world incidents of recent years such as the GM's decade-long delay of initiating the Cobalt recall, XL's discarding of tons of beef products to landfills, and Samsung's multi-billion-dollar settling of potentially explosive Galaxy Note 7s have demonstrated that poor recall management can result in great harm to the company, the customers, and the public at large. Our research has practical implications for improving product recall management using optimization tools.

The first essay investigates the "when to initiate the product recall" problem in a dynamic decision making process. Using our dynamic programming model that updates the estimation for product defect rate, we solve for a threshold curve that defines the maximum number of returned products at each stage beyond which initiating product recall is optimal. Our numerical experiments show that threshold curves are sensitive to costs of managing product returns and recalls. That is, if managing recalls is much more expensive than managing returns, then managers tend to "wait-and-see" and, thereby, take higher recall risks.

Our primary contribution to the literature is combining dynamic programming and the conjugate property of beta distributions and Bernoulli processes to model and solve the product recall timing optimization problem. Due to computational complexity, our proposed dynamic programming model faces the limitation of problem size. We adopt the simulation method using the structure of threshold curve to solve large-sized problems to balance solution accuracy and computation time.

The second essay explores "how to reduce product recall risks" using production quality investment. Our primary contribution is that we model and solve this optimization problem by extending the classic single-period stochastic newsvendor problem. Introducing a second decision variable of quality level, we capture its relationships with recall probability and production cost with monotone functions. Since the adopted function forms are very general, our model is neither convex nor concave. Our parametric studies, however, reveal some interesting features such as the observation that optimal ordering quantity and quality level have a conflicting relationship; that is, most parameter changes lead one to increase and the other to decrease. Another finding is that increased profitability discourages quality level, which can be explained by viewing quality level as a cost factor since production cost is a non-decreasing function of quality. We further extend the model of internal suppliers to multiple external suppliers. Our numerical experiments suggest external suppliers satisfy their demands independently and respond very little to the potential requirements of covering other suppliers' demand in the case of product recall.

The third essay seeks to answer "how to reduce recall cost" from the perspective of locating reprocessing centers. Our primary contribution is that we model and solve a two-stage stochastic mixed integer programming model that combines closed-loop network design and disruption management. We adopt a scenario-based approach to describe the uncertainty of product recalls and devise a Lagrangian-relaxation-based algorithm to solve the uncapacitated case. Our numerical experiments show that designing networks with product recall concerns reduces worst-case regrets. Furthermore, considering potential recalls in the location design reduces total expected costs in the long run, given relative tight capacity abundance, inexpensive to expensive reverse flow cost, and scarce to sufficient facility availability.

The research in this thesis can be extended in several directions. To address large-sized recall timing optimization in the dynamic programming model, we can adopt the neurodynamic programming method studied by Bertsekas and Tsitsiklis [7] to find an approximate function form for the model objective. To extend our efforts for reducing recall risks with quality control, we can devise a random price-dependent demand function—as summarized by N. Petruzzi and M. Dada [33]—instead of assuming both demand and price are exogeneous for higher cross-functional effectiveness. Our closed loop network design model requires accurate estimation of product recall probability; we can extend this model with accuracy estimation of predicting product recall probability as studied by M. Lim et al. [23].

References

[1] R. Akkerman, P. Farahani, and M. Grunow. Quality, safety and sustainability in food distribution: A review of quantitative operations management approaches and challenges. *OR Spectrum*, 32(4):863–904, 2010.

[2] K. J. Arrow, T. Harris, and J. Marschak. Optimal inventory policy. *Econometrica*, 19:250–272, 1951.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2013.

[4] K. A. Beauchamp and R. S. Littlefield. The Maple Leaf Foods recall: best practice interaction during a food-related crisis. *International Journal of Business Continuity and Risk Management*, 3(1):1–18, 2012.

[5] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.

[6] B. Berman. Planning for the inevitable product recall. *Business Horizons*, 42(2):69–78, 1999.

[7] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996.

[8] G. H. Chao, S. M. R. Iravani, and R. C. Savaskan. Quality improvement incentives and product recall cost sharing contracts. *Management Science*, 55(7):1122–1138, 2009.

[9] S. Charlebois, M. Von Massow, and W. Pinto. Food recalls and risk perception: An exploratory case of the XL Foods and the biggest food recall in Canadian history. *Journal of Food Products Marketing*, 21(1):27–43, 2015.

[10] V. Chvátal. *Linear programming*. Macmillan, 1983.

[11] M. Dada, N. C. Petruzzi, and L. B. Schwarz. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management*, 9(1):9–32, 2007.

[12] M. S. Daskin. *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons, 2011.

[13] A. F. Daughety and J. F. Reinganum. Product safety: liability, R&D, and signaling. *The American Economic Review*, pages 1187–1206, 1995.

[14] W. N. Davidson and D. L. Worrell. Research notes and communications: The effect of product recall announcements on shareholder wealth. *Strategic management journal*, 13(6):467–473, 1992.

[15] P. Desmarchelier, N. Fegan, N. Smale, and A. Small. Managing safety and

quality through the red meat chain. *Meat Science*, 77(1):28–35, 2007.

[16] M. L. Fisher. The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18, 1981.

[17] M. Fleischmann, P. Beullens, J. M. Bloemhof-Ruwaard, and L. N. Wassenhove. The impact of product recovery on logistics network design. *Production and Operations Management*, 10(2):156–173, 2001.

[18] N. H. Hooker, R. M. Nayga, and J. W. Siebert. The impact of HACCP on costs and product exit. *Journal of Agricultural and Applied Economics*, 34(01):165–174, 2002.

[19] M. Hora, H. Bapuji, and A. V. Roth. Safety hazard and time to recall: The role of recall strategy, product defect type, and supply chain player in the US toy industry. *Journal of Operations Management*, 29(7):766–777, 2011.

[20] G. Jarrell and S. Peltzman. The impact of product recalls on the wealth of sellers. *Journal of Political Economy*, 93(3):512–536, 1985.

[21] L. J. Krajewski, L. P. Ritzman, and M. K. Malhotra. *Operations management: Processes and supply chains*, volume 1. Pearson New York, 2013.

[22] D. H. Lee and M. Dong. A heuristic approach to logistics network design for end-of-lease computer products recovery. *Transportation Research Part E: Logistics and Transportation Review*, 44(3):455–474, 2008.

[23] M. Lim, A. Bassamboo, S. Chopra, and M. Daskin. Facility location decisions with random disruptions and imperfect estimation. *Manufacturing & Service Operations Management*, 15(2):239–249, 2013.

[24] O. Listeş. A generic stochastic model for supply-and-return network design. *Computers and Operations Research*, 34(2):417–442, 2007.

[25] Z. Lu and N. Bostel. A facility location model for logistics systems including reverse flows: The case of remanufacturing activities. *Computers and Operations Research*, 34(2):299–323, 2007.

[26] R. D. Maiorescu. Crisis management at general motors and toyota: An analysis of gender-specific communication and media coverage. *Public Relations Review*, 42(4):556–563, 2016.

[27] A. M. Marino. A model of product recalls with asymmetric information. *Journal of Regulatory Economics*, 12(3):245–265, 1997.

[28] T. L. Marsh, T. C. Schroeder, and J. Mintert. Impacts of meat product recalls on consumer demand in the usa. *Applied Economics*, 36(9):897–909, 2004.

[29] A. Marucheck, N. Greis, C. Mena, and L. Cai. Product safety and security in the global supply chain: Issues, challenges and research opportunities. *Journal*

of Operations Management, 29(7):707–720, 2011.

[30] H. Min. A bicriterion reverse distribution model for product recall. *Omega*, 17(5):483–490, 1989.

[31] H. Min and H. J. Ko. The dynamic design of a reverse logistics network from the perspective of third-party logistics service providers. *International Journal of Production Economics*, 113(1):176–192, 2008.

[32] M. Parlar. *Interactive Operations Research with Maple: Methods and Models.* Birkhäuser, Boston, 2000.

[33] N. C. Petruzzi and M. Dada. Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2):183–194, 1999.

[34] M. S. Pishvaee, M. Rabbani, and S. A. Torabi. A robust optimization approach to closed-loop supply chain network design under uncertainty. *Applied Mathematical Modelling*, 35(2):637–649, 2011.

[35] W. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* Wiley-Interscience, New Jersey, 2007.

[36] L. Qi. A continuous-review inventory model with random disruptions at the primary supplier. *European Journal of Operational Research*, 225(1):59–74, 2013.

[37] L. Qi, Z. J. M. Shen, and L. V. Snyder. The effect of supply disruptions on supply chain design decisions. *Transportation Science*, 44(2):274–289, 2010.

[38] M. Ramezani, M. Bashiri, and R. Tavakkoli-Moghaddam. A new multi-objective stochastic model for a forward/reverse logistic network design with responsiveness and quality level. *Applied Mathematical Modelling*, 37(1):328–344, 2013.

[39] S. Ross. *A First Course in Probability.* Macmillan, New York, 1976.

[40] A. P. Ruszczynski and A. Shapiro. *Stochastic Programming.* Elsevier, 2003.

[41] M. I. A. Salema, A. P. Barbosa-Povoa, and A. Q. Novais. A warehouse-based design model for reverse logistics. *Journal of the Operational Research Society*, 57(6):615–629, 2006.

[42] M. I. G. Salema, A. P. Barbosa-Povoa, and A. Q. Novais. An optimization model for the design of a capacitated multi-product reverse logistics network with uncertainty. *European Journal of Operational Research*, 179(3):1063–1077, 2007.

[43] V. Salin and N. H. Hooker. Stock market reaction to food recalls. *Review of Agricultural Economics*, 23(1):33–46, 2001.

[44] L. J. Savage. The theory of statistical decision. *Journal of the American Statistical association*, 46(253):55–67, 1951.

[45] R. C. Savaskan, S. Bhattacharya, and L. N. Van Wassenhove. Closed-Loop Supply Chain Models with Product Remanufacturing. *Management Science*, 50(2):239–252, 2004.

[46] A. D. Sezer and Ç. Haksöz. Optimal Decision Rules for Product Recalls. *Mathematics of Operations Research*, 37(3):399–418, 2012.

[47] L. V. Snyder and M. S. Daskin. Reliability models for facility location: The expected failure cost case. *Transportation Science*, 39(3):400–416, 2005.

[48] L. V. Snyder, M. P. Scaparra, M. S. Daskin, and R. L. Church. Planning for disruptions in supply chain networks. *Tutorials in Operations Research*, 2:234–257, 2006.

[49] S. Thirumalai and K. K. Sinha. Product recalls in the medical device industry: an empirical exploration of the sources and financial consequences. *Management Science*, 57(2):376–392, 2011.

[50] M. R. Thomsen and A. M. McKenzie. Market incentives for safe foods: an examination of shareholder losses from meat and poultry recalls. *American Journal of Agricultural Economics*, 83(3):526–538, 2001.

[51] A. R. Valukas. Report to board of directors of General Motors company regarding ignition switch recalls. *Jenner & Block, Tech. Rep*, 2014.

[52] L. Welling. A theory of voluntary recalls and product liability. *Southern Economic Journal*, pages 1092–1111, 1991.

[53] R. L. Winkler. *Introduction to Bayesian Inference and Decision.* Holt, Rinehart and Winston, New York, 1972.