

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays on Supply Chain Management with Model Uncertainty

Permalink

<https://escholarship.org/uc/item/75r7x7kw>

Author

Lu, Mengshi

Publication Date

2014

Peer reviewed|Thesis/dissertation

Essays on Supply Chain Management with Model Uncertainty

by

Mengshi Lu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering – Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Shen, Chair
Professor Andrew Lim
Professor Candace Yano
Professor Bin Yu

Spring 2014

Essays on Supply Chain Management with Model Uncertainty

Copyright 2014
by
Mengshi Lu

Abstract

Essays on Supply Chain Management with Model Uncertainty

by

Mengshi Lu

Doctor of Philosophy in Engineering – Industrial Engineering and Operations Research

University of California, Berkeley

Professor Zuo-Jun Shen, Chair

Traditional supply chain management models typically require complete model information, including structural relationships (e.g., how pricing decisions affect customer demand), probabilistic distributions, and parameters. However, in practice, the model information may be uncertain. My dissertation research seeks to address model uncertainty in supply chain management problems using data-driven and robust methods. Incomplete information typically comes in two forms, namely, historical data and partial information. When historical data are available, data-driven methods can be used to obtain decisions directly from data, instead of estimating the model information and then using these estimates to find the optimal solution. When partial information is available, robust methods consider all possible scenarios and make decisions to hedge against the worst-case scenario effectively, instead of making simplified assumptions that could lead to significant loss.

Chapter 1 provides an overview of model uncertainty in supply chain management, and discusses the limitations of the traditional methods. The main part of the dissertation is on the application of data-driven and robust methods to three widely-studied supply chain management problems with model uncertainty.

Chapter 2 studies the reliable facility location problem where the joint-distribution of facility disruptions is uncertain. For this problem, usually, only partial information in the form of marginal facility disruption probabilities is available. Most existing models require the assumption that the disruptions at different locations are independent of each other. However, in practice, correlated disruptions are widely observed. We present a model that allows disruptions to be correlated with an uncertain joint distribution, and apply distributionally-robust optimization to minimize the expected cost under the worst-case distribution with the given marginal disruption probabilities. The worst-case distribution has a practical interpretation, and its sparse structure allows us to solve the problem efficiently. We find that ignoring disruption correlation could lead to significant loss. The robust method can significantly reduce the regret from model misspecification. It outperforms the traditional approach even under very mild correlation. Most of the benefit of the robust model can be captured at a relatively small cost, which makes it easy to implement in practice.

Chapter 3 studies the pricing newsvendor problem where the structural relationship between pricing decisions and customer demand is unknown. Traditional methods for this problem require the selection of a parametric demand model and fitting the model using historical data, while model selection is usually a hard problem in itself. Furthermore, most of the existing literature on pricing requires certain conditions on the demand model, which may not be satisfied by the estimates from data. We present a data-driven approach based only on the historical observations and the basic domain knowledge. The conditional demand distribution is estimated using non-parametric quantile regression with shape constraints. The optimal pricing and inventory decisions are determined numerically using the estimated quantiles. Smoothing and kernelization methods are used to achieve regularization and enhance the performance of the approach. Additional domain knowledge, such as concavity of demand with respect to price, can also be easily incorporated into the approach. Numerical results show that the data-driven approach is able to find close-to-optimal solutions. Smoothing, kernelization, and the incorporation of additional domain knowledge can significantly improve the performance of the approach.

Chapter 4 studies inventory management for perishable products where a parameter of the demand distribution is unknown. The traditional separated estimation-optimization approach for this problem has been shown to be suboptimal. To address this issue, an integrated approach called operational statistics has been proposed. We study several important properties of operational statistics. We find that the operational statistics approach is consistent and guaranteed to outperform the traditional approach. We also show that the benefit of using operational statistics is larger when the demand variability is higher. We then generalize the operational statistics approach to the risk-averse newsvendor problem under the conditional value-at-risk (CVaR) criterion. Previous results in operational statistics can be generalized to maximize the expectation of conditional CVaR. In order to model risk-aversion to both the uncertainty in demand sampling and the uncertainty in future demand, we introduce a new criterion called the total CVaR, and find the optimal operational statistic for this new criterion.

To my parents

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Model Uncertainty in Supply Chain Management	1
2 Reliable Facility Location Design with Distributional Uncertainty	4
2.1 Introduction	4
2.2 Literature Review	6
2.3 Model and Formulation	9
2.4 Numerical Results	14
2.5 Extensions to Other Reliable Facility Location Problems	24
2.6 Summary and Future Directions	26
3 Joint Pricing-Inventory Management with Structural Uncertainty	28
3.1 Introduction	28
3.2 Literature Review	30
3.3 A Data-Driven Approach Based on Isotonic Quantile Regression	33
3.4 Smoothing, Concavity Constraint and Kernelization	40
3.5 Numerical Experiments	50
3.6 Summary and Future Directions	53
4 Inventory Management for Perishable Goods with Parameter Uncertainty	54
4.1 Introduction	54
4.2 Literature Review	55
4.3 Background on Operational Statistics	57
4.4 Properties of Operational Statistics	58
4.5 Operational Statistics under Risk Aversion	63
4.6 Summary and Future Directions	67
A Proofs	68

B Data and Supplementary Results	78
B.1 Severe Weather Hazard Data for the Supply Chain Design Example	78
B.2 Selected Reliable Facility Location Results Using the 88-Node Data Set	78
Bibliography	81

List of Figures

2.1	48-hour impact probability forecast of Hurricane Sandy at the time of its landfall (NOAA, 10-29-2012)	5
2.2	The optimal design under independent disruptions (Design I)	15
2.3	The optimal design under worst-case correlated disruptions (Design R)	16
2.4	Impact of important factors on expected regret	20
2.5	Expected cost under different degrees of disruption correlation	21
2.6	Benefit and cost of the robust design with different weights	22
2.7	Percentage of maximum benefit captured and maximum cost incurred by different weights	23
3.1	Examples of isotonic quantile regression estimates	37
3.2	Quantile path and breakpoints in parametric programming	38
3.3	Number of breakpoints in isotonic quantile regression	39
3.4	Estimated optimal inventory level, expected profit function, and optimal price	40
3.5	Quantile path and breakpoints using smoothing	42
3.6	Improved data-driven solutions using smoothing	42
3.7	Cross validation risk of different smoothing parameters	44
3.8	Quantile path and breakpoints with concavity constraint	45
3.9	Improved data-driven solutions with concavity constraint	46
3.10	Quantile path and breakpoints using kernelization	49
3.11	Improved data-driven solutions using kernelization	49
4.1	Comparison of operational statistics and separated estimation-optimization solutions (exponential distribution)	61
4.2	Comparison of the optimality gaps for the operational statistics and separated estimation-optimization approaches (exponential distribution)	61
4.3	Impact of demand coefficient of variation on the benefit of operational statistics (shifted exponential distribution)	62
4.4	Total CVaR achieved by operational statistics and separated estimation-optimization	66

List of Tables

2.1	Summary of literature on reliable facility location	7
2.2	Optimal location designs of the two models	16
2.3	Comparison of the performance of the two designs	16
2.4	Levels for important factors	18
2.5	Selected results for the 49-node data set	19
3.1	Summary of the numerical experiment results	51
3.2	Summary of the numerical experiment results with demand concavity	52
B.1	Estimated marginal disruption probability for the supply chain design example.	79
B.2	Selected results for the 88-node data set.	80

Acknowledgments

I am deeply indebted to my advisor, Professor Zuo-Jun “Max” Shen, for his guidance, encouragement, and support. He is not only a great research advisor, but also a great mentor. His advice on research and academic career has been invaluable to me. I would also like to thank my dissertation committee members, Professor Andrew Lim, Professor Candace Yano, and Professor Bin Yu. My research benefited greatly from their expertise. Their helpful comments and suggestions significantly improved the exposition of my dissertation. I am also grateful to Professor George Shanthikumar whose research in operations management under model uncertainty inspired me to focus my research in this area. As a PhD student, I spent one year conducting research at the Hewlett Packard Laboratories. I would like to thank Dr. Kemal Güler for being a great mentor and a cordial friend, and Professor Pin Ng for his guidance on quantile regression. I am also grateful to Dr. Fereydoon Safai for supporting my research at the Hewlett Packard Laboratories.

I would also like to thank the other IEOR faculty members, especially Professor Ying-Ju Chen, for his advice on my research, and Professor Rhonda Righter, Professor Lee Schruben, and Professor Jon Burgstone, for their help with my teaching. I am also very thankful to the IEOR staff, especially Mr. Michael Campbell, Ms. Anayancy Paz, and Mr. Jay Sparks. I had an amazing time while living in Berkeley. I would like to thank my dear friends, including Yong, Ye, Tianhu, Kai-Chuan, Yifen, Huaning, Renmin, Minzhi, Yangqing, Sizhu, and many others.

I am forever grateful to my parents for their unconditional support during my long endeavor as a graduate student. I am also grateful to my mother-in-law and my wife’s aunt, Tongyu, for traveling to the United States to help us during the most difficult times. Last but not least, my special thanks go to my wife, Qiuting, for always standing by my side during my ups and downs, and for bringing our daughter, Yihan, into this world. I am so fortunate to be accompanied by her in Berkeley and in the life-long journey ahead of us.

Chapter 1

Model Uncertainty in Supply Chain Management

Dealing with uncertainty effectively is one of the fundamental motivations for supply chain management research. Sources of uncertainty that have been widely studied include customer demand, production yield, fluctuations in the leadtimes, unreliable suppliers or facilities, etc. However, most of the traditional supply chain management models assume that the variable(s) and/or parameter(s) of interest, although uncertain, can be characterized probabilistically using the model information. There are typically three types of model information, namely, structural relationships, probabilistic distributions, and parameters.

To illustrate the three types of model information, consider the classical pricing news vendor problem as an example. In this problem, the retailer jointly determines the selling price and the order quantity to maximize the expected profit. The customer demand is uncertain and depends on the selling price. To characterize customer demand, we first need to know the structural relationship between price and demand. This is typically in the form of a demand function. For example, we can assume the demand, D , is a linear function of price, p , plus some random factor, ε , i.e.,

$$D(p, \varepsilon) = \alpha + \beta p + \varepsilon,$$

where α and β are the coefficients of the linear mean demand function. Then, given the price and a realization of the random factor, the demand is determined. Second, we need to know the probabilistic distribution of the random factor. For example, we can assume that the random factor is normally distributed, i.e.,

$$\varepsilon \sim \mathcal{N}(\mu, \sigma^2),$$

where μ and σ^2 are the mean and variance of the normal distribution, respectively. Under certain circumstances, there may be multiple interdependent random factors. In such a case, we also need to know their joint distribution. Finally, we need to know the parameters, including the coefficients, α and β , in the demand function, as well as the mean, μ , and variance, σ^2 , of the random factor. With the complete model information, the customer demand

is characterized probabilistically. We can then derive the expected profit and optimize the pricing and inventory decision.

In practice, some or even all of the model information may be unknown. In the presence of model uncertainty, decisions need to be made based on the information that is available. We consider two types of available information. The first type is historical data. For example, we may have price and demand observations from the previous periods. In this case, the traditional method is to estimate the model information from the data. The estimate is then used to substitute for the unknown model information in the optimal solution. The second type of limited information is partial information. For example, when there are multiple random factors with an unknown joint distribution, we may know the marginal distribution of each random factor. In this case, the traditional method is to make simplifying assumptions based on the available partial information. For example, a common practice when only the marginal distribution is known is to assume the random factors are independent. Then, the joint distribution can be constructed as the product of the marginal distributions.

There are several limitations of the traditional methods for dealing with model uncertainty. First, estimating the model information from data is not an easy task. There are usually many available probabilistic models. Selecting a good model usually requires training or technical support that most supply chain managers do not have. If the model is not selected properly, the estimates and the resulting decisions may be highly suboptimal. Second, the model selection and estimation process needs to be repeated for each new data set, which may entail an extensive amount of work. Third, classical supply chain management models typically require certain conditions on the model input. For example, in order to use classical results in pricing, we usually need the demand function to satisfy certain price elasticity conditions and the distribution of the random factor to have an increasing failure rate. These conditions may not be satisfied by the estimates from data. Finally, even when model information can be properly estimated, separation of estimation and optimization still results in suboptimality. When only partial information is available, traditional methods can also be problematic. Assumptions based on the available information can be hard to verify in practice. If an oversimplified model is used due to strong assumptions, the corresponding decisions can be highly suboptimal.

Data-driven and robust methods can be applied to effectively address model uncertainty. When historical data are available, data-driven methods derive decisions from data without using a specific parametric model or separating estimation and optimization. They do not require model selection and fitting by managers, or strong conditions on the model input, and thus can be directly applied in practice. When partial information is available, robust methods consider the worst-case among all possibilities based on the available information. Thus, it avoids making ungrounded assumptions, and can prevent massive loss due to model misspecification.

My dissertation consists of essays on the application of data-driven and robust methods for three widely studied supply chain management problems with model uncertainty. Chapter 2 is on the reliable facility location problem under facility disruption risk. For this problem, usually, only the partial information, i.e., the marginal facility disruption proba-

bility, is available, while the joint distribution for all the facilities is unknown. Using robust optimization, we avoid making the assumption that the disruptions are independent. We consider all the joint distributions with the given marginal disruption probability, and derive the worst-case distribution in closed form. We find that the robust method can significantly reduce the regret from model misspecification. It outperforms traditional approaches for this problem even under very mild correlation. Solving the robust model also requires much less computational effort than the traditional approach. Most of the benefit of the robust model can be captured at a relatively small cost, which makes it easy to implement in practice.

Chapter 3 is on the pricing newsvendor problem where the structural relationship between price and customer demand is unknown. We present a data-driven approach that does not require any parametric demand model. Instead, the approach is based only on historical observations and basic domain knowledge. The conditional demand distribution is estimated using non-parametric quantile regression with shape constraints. The optimal pricing and inventory decisions are determined numerically using the estimated quantiles. Smoothing and kernelization methods are used to achieve regularization and enhance the performance of the approach. Additional domain knowledge, such as demand concavity, can also be incorporated in the approach. Numerical results show that the data-driven approach is able to find close-to-optimal solutions. Smoothing, kernelization, and the incorporation of additional domain knowledge can significantly improve the performance of the approach.

Chapter 4 is on inventory management for perishable products where a parameter of the demand distribution is unknown. The traditional separated estimation-optimization approach for this problem is shown to be suboptimal. We study several important properties of an integrated approach called operational statistics. We find that the operational statistics approach is consistent and guaranteed to outperform the separated estimation-optimization approach. The benefit of using operational statistics increases as the demand variability increases. We then generalize the operational statistics approach to the risk-averse newsvendor problem under the conditional value-at-risk (CVaR) criterion. Previous results in operational statistics can be generalized to solve the problem of maximizing the expectation of conditional CVaR. In order to model risk-aversion to both demand sampling risk and future demand uncertainty risk, we introduce a new criterion called the total CVaR, and find the optimal operational statistic for this new criterion.

Chapter 2

Reliable Facility Location Design with Distributional Uncertainty

2.1 Introduction

In the recent years, supply chain disruptions have caused significant losses due to facility damage and production or service interruption. Designing reliable supply chains when facilities are subject to random disruptions has gained a lot of attention from industry and academia. For example, IBM has launched the Business Continuity and Resilience Service to help companies evaluate their disruption risk and improve their resilience using optimized planning and design. In operations research and management sciences, the reliable facility location problem has been extensively studied (e.g., Snyder and Daskin, 2005; Cui, Ouyang, and Shen, 2010; Lim et al., 2010). In this problem, the decision maker needs to design a supply chain network, where the facilities will be disrupted according to some probabilistic distribution. (It is assumed that the arcs of the network, i.e., transportation links between facilities and customers, are not disrupted.) Customers can only be served by available facilities. Unlike the classical facility location models, customer assignment, and thus, the transportation cost, in the reliable facility location problem is random, and depends on the joint distribution of the disruptions. The decision maker seeks an optimal design which minimizes the total expected cost.

In most of the existing reliable facility location literature, disruptions at different locations are assumed to be independent. However, in practice, correlated disruptions are widely observed. For example, consider the disruptions caused by Hurricane Sandy in October, 2012. Figure 2.1 shows the 48-hour forecast by the National Oceanic and Atmospheric Administration (NOAA) on the impact probability of Hurricane Sandy at the time of its landfall. The forecast consisted of a number of regions where the hurricane could cause severe impact with specific probability. Consider a customer in Columbus, OH, who is served primarily by a DC in Cleveland, OH, and backed-up by another DC in Pittsburgh, PA. As shown in the figure, the disruption probabilities of the DCs can be estimated as 40%

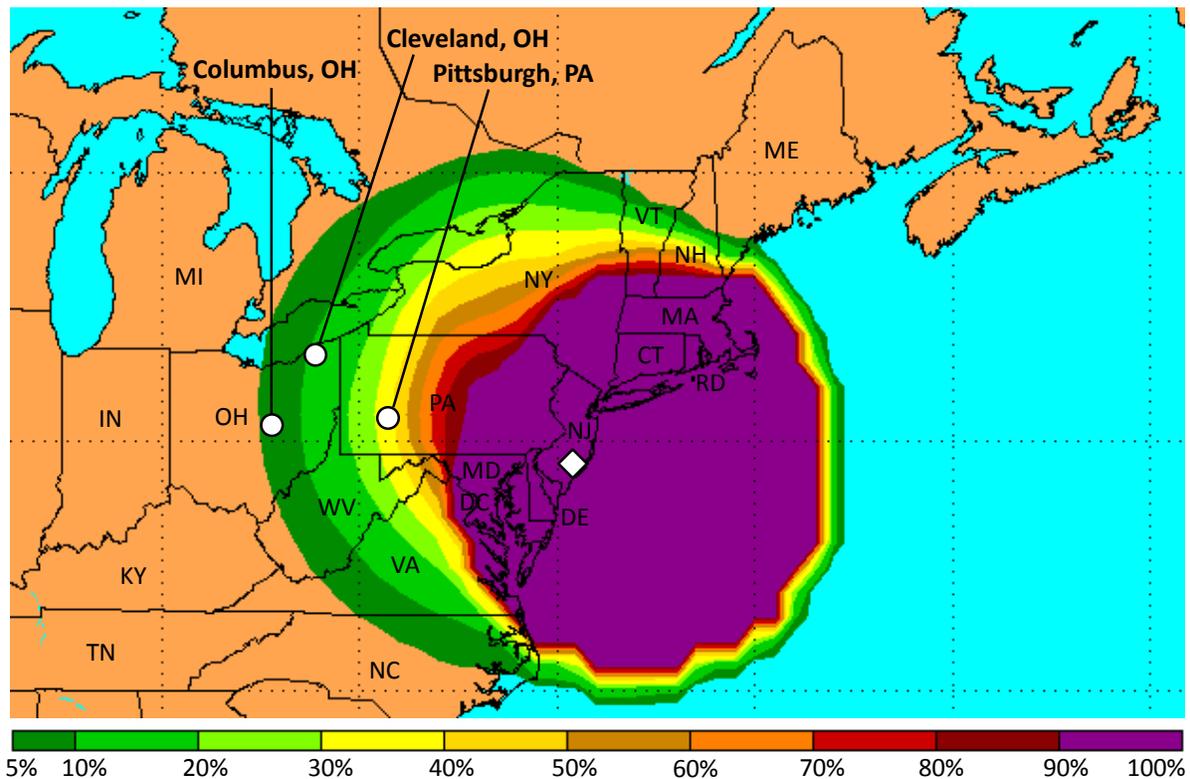


Figure 2.1: 48-hour impact probability forecast of Hurricane Sandy at the time of its landfall (NOAA, 10-29-2012)

Note: a location is impacted if it has tropical storm force surface wind (1-minute average speed ≥ 39 mph). “ \diamond ”: the hurricane center, “ \circ ”: facility or customer.

and 20%, respectively. If disruptions are assumed to be independent, the customer faces a fairly low risk of disruption with only 8% probability. However, under this circumstance, if Cleveland is impacted, Pittsburgh will also be impacted with a high probability, i.e., the disruptions at these two facilities are positively correlated. As a result, the customer could actually face a much higher risk, with disruption probability close to 20%.

The previous example shows that disruption correlation can significantly affect the magnitude of the disruption risk faced by the supply chain. As we will show later in this chapter, it also affects the optimal facility location design. However, due to the difficulty in estimation, modeling, and optimization, most of the existing literature on reliable facility location design only considered independent disruptions. In this chapter, we present a distributionally-robust optimization model to incorporate correlated disruptions. We assume that the disruptions have an unknown joint distribution, and minimize the expected cost under the worst-case distribution with given marginal disruption probabilities. Using the structural property of a class of widely-studied reliable facility location problems, we are

able to derive the worst-case distribution in closed-form, which has a practical interpretation. The sparse structure of the worst-case distribution also allows us to transform this seemingly complicated problem to a much simpler equivalent problem, and solve it efficiently.

We compare the optimal solutions of the robust model with those of the traditional model, which is based on the assumption of independent disruptions. We are particularly interested in the regret of the models, which is the increase in cost when the optimal solution of one model is erroneously used for the other model. We find that ignoring disruption correlation could lead to significant losses. On the other hand, the regret from applying the robust model under independent disruptions is much lower. As key factors, such as source disaster probability, disruption propagation effect, and service interruption penalty, increase, the regret of the traditional optimal design increases dramatically, while the regret of the robust design only increases mildly, or largely stays the same. In practice, we expect that the disruptions are positively correlated, but the correlation is smaller than the worst-case. We compare the two models under different degrees of correlation, and find that even though the robust model is based on the worst-case correlation, it still outperforms the traditional model when disruptions are only mildly correlated. We also consider a weighted-average objective consisting of the worst-case expected cost and the normal operating cost with no disruption. We find that most of the benefit of the robust model can be captured at a very small cost.

Given these advantages, we believe this robust model can serve as a promising alternative approach for reliable facility location problems. It does not require any additional model input, and thus can be applied directly to real-world problems that are currently being solved by the traditional approach, which is based on the assumption of independent disruptions. The robust model also requires much less computational effort, and thus can effectively handle large-scale problems.

The remainder of the chapter is organized as follows. Section 2.2 reviews related literature. In Section 2.3, we present the distributionally robust reliable facility location model and its equivalent formulation. Section 2.4 shows the numerical results. Section 2.5 discusses how the robust approach can be applied to other reliable facility location problem. Section 2.6 summarizes the results and discusses directions for future work.

2.2 Literature Review

In this section, we briefly review existing reliable facility location models, and discuss why most of these models are not applicable for correlated disruptions. We then review the few papers that either incorporated correlated disruptions or considered interdependence between locations, and discuss how our model differs from these models.

As noted by Snyder et al. (2012), there are two major streams of reliable facility location models: stochastic (S) models and robust (R) models. The stochastic models further fall into four main categories, namely, Scenario-Based (SB) models, Implicit Formulation (IF) models, Reliable Backup (RB) models, and Continuum Approximation (CA) models. For

Table 2.1: Summary of literature on reliable facility location

Category		Literature
S	SB	
	IF	Snyder and Daskin (2005), Berman, Krass, and Menezes (2007), Shen, Zhan, and Zhang (2011), Cui, Ouyang, and Shen (2010), Aboolian, Cui, and Shen (2013), Chen, Li, and Ouyang (2011), Li and Ouyang (2011), Li and Ouyang (2012), Li, Ouyang, and Peng (2013)
	RB	Lim et al. (2010), Li, Zeng, and Savachkin (2013), An, Zhang, and Zeng (2011), Liang, Shen, and Xu (2013)
	CA	Cui, Ouyang, and Shen (2010), Li and Ouyang (2010), Lim et al. (2013), Berman, Krass, and Menezes (2013)
	Other	Qi and Shen (2007), Qi, Shen, and Snyder (2010), Mak and Shen (2012)
R	IM	Church, Scaparra, and Middleton (2004), Church and Scaparra (2007), Scaparra and Church (2008a); Scaparra and Church (2008b), Losada et al. (2012), Liberatore, Scaparra, and Daskin (2012), An et al. (2012)
	Other	Snyder and Daskin (2006), Peng et al. (2011)

robust models, most of the existing literature is based on the Interdiction Median (IM) model. Table 2.1 summarizes the literature in these categories. For a more comprehensive and detailed review, please refer to Snyder et al. (2012). From the table, we can see that the IF model has been the most popular approach for stochastic reliable facility location. Thus, in this chapter, we refer to the IF model as the traditional model.

Next, we discuss why most of the existing models are not applicable or suitable for correlated disruptions. The SB model can incorporate correlated disruptions using sample average approximation (SAA). However, it has been shown that SAA performs poorly for independent disruptions (e.g., Shen, Zhan, and Zhang, 2011). We expect its performance would still be unsatisfactory, if not worse, for correlated disruptions. The IF model is based on calculating the probability of a customer being served by each facility, which requires that the disruptions are independent. The RB model assumes each customer is backed-up by a fixed fortified facility under all disruption scenarios. If the fortified facilities have infinite capacity (which is assumed by all existing literature using the RB model), disruption correlation will not affect the expected cost. In the capacitated case, disruption correlation does affect the expected cost. However, it only captures the effect of correlation on the probability of the aggregate demand exceeding the capacity constraint. It does not capture the effect of correlation on the probability of customers being rerouted to distant facilities, which is the main focus of reliable facility location models. The existing robust models consider all possible disruption scenarios. But they do not consider the probabilistic distribution of the disruption scenarios. Thus, they cannot model disruption correlation.

To our knowledge, CA is the only approach that has been successfully applied to incorporate correlated disruptions. Li and Ouyang (2010) considered the CA counterpart of the

IF model given the conditional probability of the disruptions. They found that the expected cost is higher when disruptions are positively correlated. Their numerical study shows that the impact of correlation on the expected cost can be significant when both the disruption probability and the service interruption penalty are high. Lim et al. (2013) considered the CA counterpart of the RB model with capacitated backup facilities. The main purpose is to study the effect of misspecifying the disruption probability and/or correlation on the relative regret. They found that the expected cost is increasing in the correlation and decreasing in the capacity. Their numerical result shows that joint underestimation of disruption probability and correlation results in higher loss compared to joint overestimation. In related work, Berman, Krass, and Menezes (2013) considered the continuous 2-median and 2-center problems restricted to a unit line segment. They derived in closed-form the trajectory of optimal locations as a function of the disruption probability and correlation.

The major difference between our model and the CA-based models is that our model is a discrete location model, while the CA model is a continuous location model. The continuous model requires that the demand can be well approximated by a continuous function, and that the potential locations are not restricted to given candidate sites. While these conditions may hold under certain circumstances (for example, individual customers within an urban area can be well approximated by a continuous function), they may not hold under many other circumstances. We consider a detailed supply chain design problem. The customers are distributed across the nation, and thus the demand is hard to approximate using a continuous function. Also, the potential locations for the facilities (e.g., warehouses and distribution centers) are typically restricted to a number of candidate sites. Thus, we believe a discrete model is more suitable for this setting.

Given the difference in the nature and specific settings of the models, it may not be completely appropriate to directly compare the results and insights from this chapter and those from the CA-based papers. Nonetheless, we notice the following key differences. First, in contrast to Li and Ouyang (2010) who found that the regret of ignoring correlation is usually not significant, we find that such regret is significant in our real-world motivated case study and most of our simulated examples, and it is also much higher than the regret from using the robust design for independent disruptions. Also, Li and Ouyang (2010) found that the number of opened facilities is smaller when disruptions are correlated, while we find the opposite result. We think these differences are probably due to the different nature (i.e., discrete vs. continuous) of the models and the difference in the correlation structure. Second, Lim et al. (2013) found that the effect of misspecification in disruption correlation alone is very limited. We find that misspecification of correlation alone can also result in significant loss, and overestimating correlation (i.e., assuming worst-case correlation) is in general better than underestimating (i.e., assuming independence). We think these differences are probably due to the fact that Lim et al. (2013) considered the CA counterpart of the RB model with capacitated backup, which, as we mentioned, does not reflect the effect of correlation on rerouting customers to distant facilities.

In addition to the CA-based models, the literature includes discrete location models that considered specific (deterministic) interdependence structure between locations. Liberatore,

Scaparra, and Daskin (2012) considered one type of interdependence known as the “ripple effect”, where a disruption at one location causes a nearby facility to lose a fixed proportion of capacity. They incorporated the ripple effect in the IM model with fortification decisions. Our model differs from Liberatore, Scaparra, and Daskin (2012) in that the IM model is for determining the worst-case disruption scenarios for a given design, while our model is for determining the optimal design. Another difference is that we consider correlated random disruptions, while Liberatore, Scaparra, and Daskin (2012) considered a deterministic interdependence structure between locations. Li, Ouyang, and Peng (2013) considered another type of interdependence known as “supporting station”, where different locations may require resources provided by the same supporting station. Thus, independent disruptions to the supporting stations may result in correlated disruptions to the facilities. Technically speaking, the supporting station model is still a model with independent disruptions. Our model does not require the special structure of supporting stations, and thus can be applied under more general settings.

In summary, our model significantly differs from the existing literature. In contrast to the CA-based models, our model is a discrete model which is applicable under more general problem settings. We also draw new insights from the numerical results. Compared with the models of Liberatore, Scaparra, and Daskin (2012) and Li, Ouyang, and Peng (2013), our model is based on correlated random disruptions instead of special interdependence structure.

2.3 Model and Formulation

In this section we focus on the reliable uncapacitated fixed-charge location (RUFL) problem, as an example to illustrate the distributionally-robust optimization model for reliable facility location design, and show how it can be transformed to an equivalent problem. The same approach applies to other widely studied reliable facility location problems, including the p -median problem, the capacitated fixed-charge location problem, and the multi-allocation hub location problem.

Consider the problem of locating facilities at a set $\mathcal{J} = \{1, \dots, J\}$ of candidate locations to serve a set $\mathcal{I} = \{1, \dots, I\}$ of customers. Let d_i denote the demand of customer $i \in \mathcal{I}$, and f_j the fixed cost of opening a facility at location $j \in \mathcal{J}$. Serving customer i from a facility at location j incurs unit transportation cost c_{ij} . Let $\mathbf{x} = (x_0, x_1, \dots, x_J)$ denote the facility location decision, where $x_j = 1$ if facility is opened at location j , and $x_j = 0$ otherwise. The facilities are subject to random disruptions. Let $\boldsymbol{\xi} = (\xi_0, \xi_1, \dots, \xi_J)$ denote the disruption scenario, where $\xi_j = 0$ if location j is disrupted, and $\xi_j = 1$ if it is online, i.e., not disrupted. We will sometimes, for convenience, use the set of online locations, S , to denote the disruption scenario, with the correspondence

$$S(\boldsymbol{\xi}) = \{j \in \mathcal{J} : \xi_j = 1\},$$

and

$$\boldsymbol{\xi}(S) = (I(0 \in S), I(1 \in S), \dots, I(J \in S)),$$

where $I(\cdot)$ is the indicator function.

Given \mathbf{x} and $\boldsymbol{\xi}$, each customer is either assigned to an available (i.e., opened and online) facility (with $y_{ij} = 1$ if customer i is assigned to facility j , and $y_{ij} = 0$ otherwise), or its service is interrupted. In order to model service interruptions, a virtual facility 0 is added to \mathcal{J} . $y_{i0} = 1$ means customer i 's service is interrupted, with $c_{i,0}$ being the unit penalty cost. The virtual facility is never disrupted, i.e., $\xi_0 \equiv 1$, and its fixed cost $f_0 = 0$. Note that in the RUFL model, the facilities are uncapacitated. Thus, service interruptions, if any, are not due to limited supply capacity. In Section 2.5, we show how to handle limited capacity.

Let $h(\mathbf{x}, \boldsymbol{\xi})$ denote the transportation and penalty cost under the optimal customer assignment/interruption decisions, given location design \mathbf{x} and disruption scenario $\boldsymbol{\xi}$, i.e.,

$$\begin{aligned} h(\mathbf{x}, \boldsymbol{\xi}) = \min & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_i c_{ij} y_{ij} \\ \text{s.t.} & \sum_{j \in \mathcal{J}} y_{ij} = 1, \quad \forall i \in \mathcal{I} \\ & y_{ij} \leq x_j \xi_j, \quad \forall i \in \mathcal{I}, \quad \forall j \in \mathcal{J} \\ & y_{ij} \in \{0, 1\}, \quad \forall i \in \mathcal{I}, \quad \forall j \in \mathcal{J} \end{aligned} \quad (2.1)$$

Let $p(\boldsymbol{\xi})$ be the joint distribution of the disruptions, i.e., $p(\boldsymbol{\xi})$ is the probability that disruption scenario $\boldsymbol{\xi}$ occurs, the RUFL problem is defined as

$$(\text{RUFL}) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{j \in \mathcal{J}} f_j x_j + \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\},$$

where $\mathcal{X} = \{\mathbf{x} : x_j \in \{0, 1\}, \forall j \in \mathcal{J}\}$. Traditional RUFL models (e.g., Snyder and Daskin, 2005; Cui, Ouyang, and Shen, 2010) consider the special case where disruptions are independent, i.e.,

$$p(\boldsymbol{\xi}) = \prod_{j \in \mathcal{J}} (1 - q_j)^{\xi_j} (q_j)^{1 - \xi_j},$$

where q_j is the marginal disruption probability of location j .

In distributionally-robust optimization, instead of assuming some specific joint distribution, we assume $p(\boldsymbol{\xi})$ to be uncertain, but within a distributional uncertainty set. In specific, we consider the set of all joint distributions such that the marginal disruption probability of location j is equal to q_j , i.e.,

$$\mathcal{P} = \left\{ p \left| \begin{array}{l} \sum_{S: j \in S} p(S) = 1 - q_j, \quad \forall j \in \mathcal{J} \\ p(S) \geq 0, \quad \forall S \subseteq \mathcal{J} \\ p(S) = 0, \quad \forall S, 0 \notin S \end{array} \right. \right\}.$$

Recall that the virtual facility is never disrupted, i.e., $0 \in S$ for any disruption scenario S , and the disruption probability $q_0 = 0$. Thus, we have the following constraint

$$\sum_{S:0 \in S} p(S) = 1 - q_0 = 1.$$

This constraint guarantees that $p(S)$ is a probability measure. Also, Note that the distributional uncertainty set \mathcal{P} does not require any additional model input other than the marginal disruption probability. Thus, it is possible to directly compare the robust model with the traditional model.

The distributionally-robust reliable uncapacitated fixed-charge location (DR-RUFL) problem minimizes the expected cost under the worst-case distribution in \mathcal{P} , i.e., the one that leads to the maximum expected cost,

$$(\text{DR-RUFL}) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{j \in \mathcal{J}} f_j x_j + \max_{p \in \mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\}. \quad (2.2)$$

Distributionally-robust optimization has been extensively studied and applied to various problems. For a review, please refer to Bertsimas, Brown, and Caramanis (2011). More specifically, our model falls into the category of marginal moment models studied by Bertsimas, Natarajan, and Teo (2004). Agrawal et al. (2010); Agrawal et al. (2012) also studied the marginal moment models. Their focus is to derive an upper bound on the regret from ignoring correlation for a class of problems. Most reliable facility location models are not in this class, which means ignoring correlation can result in substantial regret.

Considering the worst-case distribution is certainly conservative. However, we believe it can usually be justified in practice. First, previous studies suggest that in supply chain risk management, managers are more concerned of the “maximum exposure”, i.e., the worst-case (Tang, 2006). Second, as we will discuss later, the worst-case distribution for the DR-RUFL problem has a practical interpretation. Under certain circumstances, we expect that it is closer to the actual distribution than the independent distribution. Third, since the actual distribution is typically unknown, given only the marginal probability, one could either assume the disruptions are independent, or apply the DR-RUFL model. Our numerical results in Section 2.4 show that the latter option usually outperforms the former. Furthermore, the optimal design under the worst-case distribution is not expensive to implement in practice, and much of its benefit can be achieved at a relatively low cost.

Equivalent Formulation of DR-RUFL

The DR-RUFL problem in (2.2) is a mini-max formulation. The inner problem has the goal of choosing the worst disruption distribution p for a given design \mathbf{x} , which can be formulated

as a linear program

$$\begin{aligned}
 \max \mathbb{E}_p[h(\mathbf{x}, S)] &= \max \sum_{S \subseteq \mathcal{J}} h(\mathbf{x}, S)p(S) \\
 \text{s.t.} \quad &\sum_{S: j \in S} p(S) = 1 - q_j, \quad \forall j \in \mathcal{J} \\
 &p(S) \geq 0, \quad \forall S \subseteq \mathcal{J} \\
 &p(S) = 0, \quad \forall S, 0 \notin S
 \end{aligned} \tag{2.3}$$

This linear program has 2^J variables, which could still make the DR-RUFL problem computationally intractable. However, due to a property of RUFL, we can derive the worst-case distribution in a closed-form that does not depend on \mathbf{x} or $h(x, S)$. The DR-RUFL problem can then be transformed into a much simpler equivalent problem and solved efficiently.

First, we need to show that with any given \mathbf{x} , the cost function $h(\mathbf{x}, S)$ in (2.1) is supermodular in S . A set function g is said to be supermodular if for any $S, T \subseteq \mathcal{J}$,

$$g(S \cap T) + g(S \cup T) \geq g(S) + g(T).$$

g is supermodular if and only if for any $S \subset T \subset \mathcal{J}$, and any $j \in \mathcal{J} \setminus T$,

$$g(S \cup \{j\}) - g(S) \leq g(T \cup \{j\}) - g(T). \tag{2.4}$$

Condition (2.4) is known as the condition of increasing differences. $g(S \cup \{j\}) - g(S)$ is the difference in function value from augmenting subset S with $j \in \mathcal{J} \setminus S$. Similarly, $g(T \cup \{j\}) - g(T)$ is the difference in function value from augmenting subset T with $j \in \mathcal{J} \setminus T$. The differences are increasing if for any $S \subset T$, $g(S \cup \{j\}) - g(S) \leq g(T \cup \{j\}) - g(T)$. Many reliable facility location problems have increasing differences. The intuition is that having additional available facilities has diminishing marginal returns. For the RUFL problem, we have the following lemma.

Lemma 1 (Supermodularity). *For any $\mathbf{x} \in \mathcal{X}$, the cost function $h(\mathbf{x}, S)$ given in (2.1) is supermodular in S .*

Using supermodularity, we can derive the worst-case distribution. Without loss of generality, assume the facilities are indexed in ascending order of marginal disruption probabilities, i.e.,

$$0 \equiv q_0 \leq q_1 \leq \dots \leq q_J \leq q_{J+1} \equiv 1.$$

Consider $J + 1$ disruption scenarios denoted by $\xi^0, \xi^1, \dots, \xi^J$. The s -th scenario is defined as

$$\xi^s = (\xi_0^s, \xi_1^s, \dots, \xi_J^s),$$

where $\xi_j^s = I(j \leq s)$ for all $j \in \mathcal{J}$, and $I(\cdot)$ is the indicator function. In other words, in the s -th scenario, the less reliable locations $s + 1, \dots, J$ are disrupted, and the more reliable locations $0, 1, \dots, s$ are online.

We will show that in the worst-case distribution, only the $J + 1$ disruption scenarios we just defined will have non-zero probability. All the other disruption scenarios will have zero probability. Also, the worst-case distribution does not depend on the location design \mathbf{x} , and can be found in closed-form. The following lemma is due to Edmonds (1971) and Agrawal et al. (2010). It can also be shown using basic linear program duality.

Lemma 2 (Worst-case disruption distribution). *In the worst-case disruption distribution for DR-RUFL, only disruption scenarios $\xi^0, \xi^1, \dots, \xi^J$ may have nonzero probabilities, and the probability of scenario ξ^s is equal to $q_{s+1} - q_s$ for all $s = 0, 1, \dots, J$.*

To better understand Lemma 2, consider the case where a hazard originates at a source, and propagates along certain direction (for example, an earthquake). Assume there are J impact regions which are indexed in ascending order of impact probabilities, (e.g., region 1 is the outermost region, and region J is the innermost region), and assume there is a candidate location in each region. In scenario ξ^s , facilities in regions $s + 1, \dots, J$ are disrupted, i.e., the disruption has propagated far enough to reach region $s + 1$, and thus all regions that are closer to the hazard source. On the other hand, facilities in regions $1, \dots, s$ are online, i.e., the disruption has not propagated far enough to reach region s , and thus all regions beyond it. If we assume the disruption cannot “jump” to a further region without impacting all regions closer to the hazard source, then we can see that only scenarios ξ^s , $s = 0, 1, \dots, J$ are possible. The probability of ξ^s is the probability of reaching region $s + 1$ but not reaching region s , which is equal to $q_{s+1} - q_s$.

We would like to point out that in the DR-RUFL model, we do not make any assumption on the structure of the disruption. The propagation example we just described is only a practical interpretation of the worst-case distribution. In this situation, the worst-case distribution is really close to the actual distribution. There are certainly other disruption structures. The worst-case distribution can still be used to approximate the unknown actual distribution, and we will show its performance is better than the traditional model in the numerical study presented later in this chapter.

A direct result from Lemma 2 is the worst-case correlation. Let ρ_{jk}^* be the worst-case correlation between locations j and k , with $j < k$. It is easy to verify that

$$\rho_{jk}^* = \sqrt{\frac{q_j(1 - q_k)}{q_k(1 - q_j)}}. \quad (2.5)$$

Two observations can be made. First, as a result of supermodularity, the worst-case correlation achieves the maximum correlation with the given marginal disruption probability. Second, the correlation is stronger between locations with similar marginal disruption probabilities. We think this partially reflects practical situations, since facilities that are close to each other tend to have similar disruption probabilities, and they are also more likely to be disrupted at the same time due to common hazards.

Another observation from Lemma 2 is that the worst-case disruption distribution only depends on the marginal disruption probability, but not on the transportation cost. Recall

that traditional RUFL models are based on the implicit formulation (IF) model where customers are assigned to multiple backup facilities with different backup levels. The level r backup facility will only be used, if level 1 through level $r - 1$ backup facilities are disrupted. Under independent disruptions, it is optimal to assign backup facilities level by level in increasing order of transportation cost without considering reliability, given that the number of backup level is large enough (Cui, Ouyang, and Shen, 2010). However, under the worst-case correlated distribution, if the level r backup facility is less reliable than the level $r - 1$ backup facility, it will be disrupted whenever the level $r - 1$ facility is disrupted. Thus, assigning a less reliable facility as a higher level backup is meaningless. This shows when disruptions are correlated, one needs to consider both transportation cost and reliability in determining backup levels.

Using the worst-case disruption distribution, we obtain an equivalent formulation of the DR-RUFL problem, which we refer to as the worst-case reliable uncapacitated fixed-charge location (WC-RUFL) problem.

Proposition 3 (Equivalent formulation). *The DR-RUFL problem is equivalent to*

$$(WC-RUFL) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{j \in \mathcal{J}} f_j x_j + \sum_{s \in \mathcal{J}} (q_{s+1} - q_s) h(\mathbf{x}, \boldsymbol{\xi}^s) \right\}.$$

The WC-RUFL problem is a stochastic program with only $J + 1$ scenarios, and thus can be solved efficiently using standard methods such as Benders decomposition (e.g., Magnanti and Wong, 1981).

2.4 Numerical Results

In this section, we use numerical results to show the advantage of the distributionally-robust model over the traditional model that assumes independent disruptions. First, we will present an example motivated by a real-world situation and show how considering disruption correlation affects the optimal location design. Then, we will compare the two models with numerical experiments and draw managerial insights.

Example: Supply Chain Design under Severe Weather Hazards

We consider the case where a large nation-wide company is planning its distribution center (DC) network to serve retail stores that replenish from the DCs. The customers are represented by the 48 states in the contiguous US and Washington, D.C., and the demand is proportional to the state population. The fixed cost for a DC is proportional to the median home price, and the unit transportation cost is proportional to the Great Circle Distance (calculated using the geographic coordinates of the state capitals). Specifically, we use the



Figure 2.2: The optimal design under independent disruptions (Design I)

49-node data set described in Daskin (1995). This data set has been widely used in location analysis, especially for designing supply chains. It has also been used in reliable facility location problems (e.g., Snyder and Daskin, 2005).

Based on previous experience, disruptions to the DCs are mainly caused by severe weather hazards, e.g., tornadoes, storms, etc. It is assumed that disrupted DCs will not be able to serve customer demand for the entire planning horizon (a quarter). The company has obtained the severe weather hazard data from the Storm Prediction Center of the NOAA. Using this data set, it can estimate the marginal disruption probabilities. More details can be found in Appendix B.1. Although the company tries its best to fulfill all the demand, service interruption is still possible under a severe disruption. It is estimated that there is a \$40,000 penalty cost for each unit of unfulfilled customer demand. Given all these data, the company seeks to design a reliable supply chain to minimize the total cost consisting of the fixed cost and the expected transportation/penalty cost.

Since only the marginal disruption probabilities are available, the manager is faced with two options. The first option is to assume that the disruptions are independent and apply the traditional RUFL model (e.g., Cui, Ouyang, and Shen, 2010). The optimal design (I) is shown in Figure 2.2. The second option is to consider all joint distributions with the same marginal disruption probability and apply the DR-RUFL model. The optimal design (R) is shown in Figure 2.3. The details of the two designs can be found in Table 2.2. We see that the two designs only differ by two facilities (and the customer assignments related to these facilities). The California (CA) and Michigan (MI) facilities in design I are moved to Nevada (NV) and West Virginia (WV), respectively, in design R. Table 2.3 compares the performance of the two designs. When there is no disruption, or when the disruptions are independent, design I performs slightly better. Implementing design R will increase the expected cost by



Figure 2.3: The optimal design under worst-case correlated disruptions (Design R)

Table 2.2: Optimal location designs of the two models

Design I			Design R		
Location	% Disruption	% Demand	Location	% Disruption	% Demand
PA	6.96	28.93	–	–	25.52
TX	7.95	8.76	–	–	–
AL	10.80	16.58	–	–	15.17
IA	12.88	15.27	–	–	–
CA	1.45	18.56	NV	0.83	18.56
MI	6.99	11.89	WV	3.45	16.71

Note: % Disruption: marginal disruption probability, % Demand: proportion of total demand served, “–”: same as design I.

Table 2.3: Comparison of the performance of the two designs

Performance	Design I			Design R		
	Cost	Increase	% Increase	Cost	Increase	% Increase
No disruption	857,166	–	–	889,919	32,753	3.82
Independent disruption	927,027	–	–	945,836	18,808	2.03
Worst-case correlated	2,495,053	663,401	36.22	1,831,652	–	–

Note: Increase: increase in cost compared to the design with the lower cost, % Increase: relative increase, “–”: not applicable.

4% and 2%, respectively. However, under the worst-case distribution, design R performs much better than design I, and implementing design I will increase the expected cost by over 30%.

To understand why design I results in such high additional costs, consider the Pennsylvania (PA) facility which serves the most populous New England and Mid-Atlantic regions. In design I, almost 30% of the total demand is served by the PA facility. When PA is disrupted, most of these customers will be rerouted to the MI facility. However, since PA and MI have marginal disruption probabilities 5.01% and 5.03%, the correlation between them can be very high. As a result, MI usually fails to serve as a backup facility. On the other hand, in design R, most of the customers served by PA are backed-up by the much more reliable WV facility. The correlation between PA and WV has a much lower upper bound. Thus, WV is a much more effective backup than MI is.

The worst-case correlation is a conservative estimate. Under the actual correlation, the regret of design I will be smaller. (The regret of design R will also be smaller.) However, we expect that the actual correlation between PA and MI is still relatively high, as these two locations are more likely to be affected by common hazard originating from the Great Lakes. We also expect that the robust design is more favorable in practice, since the extra cost (i.e., the increase in cost when there are no disruptions or when the disruptions are independent) is small but the potential savings is huge. As we mentioned in Section 2.3, there are also other reasons to consider the worst-case distribution rather than the independent distribution. In the next subsection, we show this by numerical experiments using simulated data.

Numerical Results

We would like to compare the robust model and the traditional model in a more comprehensive numerical study. Instead of using the severe weather hazard data from the NOAA, we generate the disruption probabilities in the same way as Cui, Ouyang, and Shen (2010). Let α be the probability that a disastrous event occurs at a certain source. The disaster then propagates and causes disruptions to facilities at different distances from the source. The marginal disruption probability decreases exponentially in the distance. Let D_j be the distance of location j from the source, then the marginal disruption probability of location j is given by

$$q_j = \alpha e^{-D_j/\theta},$$

where θ is a parameter that measures the strength of the disruption propagation effect. The source disaster probability α , the disruption propagation factor θ , along with the service interruption penalty ω , are the key factors that significantly affect the cost and the optimal design. For each factor, we consider three levels, which gives us 27 different combinations, as shown in Table 2.4.

We use the same demand, fixed cost, and transportation cost data as in the previous example, i.e., the 49-node data set in Daskin (1995). Similar results were found using a larger data set in Daskin (1995). Those results are available in Appendix B.2. The robust model

Table 2.4: Levels for important factors

factor	low	medium	high
α	0.1	0.2	0.3
θ	200	400	800
ω	20000	40000	80000

is solved using an accelerated Benders decomposition algorithm. The traditional model is solved using the search-and-cut (SnC) algorithm in Aboolian, Cui, and Shen (2013), which to our knowledge is the state-of-art for the traditional RUFL model. Both algorithms are implemented and tested using ILOG Cplex 12.4 with MATLAB R2009b on a Intel Core i7-930 2.80 GHz quad core processor running 64-bit Windows 7. The SnC algorithm uses 4 levels of backup and a neighborhood size of 3 (for details please refer to Aboolian, Cui, and Shen, 2013), and is solved to a 0.1 % optimality gap or 7200 seconds maximum runtime, whichever occurs first.

Table 2.5 summarizes the solutions under different α , ω , and θ . The subscript R represents the robust model and the subscript I represents the traditional with independent disruptions. n is the number of open facilities in the optimal solution. z is the optimal expected cost. Δz is the regret, i.e., the increase in cost when the optimal solution under one disruption distribution is erroneously used in the other disruption distribution. For example, Δz_R is the regret if the optimal solution under independent disruptions is used when the disruptions are actually worst-case correlated. $\% \Delta z$ is the percentage relative regret, i.e., $\% \Delta z = 100 \times \Delta z / z$. CPU and GAP are the computation time and optimality gap, respectively, when the algorithm terminates.

From Table 2.5, we have several observations. First, comparing columns n_R and n_I , we see that the number of opened facilities in the robust solution is greater than or equal to that of the independent solution for all instances. This shows that more facilities are required to mitigate correlated disruptions. Second, from columns Δz_R and $\% \Delta z_R$, we see that failing to consider disruption correlation could lead to significant loss, with an average regret of 187,000 (11.98%). For some of the instances, the relative regret is more than 20%. On the other hand, from columns $\% \Delta z_I$ and $\% \Delta z_I$, we see that although assuming the worst-case correlation is conservative, it does not lead to a significant cost increase even when disruptions are independent, with an average regret of 25,000 (2.76%), and the relative regret is less than 8% for all instances. Finally, we consider the computational performance. Comparing columns CPU $_R$ and CPU $_I$, and columns GAP $_R$ and GAP $_I$, we see that the robust model requires much less computational effort than the traditional model. This gives the robust model a great advantage for solving large-scale problems.

From Table 2.5, we also see that the performance of the solutions is significantly affected by the parameters α , ω , and θ . In Figure 2.4, we show the impact of these factors on the regret. Consider the source disaster probability α , for example. For each level of α , we consider different combinations of the other two factors, i.e., ω and θ , and compare the

Table 2.5: Selected results for the 49-node data set

α	ω	θ	n_R	z_R	Δz_R	$\% \Delta z_R$	CPU $_R$	GAP $_R$	n_I	z_I	Δz_I	$\% \Delta z_I$	CPU $_I$	GAP $_I$
0.1	20000	200	6	8.66	0.00	0.04	2.64	0.00	5	8.63	0.00	0.00	26.08	0.02
0.2	20000	200	6	8.74	0.00	0.04	1.98	0.00	5	8.68	0.00	0.02	170.57	0.00
0.3	20000	200	6	8.82	0.00	0.06	2.81	0.00	5	8.74	0.00	-0.04	200.14	0.07
0.1	40000	200	6	8.66	0.00	0.04	2.80	0.00	5	8.63	0.00	0.00	30.32	0.02
0.2	40000	200	6	8.75	0.00	0.04	2.54	0.00	5	8.68	0.00	0.02	194.73	0.00
0.3	40000	200	6	8.83	0.00	0.06	2.36	0.00	5	8.74	0.00	-0.04	223.67	0.07
0.1	80000	200	6	8.67	0.00	0.04	2.80	0.00	5	8.63	0.00	0.00	33.42	0.02
0.2	80000	200	6	8.76	0.00	0.04	2.84	0.00	5	8.68	0.00	0.02	217.05	0.00
0.3	80000	200	6	8.86	0.00	0.06	2.47	0.00	5	8.74	0.00	-0.04	251.07	0.07
0.1	20000	400	6	9.39	0.02	0.20	2.14	0.00	6	8.74	-0.01	-0.08	348.97	0.10
0.2	20000	400	7	10.13	0.09	0.93	4.11	0.00	6	8.90	0.23	2.60	1772.63	0.09
0.3	20000	400	7	10.80	0.24	2.24	2.84	0.00	6	9.05	0.24	2.70	4905.02	0.07
0.1	40000	400	7	9.76	0.10	1.05	3.20	0.00	6	8.74	0.22	2.57	396.79	0.10
0.2	40000	400	7	10.70	0.43	3.98	2.57	0.00	6	8.90	0.23	2.60	2013.53	0.09
0.3	40000	400	7	11.62	0.77	6.64	3.56	0.00	6	9.05	0.39	4.35	5551.85	0.07
0.1	80000	400	7	10.33	0.43	4.21	2.75	0.00	6	8.74	0.22	2.57	442.22	0.10
0.2	80000	400	7	11.75	1.18	10.07	3.69	0.00	6	8.90	0.38	4.28	2244.51	0.09
0.3	80000	400	7	13.15	1.95	14.84	3.88	0.00	6	9.05	0.39	4.35	6184.70	0.07
0.1	20000	800	7	13.36	0.73	5.47	2.89	0.00	6	8.94	0.35	3.89	1225.04	0.00
0.2	20000	800	7	17.75	1.80	10.16	2.12	0.00	6	9.30	0.33	3.50	7237.74	1.13
0.3	20000	800	8	21.94	3.11	14.16	2.94	0.00	6	9.68	0.78	8.09	7234.39	10.36
0.1	40000	800	7	16.90	1.90	11.27	3.05	0.00	6	8.94	0.35	3.89	1394.15	0.00
0.2	40000	800	7	24.85	4.15	16.72	2.87	0.00	6	9.30	0.33	3.50	7219.95	1.43
0.3	40000	800	8	32.57	6.63	20.35	2.76	0.00	6	9.68	0.78	8.09	7273.99	6.63
0.1	80000	800	7	23.99	4.25	17.73	2.61	0.00	6	8.94	0.35	3.89	1552.36	0.00
0.2	80000	800	7	39.03	8.85	22.68	2.88	0.00	6	9.30	0.33	3.50	7304.21	1.64
0.3	80000	800	8	53.85	13.68	25.40	2.73	0.00	6	9.68	0.78	8.09	7209.54	7.02
Average			6.74	15.58	1.87	11.98	2.85	0.00	5.67	8.96	0.25	2.76	2698.47	1.08

Note: R : robust model, I : independent model, n : number of facilities, z : expected cost ($\times 10^5$), Δz : regret ($\times 10^5$), $\% \Delta z$: relative regret (%), CPU: computation time (s), GAP: optimality gap at termination (%).

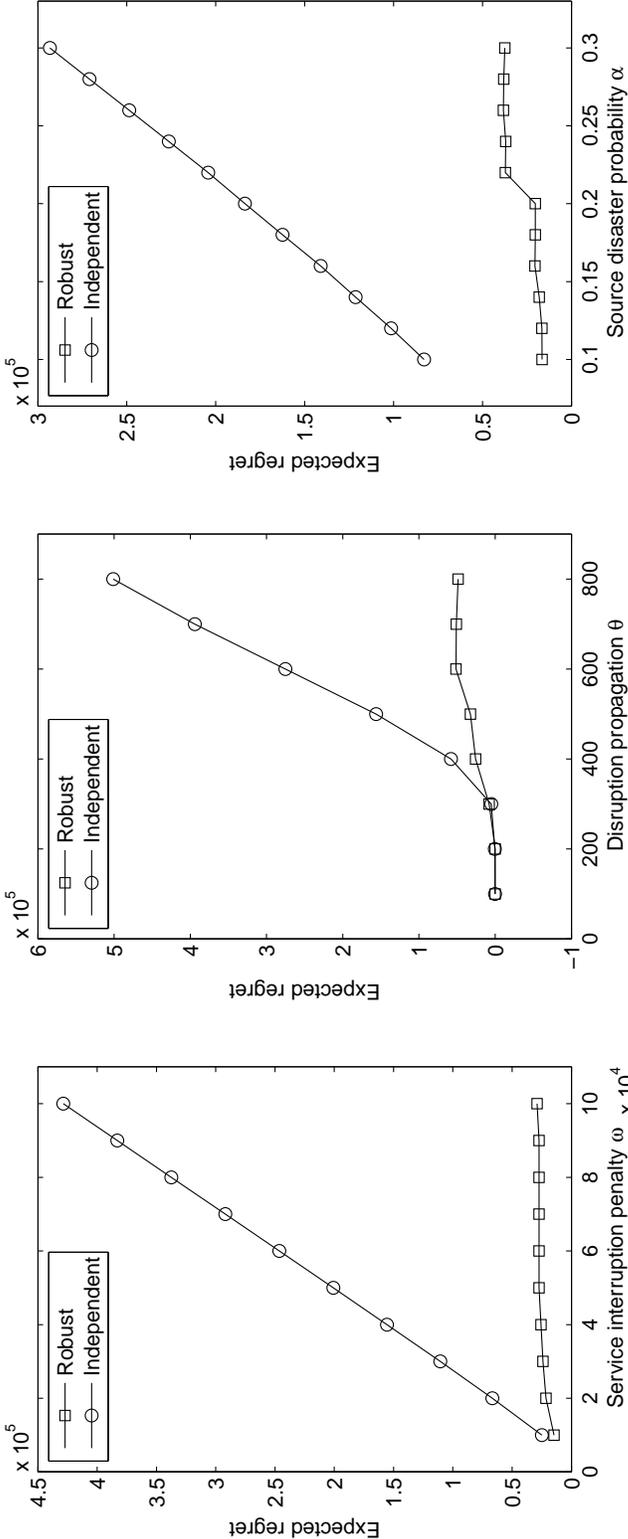


Figure 2.4: Impact of important factors on expected regret

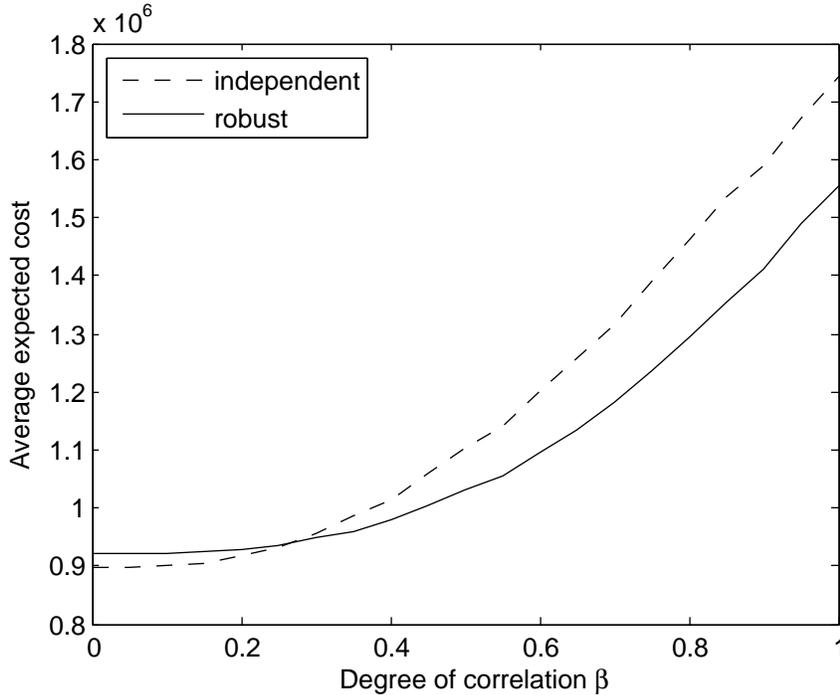


Figure 2.5: Expected cost under different degrees of disruption correlation

average regret. We see that the regret of the independent solution increases dramatically as α increases. On the other hand, the regret of the robust solution only increases mildly. Similar results are observed for ω and θ .

We have compared the robust solution and the independent solution under two extreme distributions, i.e., disruptions are either independent or worst-case correlated. In practice, we expect that the actual distribution is between the two extreme cases, i.e., the disruptions are positively correlated, but the correlation is smaller than the worst-case. We also like to compare the performance of the robust and independent solutions under these intermediate cases. Assume the disruption correlation between location j and k is given by $\beta\rho_{jk}^*$, where ρ_{jk}^* is the worst-case correlation given in (2.5), and $\beta \in [0, 1]$ is a parameter that controls the degree of correlation. When $\beta = 0$ or 1 , the joint distribution reduces to the independent or the worst-case distribution, respectively. We use simulation to evaluate the expected cost of the optimal robust solutions and independent solutions under different β . The average cost is shown in Figure 2.5. We see that even though the independent model has a slightly lower expected cost when the degree of correlation is close to zero, the robust model starts to outperform the independent model under mildly correlated disruptions (e.g., $\beta = 0.3$), and has a substantial advantage when the correlation is relatively high.

One common criticism of robust optimization is that it focuses on the worst case, and thus can be overly conservative and too expensive to implement in practice. In order to

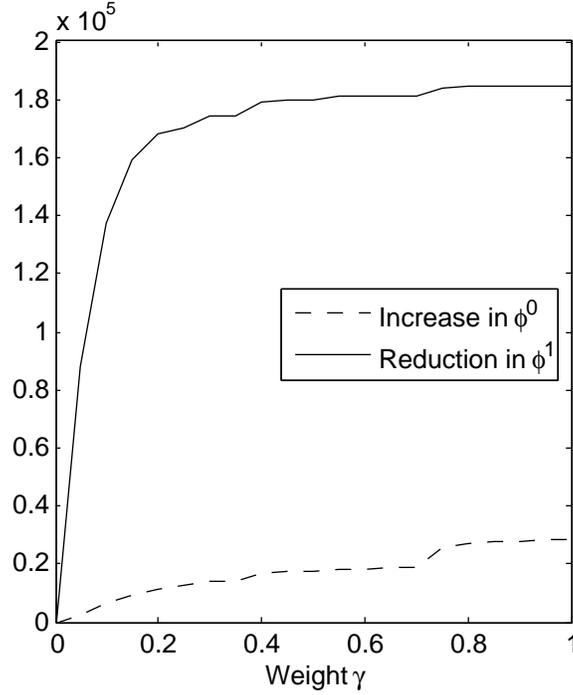


Figure 2.6: Benefit and cost of the robust design with different weights

address this issue, we consider a weighted-average objective function

$$\phi^\gamma(\mathbf{x}) = \gamma\phi^1(\mathbf{x}) + (1 - \gamma)\phi^0(\mathbf{x}),$$

where ϕ^1 is the expected cost of the DR-RUFL problem, and ϕ^0 is the normal operating cost, i.e., the cost when there is no disruption. $\gamma \in [0, 1]$ is a parameter that measures the degree of conservativeness. By Proposition 3,

$$\phi^1(\mathbf{x}) = \sum_{j \in \mathcal{J}} f_j x_j + \sum_{s=0}^J (q_{s+1} - q_s) h(\mathbf{x}, \boldsymbol{\xi}^s).$$

Recall that in scenario $\boldsymbol{\xi}^J$, no facility is disrupted. Thus,

$$\phi^0(\mathbf{x}) = \sum_{j \in \mathcal{J}} f_j x_j + h(\mathbf{x}, \boldsymbol{\xi}^J).$$

As a result, the weighted-average objective can be easily incorporated by adjusting the disruption probabilities, i.e.,

$$\phi^\gamma(\mathbf{x}) = \sum_{j \in \mathcal{J}} f_j x_j + \sum_{s=0}^{J-1} \gamma (q_{s+1} - q_s) h(\mathbf{x}, \boldsymbol{\xi}^s) + (1 - \gamma q_J) h(\mathbf{x}, \boldsymbol{\xi}^J).$$

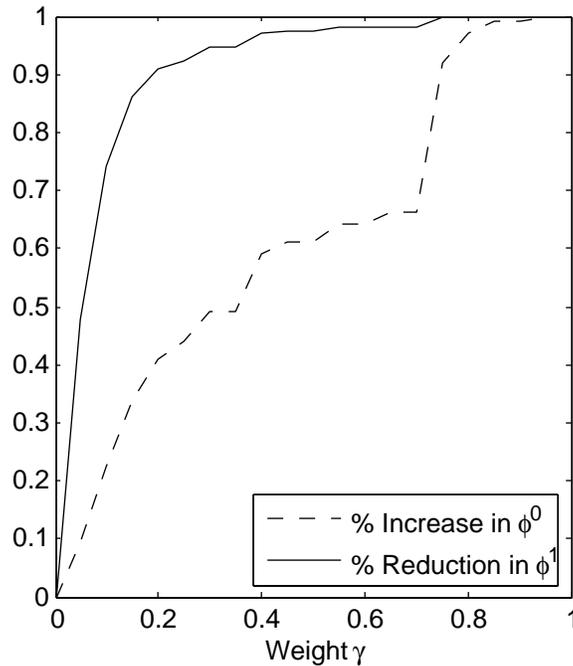


Figure 2.7: Percentage of maximum benefit captured and maximum cost incurred by different weights

Let $\mathbf{x}^\gamma = \operatorname{argmin} \{\phi^\gamma(\mathbf{x})\}$, i.e., the optimal solution with parameter γ . If $\gamma = 0$, \mathbf{x}^0 is the optimal solution when there is no disruption, i.e., the most cost-effective but also the least reliable design. For $\gamma > 0$, applying the reliable design \mathbf{x}^γ instead of design \mathbf{x}^0 has two effects. It will reduce the worst-case expected cost by $\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^\gamma)$, which is its benefit, while increasing the normal operating cost by $\phi^0(\mathbf{x}^\gamma) - \phi^0(\mathbf{x}^0)$, which can be considered as its cost. Figure 2.6 compares these two effects under different γ . We see that a large amount of benefit can be achieved with a relatively small cost.

On the other hand, when $\gamma = 1$, \mathbf{x}^1 is the optimal solution for the DR-RUFL problem, i.e., the most reliable but also the most conservative design. It will result in the maximum benefit $\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^1)$ and the maximum cost $\phi^0(\mathbf{x}^0) - \phi^0(\mathbf{x}^1)$. For $0 < \gamma < 1$, the ratio

$$\frac{\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^\gamma)}{\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^1)}$$

is the proportion of maximum benefit captured by \mathbf{x}^γ . Similarly,

$$\frac{\phi^0(\mathbf{x}^\gamma) - \phi^0(\mathbf{x}^0)}{\phi^0(\mathbf{x}^1) - \phi^0(\mathbf{x}^0)}$$

is the proportion of maximum cost incurred by \mathbf{x}^γ . Figure 2.7 compares these two proportions for different γ . We see that over 90% of maximum the benefit can be captured at a small

conservative level (e.g., $\gamma = 0.2$), with only 40% of the maximum cost. This shows that the DR-RUFL model is not expensive to implement in practice. Managers can assign a small weight to the worst-case expected cost and still capture most of its benefit.

It is necessary to point out that most of the observations are based on the average of numerical results with parameters shown in Table 2.4. For a given problem instance, the result will depend on the specific parameter setting and the other input data.

2.5 Extensions to Other Reliable Facility Location Problems

In this section, we show how the distributionally-robust approach can be applied to other reliable facility location problems, including the reliable p -median problem, the reliable capacitated fixed-charge location problem, and the reliable multi-allocation hub location problem.

The Reliable p -Median Problem

The reliable p -median (RPM) problem is very similar to the RUFL problem except that exactly k facilities with no fixed cost are located (since p is already used to denote the joint distribution of disruptions, we use k to denote the number of facilities to be located). The distributionally-robust RPM (DR-RPM) problem is defined as

$$(\text{DR-RPM}) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \max_{p \in \mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\},$$

where $\mathcal{X} = \{\mathbf{x} : x_j \in \{0, 1\}, \forall j \in \mathcal{J}; \sum_{j \in \mathcal{J}} x_j = k + 1\}$, and $h(x, \boldsymbol{\xi})$ is the same as in the RUFL case. Thus, Lemma 1 also applies to the distributionally-robust reliable p -median (DR-RPM) problem. Similar to the interdiction median model with fortification (e.g., Church and Scaparra 2007), the DR-RPM problem can also be embedded in a facility fortification problem. This is left as a topic for future research.

The Reliable Capacitated Fixed-Charge Location Problem

The reliable capacitated fixed-charge location (RCFL) problem generalizes the RUFL problem by assuming each location has a capacity B_j . The distributionally-robust RCFL (DR-RCFL) problem is defined as

$$(\text{DR-RCFL}) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{j \in \mathcal{J}} f_j x_j + \max_{p \in \mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\},$$

where $\mathcal{X} = \{\mathbf{x} : x_j \in \{0, 1\}, \forall j \in \mathcal{J}\}$ and

$$\begin{aligned}
 h(\mathbf{x}, \boldsymbol{\xi}) = \min & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_i c_{ij} y_{ij} \\
 \text{s.t.} & \sum_{j \in \mathcal{J}} y_{ij} = 1, \forall i \in \mathcal{I} \\
 & \sum_i d_i y_{ij} \leq x_j \xi_j B_j, \forall j \in \mathcal{J} \\
 & y_{ij} \geq 0, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}
 \end{aligned} \tag{2.6}$$

We have the following lemma.

Lemma 4. *For any $\mathbf{x} \in \mathcal{X}$, the cost function $h(\mathbf{x}, \boldsymbol{\xi})$ in (2.6) is supermodular in $\boldsymbol{\xi}$.*

Due to the capacity constraint in the DR-RCFL problem, the Benders decomposition algorithm for the DR-RUFL problem may be inefficient. A cross-decomposition algorithm can be applied.

The Reliable Multi-Allocation Hub Location Problem

In fixed-charge location and p -median problems, we are interested in the flow between facilities and customers. However, in some logistics, transportation, or telecommunication systems, it is also possible that most flows occur between pairs of customers, known as origin-destination (OD) pairs. In order to achieve economies of scale, each O-D pair is connected through one or multiple interconnection facilities, known as hubs. Hub location problems are concerned with the optimal location of such facilities. We focus on the most common case where each O-D pair is connected through no more than two hubs. There are two different cases, multi-allocation and single-allocation. In the single-allocation case, each customer is connected to a fixed hub in all O-D pairs. In the multi-allocation case, a customer can be connected to different hubs in different O-D pairs.

We focus on the fixed-charge hub location problem. The same argument holds for the p -hub median problem, where exactly k hubs are located. For ease of presentation, assume the set of candidate locations is the same as the set of customers, denoted by $\mathcal{V} = \{1, \dots, V\}$. For $i, i' \in \mathcal{V}$, let $d_{ii'}$ be the flow volume between O-D pair (i, i') . For $j \in \mathcal{V}$, $x_j = 1$ if a hub is built at node j , which incurs fixed charge f_j ; $x_j = 0$ otherwise. Let $\boldsymbol{\xi}$ be the disruption scenario vector. $\xi_j = 1$ means location j is online, and $\xi_j = 0$ means it is disrupted. For O-D pair (i, i') , let $y_{ii'jj'} = 1$ if customers i and i' are connected through hubs j and j' , which incurs unit transportation cost $c_{ii'jj'}$; $y_{ii'jj'} = 0$ otherwise. Similar to the RUFL problem, service interruption is represented by a virtual facility with index 0. It has a fixed charge $f_0 = 0$, unit transportation cost $c_{ii'00}$ equal to the service interruption penalty for O-D pair (i, i') , and $c_{ii'0j} = c_{ii'j0} = \infty$ for all $j \neq 0$. The distributionally-robust reliable

multi-allocation hub location (DR-RMHL) problem is defined as

$$(\text{DR-RMHL}) \quad \min_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{j \in \mathcal{V}} f_j x_j + \max_{p \in \mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\},$$

where $\mathcal{X} = \{\mathbf{x} : x_j \in \{0, 1\}, \forall j \in \mathcal{J}\}$, and the hub allocation problem $h(\mathbf{x}, \boldsymbol{\xi})$ is given by

$$\begin{aligned} h(\mathbf{x}, \boldsymbol{\xi}) = \min & \quad \sum_{i, i', j, j' \in \mathcal{V}} d_{ii'} c_{ii'jj'} y_{ii'jj'} \\ \text{s.t.} & \quad \sum_{j, j' \in \mathcal{V}} y_{ii'jj'} = 1, \quad \forall i, i' \in \mathcal{V} \\ & \quad y_{ii'jj} + \sum_{j' \neq j} (y_{ii'jj'} + y_{ii'j'j}) \leq x_j \xi_j, \quad \forall i, i', j \in \mathcal{V} \\ & \quad y_{ii'jj'} \geq 0, \quad \forall i, i', j, j' \in \mathcal{V} \end{aligned} \tag{2.7}$$

We have the following lemma.

Lemma 5. *For any $\mathbf{x} \in \mathcal{X}$, the cost function $h(\mathbf{x}, S)$ in (2.7) is supermodular in S .*

Given supermodularity, the worst-case distribution result in Lemma 2 will still hold for the aforementioned reliable facility location problems. However, several other reliable facility location problems, including the reliable p -center problem and the reliable single-allocation hub location problem, are not supermodular. Thus, the distribution given in Lemma 2 will not be the worst-case distribution for these problems. However, it is worth mentioning that these problems are not submodular either. For these problems, ignoring disruption correlation can still result in significant loss, and the distribution given in Lemma 2 could be used as a better approximation than the independent distribution.

2.6 Summary and Future Directions

In this chapter, we present a distributionally-robust optimization model to incorporate correlated disruptions in reliable facility location design. We find that this seemingly complicated problem is actually equivalent to a much simpler problem and can be solved efficiently. Our numerical results show that this model has several advantages compared to the traditional model, which is based on the assumption of independent disruptions, and thus we believe it can serve as a promising alternative approach for reliable facility location design problems.

One limitation of our model is that it focuses on the worst-case distribution, which can be overly conservative in practice. In our future work, we plan to study a more general model where the disruption correlation is explicitly given. Also, we focus on locating facilities for regular supply chain operations and assume the demand is deterministic and not affected by the disruptions. When locating facilities for humanitarian operations in disaster

relief, the demand will be highly uncertain and depend on the disruptions. We will incorporate uncertain demand in our future work. Our model is suitable for location design in a medium-to-large area (e.g., nation-wide). New models need to be developed for problems in a relatively small area (e.g., a city). We also plan to consider the facility fortification problem under correlated disruptions and study the impact of correlation on the effect of fortification.

Chapter 3

Joint Pricing-Inventory Management with Structural Uncertainty

3.1 Introduction

The interaction between price and demand has been one of the central topics studied in business. Pricing strategies can affect customer demand effectively. Inventory decisions affect the cost to satisfy the customer demand, and thus also the effectiveness of pricing strategies (Yano and Gilbert, 2004). In order to maximize the expected profit, pricing and inventory decisions need to be jointly optimized. Coordinated or joint pricing-inventory management considers the important interface between marketing and operations management. It has been widely studied by the operations research and management sciences community (see Elmaghraby and Keskinocak, 2003; Chan et al., 2004; Yano and Gilbert, 2004; Chen and Simchi-Levi, 2012). The most basic version of joint pricing-inventory management is the pricing newsvendor problem, where a retailer jointly determines the selling price and the order quantity of a perishable product. Despite its simplified setting, the pricing newsvendor problem is the backbone of many joint-pricing inventory management models, and has been shown to have significant theoretical and practical value (see Petruzzi and Dada, 1999; Raz and Porteus, 2006; Lu and Simchi-Levi, 2013).

In joint pricing-inventory management, the most important and also the most difficult task is to model the relationship between price and demand. In most of the existing literature, the relationship is modeled using the demand function, i.e., the demand is a function $D(p, \varepsilon)$ of the selling price p and a random factor ε with distribution function F_ε . Commonly used demand function forms include

- Additive model, $D(p, \varepsilon) = d(p) + \varepsilon$,
- Multiplicative model, $D(p, \varepsilon) = \varepsilon d(p)$,
- Additive-multiplicative model, $D(p, \varepsilon) = \mu(p) + \varepsilon \sigma(p)$.

The forms of $d(p)$ or $\mu(p)$ and $\sigma(p)$ also need to be specified. Some commonly used forms include

- Linear function, $d(p) = a - bp$, $a, b > 0$,
- Exponential function, $d(p) = ae^{-bp}$, $a, b > 0$,
- Iso-elastic function, $d(p) = ap^{-b}$, $a > 0$, $b > 1$.

For the distribution function F_ε , one can choose from all the commonly used distributions in operations research and management sciences, or any distribution that is suitable for the problem.

As we have seen, there are a large number of candidate demand models. The structural properties of the problem and the optimal solution depend on the selection of the demand model. For example, it has been shown that the additive model and multiplicative model cause the optimal price to move in opposite directions compared with the optimal riskless price (i.e., the price that maximizes the expected revenue) (e.g., Mills, 1959; Karlin and Carr, 1962; Petruzzi and Dada, 1999; Salinger and Ampudia, 2011). Furthermore, most of the existing models require conditions on the demand function $D(p, \varepsilon)$ and the distribution function $F(\varepsilon)$, (or equivalently, the conditional distribution function of price $F(p, x)$, where $F(p, x) = \mathbb{P}\{D(p, \varepsilon) \leq x\}$), in order to guarantee that the expected profit function is unimodal or quasi-concave (e.g., Federgruen and Heching, 1999; Yao, Chen, and Yan, 2006; Kocabiyıkoğlu and Popescu, 2011; Lu and Simchi-Levi, 2013; Roels, 2013).

In practice, the functional forms of $D(p, \varepsilon)$ and $d(p)$, along with the distribution of the random factor, are typically unknown. They need to be estimated from historical price and demand observations with properly selected parametric forms. Although several recent studies have considered more general functional forms and less restrictive required conditions, the structural uncertainty in the relationship between price and demand could still undermine the applicability of joint pricing-inventory management models in several ways. First, selecting and fitting models can be a hard problem in itself. Managers who make the pricing and inventory decisions usually do not have sufficient training and technical support to perform this task satisfactorily. Furthermore, whenever a new data set is used, the model selection and fitting process needs to be repeated again. This can create a large amount of work for companies with a large number of products and markets, or those in a rapidly changing business environment. Second, even if the models are properly selected and fitted, the estimated demand model may not satisfy the conditions required by many of the classical joint pricing-inventory models. For example, one classical result for the pricing-newsvendor problem requires that the mean demand function have increasing price elasticity, and the distribution of the random factor have a generalized increasing failure rate (Yao, Chen, and Yan, 2006). These conditions are non-trivial, and may be violated by the estimates from historical observations. Third, if an improper parametric form for the demand model is selected, or an overly simplified assumption is made, decisions from the misspecified model can

be highly suboptimal. As a result, there can be significant loss or regret (Raz and Porteus, 2006).

In this chapter, we present a data-driven approach for the pricing-newsvendor problem to address the structural uncertainty in the relationship between price and demand. The approach does not assume any parametric demand model. Instead, it only requires information which is typically available in practice, i.e., historical price and demand observations, and the basic domain knowledge in pricing, known as the Law of Demand. The conditional demand distribution is estimated using nonparametric quantile regression with monotone shape constraints. Parametric programming is used to efficiently estimate the entire quantile path. The optimal price and order quantity are found numerically using the estimated demand quantiles. Smoothing and kernelization methods are applied to avoid overfitting and improve the quality of estimates and decisions. Additional domain knowledge, such as demand concavity with respect to price, can also be incorporated in the model. Numerical results show that the data-driven approach is able to find close-to-optimal solutions. Smoothing, kernelization, and the incorporation of additional domain knowledge can significantly improve the performance of the approach.

The rest of this chapter is organized as follows. Section 3.2 reviews the related literature. Section 3.3 presents the data-driven approach using isotonic quantile regression. Section 3.4 discusses how the approach can be improved by smoothing, kernelization, and incorporating additional domain knowledge. Section 3.5 analyzes the performance of the approach using numerical experiments. Section 3.6 summarizes the results and discusses directions for future research.

3.2 Literature Review

Literature on coordinated or joint pricing-inventory management abound. Elmaghraby and Keskinocak (2003), Chan et al. (2004), Yano and Gilbert (2004), and Chen and Simchi-Levi (2012) provided excellent comprehensive surveys on the related literature. We focus on the single period pricing newsvendor problem. Whitin (1955) was the first to study this problem. He used the conditional demand distribution to model the dependence of demand on price. Later, Mills (1959) studied the case of additive demand model, and Karlin and Carr (1962) studied the case of multiplicative demand model. Young (1978) was the first to study the more general additive-multiplicative demand model. Petruzzi and Dada (1999) presented a unified framework for additive and multiplicative demand models. Most of the later studies focus on two research questions. First, what conditions are sufficient to guarantee the unimodality or quasi-concavity of the expected profit function? Second, how does demand uncertainty affect the optimal price? In other words, how does the optimal price for the pricing newsvendor problem differ from the riskless optimal price, i.e., the optimal price under deterministic demand?

For the first question, Yao, Chen, and Yan (2006) provided an excellent survey of the demand models used in previous literature and the corresponding sufficient conditions. They

showed that to guarantee unimodality of the expected profit function, it is sufficient for the mean demand function to have increasing price elasticity, and for the distribution of the random factor to have a generalized increasing failure rate. These conditions are shown to be more general than most of those previously used in the literature. Xu, Cai, and Chen (2011) provided more details on this result, and showed several applications. Lu and Simchi-Levi (2013) considered the more general additive-multiplicative demand model. They used three conditions, namely, the log-convexity of the coefficient of variation, the log-concavity of the deterministic profit function, and the log-convexity of the random factor's expectation conditioning on having leftover inventory, to establish the log-concavity of the expected profit function. Roels (2013) derived another set of sufficient conditions, which is shown to be complementary to the set of conditions in Lu and Simchi-Levi (2013), i.e., neither set of conditions implies the other. Kocabiyikoğlu and Popescu (2011) took a different approach by considering the conditional distribution function directly. They proposed a new measure, called the lost-sales rate (LSR) elasticity, and showed that the structural properties of the problem can be fully characterized by conditions regarding the LSR elasticity.

For the second question, it is well known that additive and multiplicative demand models cause the optimal price to move in opposite directions compared with the riskless optimal price. Let p^* denote the optimal price for the pricing newsvendor problem, and p^0 the riskless optimal price. Mills (1959) found that $p^* \leq p^0$ for additive demand, and Karlin and Carr (1962) found that $p^* \geq p^0$ for multiplicative demand. Petruzzi and Dada (1999) summarized the results, and argued that the difference can be explained by the monotonicity of variance and coefficient of variation of demand. Salinger and Ampudia (2011) explained the difference using simple economics regarding the marginal cost of each expected unit sold and the price elasticity of expected sales. Young (1978) studied the effect of the additive-multiplicative demand model on the relationship between p^* and p^0 , and derived conditions regarding the conditional mean and variance, which can be used to determine the relationship. Lu and Simchi-Levi (2013) and Roels (2013) also studied the effect of the additive-multiplicative demand model on the relationship between p^* and p^0 . They showed that the relationship can be characterized using the sufficient conditions that they derived to guarantee the unimodality or quasi-concavity of the expected profit function. Kocabiyikoğlu and Popescu (2011) studied the relationship between p^* and p^0 using the conditional demand distribution function, and showed that the relationship can be characterized by conditions regarding the LSR elasticity.

There have also been various extensions to the pricing newsvendor problem. Aydin and Porteus (2008) studied joint pricing-inventory management for an assortment. Murray, Gosavi, and Talukdar (2012) studied the multi-product newsvendor problem with a shared resource constraint. Chen, Xu, and Zhang (2009) and Arcelus, Kumar, and Srinivasan (2012) considered the pricing newsvendor problem under risk-aversion. Yang, Shi, and Zhao (2011) considered the pricing newsvendor problem where the decision maker seeks to maximize the probability that a profit target is achieved. Xu, Chen, and Xu (2010) and Xu and Lu (2013) studied the impact of demand and supply uncertainty, respectively, in the pricing newsvendor problem.

All of the aforementioned papers are based on the classical demand models, i.e., the demand function or the conditional demand distribution function. An exception is the paper by Raz and Porteus (2006), who provided a fractile perspective to the pricing newsvendor problem. (We use the term “quantile” instead of “fractile” to be consistent with the related statistics literature.) They assumed that the conditional demand distribution is represented by a finite number of linear or piecewise linear quantile functions. For the linear case, they characterized the optimal pricing and inventory decisions. For the general piecewise linear case, they developed an exact approach for finding the optimal solutions numerically. Using the quantile representation, they were able to identify effects that are not usually seen when using the classical demand models. For example, the optimal price may be nonmonotone in the ordering cost. They also found that using a simplified demand model can result in substantially lower expected profit.

The quantile-based approach of Raz and Porteus (2006) greatly enhances the applicability of the pricing newsvendor model. It is also closely related to the data-driven approach in this chapter. However, we notice that their approach still has several limitations. First, it is not clear how the finite number of representative quantile levels should be selected and how the corresponding quantile functions can be estimated from data. Second, the exact approach is suitable for piecewise linear quantiles with a relatively small number of breakpoints. It is not clear how these breakpoints should be determined, or how well these piecewise linear quantiles can approximate the conditional distribution. Furthermore, in practice, there may be a large number of observations with prices between the breakpoints. Those observations are not utilized effectively. Our approach is fully data-driven in that it does not require the predetermined representative quantile levels or breakpoints. A nonparametric quantile estimate is used instead of the piecewise-linear structure. This allows us to consider a large number of unique prices, and the observations are utilized more effectively.

All of the aforementioned papers assume that the demand information is given. Very few have considered how to obtain the required information from data. Feng, Luo, and Zhang (2014) studied demand estimation for the additive-multiplicative model in multi-period dynamic pricing-inventory problem. The generalized additive model (GAM) is applied to estimate the demand function. They derived sufficient conditions under which a base stock list price policy is optimal. They also developed a constrained maximum likelihood estimation approach to obtain estimates that satisfy these conditions. The GAM approach of Feng, Luo, and Zhang (2014) makes a big step towards applying joint pricing-inventory models with additive-multiplicative demand in practice. However, it is still a parametric approach.

To the best of our knowledge, Burnetas and Smith (2000) presented the only fully data-driven approach for the pricing newsvendor problem in the existing literature. The pricing problem is modeled as a multi-armed bandit problem. For a given price, the optimal inventory level is found using stochastic approximation. The main difference between their approach and ours is that in Burnetas and Smith (2000) one needs to consider a relatively small number of feasible price decisions. For each price decision, a sufficient demand sample size is required. Our approach can handle effectively the case where there are a large

number of unique prices and very few demand observations are available for each price. Our approach also makes use of domain knowledge such as the Law of Demand. This introduces interdependence between demand observations at different prices. Thus, when estimating the demand at one specific price, information contained in all the observations is utilized effectively. In Burnetas and Smith (2000), demand observations with different prices are treated separately.

Another related work is Chu, Shanthikumar, and Shen (2009). They considered a special case of the multiplicative demand model with an exponential mean demand function and exponentially distributed random factors. The mean of the random factor and the parameter of the mean demand function are unknown, and need to be estimated from historical price-demand observations. They applied a data-driven approach for parameter uncertainty, called operational statistics, to integrate parameter estimation and optimization. Parameter uncertainty and operational statistics will be discussed in more detail in Chapter 4. For the pricing newsvendor problem, our data-driven approach applies to more general cases where the demand model does not have a known parametric form.

Data-driven methods have also been applied to inventory management and pricing, separately. For inventory management, the optimal inventory level is found by estimating the demand quantile that corresponds to the optimal newsvendor critical ratio. Papers based on this approach includes Levi, Roundy, and Shmoys (2007), Levi, Perakis, and Uichanco (2012), Huh and Rusmevichientong (2009), Huh et al. (2011), and Jain et al. (2011). More details of these papers are reviewed in Chapter 4. All of these papers considered inventory decisions solely, while we consider price-dependent demand and jointly optimize pricing and inventory decisions. For revenue management, Besbes and Zeevi (2009) developed a sampling-based algorithm for dynamic pricing. Their approach is based on the closed-form solution in Gallego and van Ryzin (1994) for the same problem with complete information. The entire selling horizon is divided into a learning phase and an optimization phase. In the learning phase the empirical demand function is estimated. In the optimization phase, the estimated demand function is used in the closed-form result in Gallego and van Ryzin (1994). Wang, Deng, and Ye (2011) improved the algorithm in Besbes and Zeevi (2009) by iterating between learning and optimization. Besbes and Zeevi (2012) developed a sampling-based algorithm for the network revenue management problem. These papers considered pricing decisions solely, but did not consider the effect of inventory decisions on the effectiveness of pricing.

3.3 A Data-Driven Approach Based on Isotonic Quantile Regression

In this section, we present a data-driven approach for the pricing-newsvendor problem. We adopt the repeated newsvendor setting, which is suitable for perishable products. At the beginning of each period, the products are ordered with unit ordering cost c . At the end of

each period, all remaining products are salvaged, i.e., no inventory is carried over. Without loss of generality, assume there is no salvage value. The customer demand is a function $D(p, \varepsilon)$ of the selling price p and a random factor ε with distribution function F_ε . Given p and order quantity y , the newsvendor achieves a random profit

$$\Phi(p, y, \varepsilon) = p \min\{y, D(p, \varepsilon)\} - cy.$$

The newsvendor jointly optimizes p and y to maximize the expected profit

$$\phi(p, y) = \int \Phi(p, y, x) dF_\varepsilon(x).$$

As we mentioned, in practice, the demand function $D(p, \varepsilon)$ and the distribution function F_ε are typically unknown or uncertain. We assume historical price-demand observations in the previous n periods, denoted by $(p_1, d_1), (p_2, d_2), \dots, (p_n, d_n)$, are available. In addition to the historical observations, we also assume the Law of Demand holds, which is the basic domain knowledge in pricing. The Law of Demand states that, holding all else constant, the demand is decreasing (non-increasing) in the price in the usual stochastic order, i.e.,

$$d_i \preceq_{st} d_j, \text{ if } p_i > p_j.$$

The Law of Demand is widely observed in business. There are exceptions where raising the price may increase the demand. Those special circumstances are not considered in this chapter.

Our data-driven approach is based only on the information available in practice, i.e., the historical price-demand observations and the Law of Demand. The approach consists of two sequential stages. In the first stage, the conditional quantile functions of the demand are estimated using nonparametric regression. In the second stage, pricing and inventory decisions are made based on the estimated quantiles. For a given price p , the optimal inventory level is given by the conditional quantile corresponding to the critical ratio $1 - c/p$. The expected profit can also be estimated using the quantiles. The optimal price can then be determined numerically as the one with the highest estimated expected profit.

Isotonic Quantile Regression

The τ -quantile of a random variable X is defined as

$$q^\tau(X) = \min\{q : \mathbb{P}\{X \leq q\} \geq \tau\}.$$

For the standard newsvendor problem, it is well known that the optimal inventory level is the τ^* -quantile of the demand distribution, where the τ^* is the critical ratio

$$\tau^* = 1 - \frac{c}{s}.$$

Given historical demand data, the optimal inventory level can be found by one-sample quantile estimation. This approach has been adopted by Levi, Roundy, and Shmoys (2007); Levi, Perakis, and Uichanco (2012).

For the pricing newsvendor problem, with a given price p , the optimal inventory level is the τ^* -quantile of conditional demand distribution. If the demand function has a known parametric form, for example,

$$D(p, \varepsilon) = \alpha + \beta p + \varepsilon, \quad (3.1)$$

we can apply quantile regression (Koenker and Bassett, 1978) to estimate the conditional quantile function of the demand. Quantile regression is analogous to ordinary least square (OLS) regression. In OLS regression, the conditional mean is estimated by minimizing the sum of squared errors. In quantile regression, the conditional quantile is estimated by minimizing a different empirical risk. In univariate linear quantile regression, we solve the following optimization problem,

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmin} \left\{ \sum_{i=1}^n \rho^\tau(d_i - \alpha - \beta p_i) \right\},$$

where $\rho^\tau(d_i - \alpha - \beta p_i)$ is the weighted absolute error, with

$$\rho^\tau(a) = (\tau - I(a < 0))a.$$

The conditional τ -quantile can then be estimated as

$$\hat{q}^\tau(p) = \hat{\alpha} + \hat{\beta}p.$$

However, this approach can only be used when the demand function $D(p, \varepsilon)$ has a known parametric form, e.g., the linear function in (3.1). When $D(p, \varepsilon)$ is unknown, using a misspecified demand model could result in highly suboptimal estimates.

To address the uncertainty in $D(p, \varepsilon)$, we can apply a nonparametric method called isotonic regression. Recall that the Law of Demand states that the demand is decreasing in the price in the usual stochastic order. This means the conditional quantile function is monotone decreasing. Let q_i be the estimate for the conditional quantile given price p_i . We have

$$q_i \leq q_j, \quad \forall i, j \text{ such that } p_i > p_j.$$

This means that the quantile estimate $\mathbf{q} = (q_1, q_2, \dots, q_n)$ is an isotonic vector, i.e., it preserves the partial order defined by $p_i > p_j$. The optimal isotonic vector estimate for the conditional quantile function is found by solving the following problem to minimize the empirical risk

$$\min_{\mathbf{q}} \left\{ \sum_{i=1}^n \rho^\tau(d_i - q_i) : \mathbf{q} \text{ isotonic} \right\}.$$

The least squares version of isotonic regression, i.e., isotonic mean regression, is widely applied in statistics and engineering. Please refer to Barlow et al. (1972) for more details. Isotonic median (i.e., 0.5-quantile) regression was studied by Cryer et al. (1972).

Without loss of generality, assume that the price-demand observations are sorted in increasing order of the price, i.e., $p_1 < p_2 < \dots < p_n$. It is easy to generalize the approach to the case where multiple observations have the same price. Define auxiliary variables u_i and v_i to represent the positive and negative errors, respectively, i.e.,

$$\begin{aligned} u_i &= \max\{d_i - q_i, 0\}, \\ v_i &= \max\{q_i - d_i, 0\}. \end{aligned}$$

Also define $q_{n+1} = 0$. We have the following linear program formulation for isotonic quantile regression

$$\begin{aligned} \min \quad & \sum_{i=1}^n \tau u_i + (1 - \tau)v_i \\ \text{s.t.} \quad & q_i + u_i - v_i = d_i, \quad \forall i \\ & q_i \geq q_{i+1}, \quad \forall i \\ & u_i, v_i \geq 0, \quad \forall i \end{aligned}$$

Throughout this chapter, we may sometimes use boldface lowercase letters to denote vectors without further declaration, e.g., $\mathbf{u} = (u_1, \dots, u_n)$. We can then write the above linear program in a more compact form.

$$\begin{aligned} \min \quad & \tau \mathbf{1}'\mathbf{u} + (1 - \tau)\mathbf{1}'\mathbf{v} \\ \text{s.t.} \quad & \mathbf{q} + \mathbf{u} - \mathbf{v} = \mathbf{d} \\ & \mathbf{U}\mathbf{q} \geq \mathbf{0} \\ & \mathbf{u}, \mathbf{v} \geq \mathbf{0} \end{aligned} \tag{3.2}$$

where $\mathbf{0}$ and $\mathbf{1}$ are vectors of proper dimension, with all elements equal to 0 or 1, respectively, and \mathbf{U} is an $n \times n$ upper bi-diagonal matrix, with all entries equal to 1 on the main diagonal, and all entries equal to -1 on the first upper diagonal, i.e.,

$$\mathbf{U} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & -1 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

Figure 3.1 shows several examples of isotonic median regression (i.e., 0.5-quantile regression) with different numbers of observations. The line in dark color shows the estimated median. The line in light color shows the actual median. We see that as the number observations increases, the estimate becomes closer to the actual median. Consistency of isotonic median regression was shown in Cryer et al. (1972).

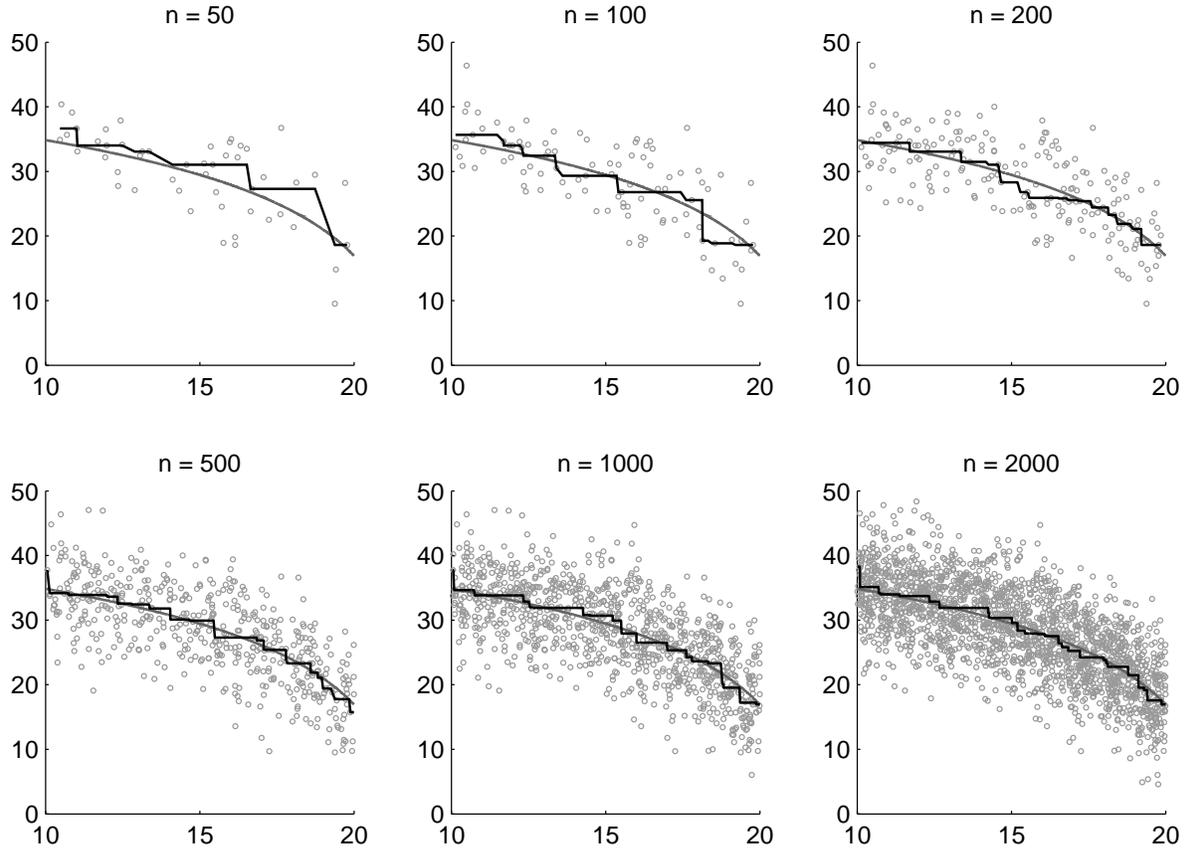


Figure 3.1: Examples of isotonic quantile regression estimates

Note: x-axis: price, y-axis: demand, n : number of observations.

Estimation of Full Quantile Path using Parametric Programming

For a given quantile level τ , the linear program in (3.2) is easy to solve. However, in order to solve the pricing-newsvendor problem, we need to estimate the full quantile path, i.e., we need to estimate the conditional quantile function for any given quantile level (technically, any given quantile level below the critical ratio $1 - c/\bar{p}$, where \bar{p} is the highest price). We can certainly focus on several prespecified quantile levels (as in Raz and Porteus, 2006), and estimate the conditional quantile function for each of the levels separately. However, it is not clear how many unique quantiles are necessary and/or sufficient and how the quantile levels should be determined.

To address this issue, we use parametric programming to efficiently compute the full quantile path. Note that the linear program in (3.2) is parameterized by the quantile level τ . Starting from the trivial case where $\tau = 0$, we find the optimal quantile estimate for the current τ . Then, based on the simplex pivoting rule, we find the next breakpoint for τ ,

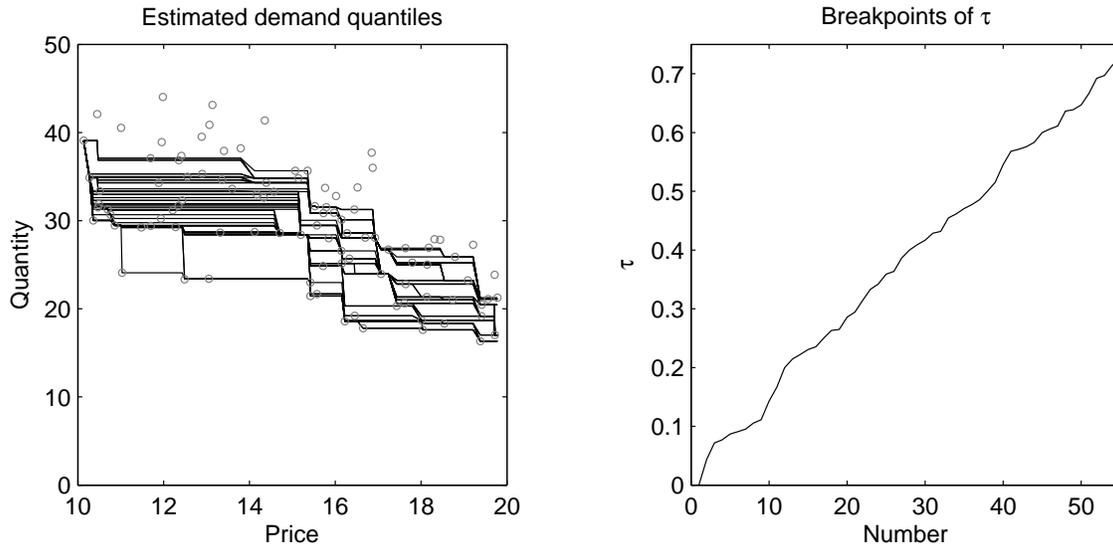


Figure 3.2: Quantile path and breakpoints in parametric programming

i.e., the largest τ such that the current optimal basis/solution remains optimal. τ is then increased to its next breakpoint, and the linear program is re-solved to optimality from the current basis. This process continues until τ reaches the highest necessary level. Since this method is well known for linear programming, we do not include the details of the algorithm. Interested readers may refer to popular textbooks on linear programming such as Bertsimas and Tsitsiklis (1997). Figure 3.2 shows an example of the full quantile path estimated using parametric programming and the breakpoints for τ .

One important question is how many breakpoints will be encountered in the parametric programming process. The number of breakpoints affects the computational efficiency. Let T_n be the number of breakpoints given n observations. For linear quantile regression, Portnoy (1991) showed that

$$T_n = O_p(n \log n),$$

where O_p is the big O in probability notation, i.e., for any $\epsilon > 0$, there exists a finite $M > 0$ such that

$$\mathbb{P} \left\{ \frac{T_n}{n \log n} > M \right\} < \epsilon.$$

We conjecture that for isotonic regression, the number of breakpoints is at most $O_p(n \log n)$. In fact, numerical results suggest that the number of breakpoints is $O_p(n)$. Figure 3.3 shows a numerical example. The solid lines show the number of breakpoints T_n as we increase the number of observations n . The dashed line shows n . We see that T_n is bounded by n in all of the sample paths.

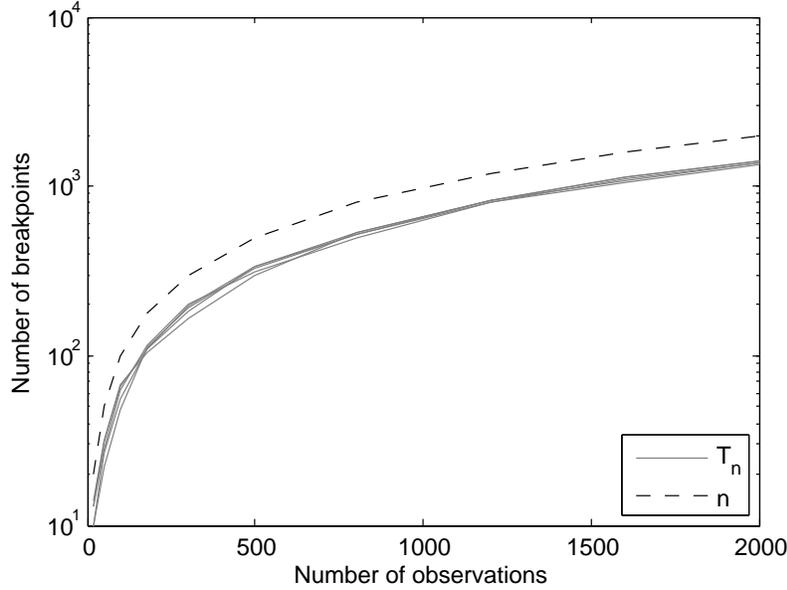


Figure 3.3: Number of breakpoints in isotonic quantile regression

Inventory and Pricing Decisions

Technically speaking, given the estimated demand quantile functions, the optimal inventory and pricing decisions can be obtained using the exact approach described in Raz and Porteus (2006). However, when there are many unique prices, the exact approach is computationally intractable. Instead, we focus on a discrete subset of feasible prices $\mathcal{P} = \{p_i, i = 1, \dots, m\}$. A simple choice for \mathcal{P} is the set of historical prices.

For a given price p , suppose there are k breakpoints for τ , denoted by τ_1, \dots, τ_k , that are smaller than or equal to the critical ratio $1 - c/p$. Let $\hat{q}^j(p)$ denote the demand quantile estimate with quantile level τ_j . The optimal order quantity for price p is given by the conditional quantile with level $1 - c/p$. Thus, it can be estimated as

$$\hat{y}(p) = \hat{q}^k(p).$$

Define $\tau_{k+1} = 1 - c/p$. Let $\varphi(p)$ denote the optimal expected profit of price p , i.e., $\varphi(p) = \phi(p, y^*(p))$, where $y^*(p)$ is the optimal inventory level for price p . $\varphi(p)$ can be estimated as

$$\hat{\varphi}(p) = p \sum_{j=1}^k (\tau_{j+1} - \tau_j) \hat{q}^j(p).$$

The optimal price in \mathcal{P} can then be found by searching over the feasible prices, i.e.,

$$\hat{p} = \operatorname{argmax}_{p \in \mathcal{P}} \{\hat{\varphi}(p)\}.$$

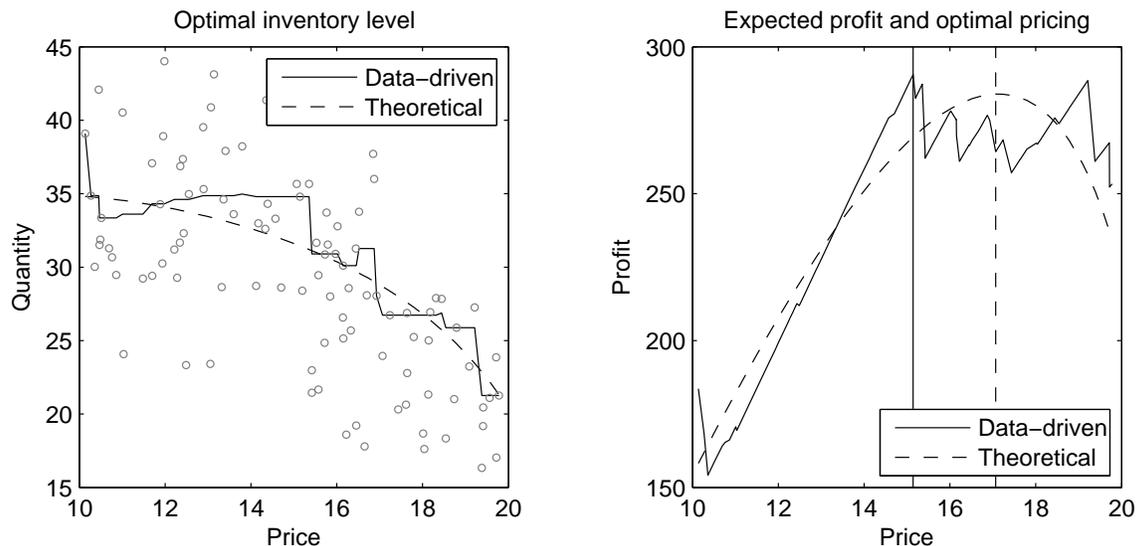


Figure 3.4: Estimated optimal inventory level, expected profit function, and optimal price

Figure 3.4 shows the estimated optimal inventory level, expected profit function, and optimal price, for the same numerical example as we used for Figure 3.2. We see that although the data-driven approach was able to find a solution that is close to the theoretical optimal solution, its performance is not satisfactory. This is because the nonparametric method in (3.2) is not regularized and leads to overfitting, i.e., the estimates and the decisions are overly sensitive to the observations. As a result, the problem is ill-posed, which can be seen from the estimated expected profit function. Small changes in the observations may cause significant changes in the decisions. We can prevent overfitting using regularization. The performance of the data-driven approach can thus be improved.

3.4 Smoothing, Concavity Constraint and Kernelization

In this section, we show how the basic data-driven approach in Section 3.3 can be improved through regularization and incorporating additional domain knowledge. We consider two regularization methods, namely, smoothing and kernelization. In some cases, we may have additional domain knowledge other than the Law of Demand. We consider one such case where the demand function is known to be concave. Later in Section 3.5, we use numerical results to show that regularization and additional domain knowledge can significantly improve the quality of the estimates and decisions.

Smoothed Demand Quantile Estimate

Isotonic quantile regression in (3.2) can be regularized by smoothing, i.e., penalizing the roughness of the estimated quantile function. One measure of roughness is the total variation of the (piecewise) derivative. Recall that the estimated quantile function is a piecewise linear function. The slope of the segment between (p_i, q_i) and (p_{i+1}, q_{i+1}) is given by

$$\frac{q_{i+1} - q_i}{p_{i+1} - p_i}.$$

Define $h_i = (p_{i+1} - p_i)^{-1}$. The total variation of the piecewise derivative is equal to the sum of the absolute differences between the slopes of consecutive linear segments, i.e.,

$$\begin{aligned} TV(\mathbf{q}) &= \sum_{i=1}^{n-2} |h_i(q_{i+1} - q_i) - h_{i+1}(q_{i+2} - q_{i+1})| \\ &= \sum_{i=1}^{n-2} |-h_i q_i + (h_i + h_{i+1})q_{i+1} - h_{i+1}q_{i+2}|. \end{aligned} \quad (3.3)$$

We can write $TV(\mathbf{q})$ in matrix form as

$$TV(\mathbf{q}) = \|\mathbf{K}\mathbf{q}\|_1,$$

where $\|\cdot\|_1$ is the \mathcal{L}_1 norm, and \mathbf{K} is an $(n-2) \times n$ matrix,

$$\mathbf{K} = \begin{bmatrix} -h_1 & h_1 + h_2 & -h_2 & 0 & \cdots & 0 \\ 0 & -h_2 & h_2 + h_3 & -h_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -h_{n-2} & h_{n-2} + h_{n-1} & -h_{n-1} \end{bmatrix}. \quad (3.4)$$

Let λ be the smoothing factor, i.e. the weight for the roughness penalty. The smoothed isotonic quantile estimate is then found by solving the following linear program,

$$\begin{aligned} \min \quad & \tau \mathbf{1}'\mathbf{u} + (1 - \tau)\mathbf{1}'\mathbf{v} + \lambda \|\mathbf{K}\mathbf{q}\|_1 \\ \text{s.t.} \quad & \mathbf{q} + \mathbf{u} - \mathbf{v} = \mathbf{d} \\ & \mathbf{U}\mathbf{q} \geq \mathbf{0} \\ & \mathbf{u}, \mathbf{v} \geq \mathbf{0} \end{aligned} \quad (3.5)$$

We note that smoothed isotonic quantile regression is equivalent to estimating the demand quantiles using quantile smoothing splines with shape constraints (Koenker, Ng, and Portnoy, 1994). The factor λ measures the strength of smoothing. When λ is sufficiently large, smoothed isotonic quantile regression reduces to linear quantile regression.

For a given λ , problem (3.5) is parameterized by τ . Thus, we can use the same parametric programming approach for problem (3.2) to efficiently compute the full quantile path.

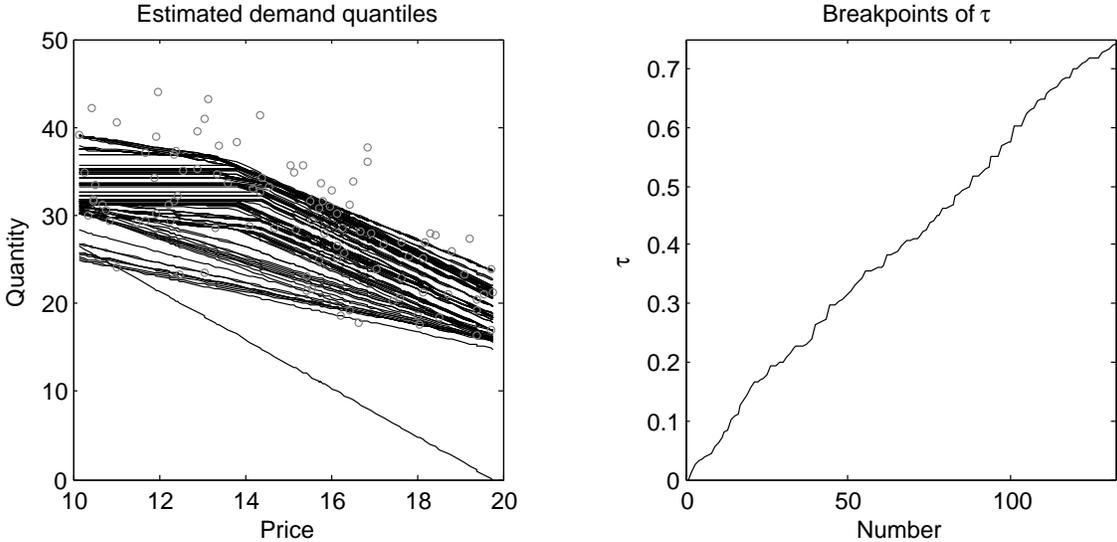


Figure 3.5: Quantile path and breakpoints using smoothing

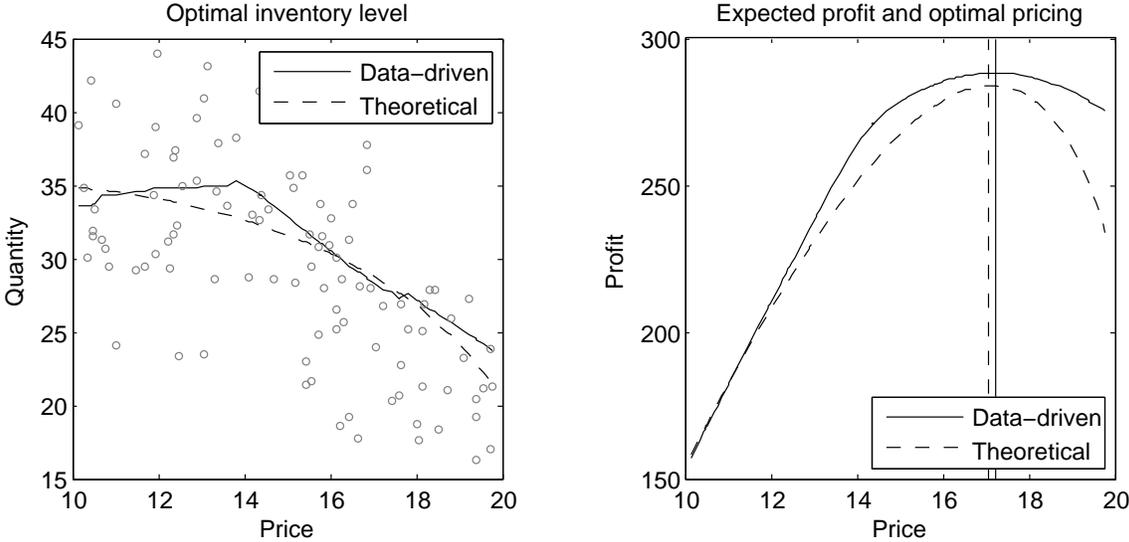


Figure 3.6: Improved data-driven solutions using smoothing

Figures 3.5 and 3.6 show the estimated demand quantiles and decisions using smoothing for the same example we used for Figures 3.2 and 3.4. Comparing these results with those without smoothing, we see that the use of smoothing improved the quality of the estimates and the decisions. (Both the estimated expected profit and the optimal price are closer to the theoretical values.) More numerical results will be presented in Section 3.5.

The performance of smoothing depends largely on the selection of λ , known as nonparametric model selection. Model selection is a very important topic in statistics. Numerous model selection methods have been proposed and studied. In this chapter we select the smoothing factor using cross validation. In cross validation, the observations are split into a training set and a validation set. For a given λ , the optimal estimate is obtained using the training set. Then, the estimate is applied to the validation set, and the empirical risk, known as the cross validation risk, is calculated. Different λ values are tested, and the one with the lowest cross validation risk is selected.

The cross validation procedure described above is standard in statistics. However, for isotonic quantile regression, this procedure can be improved. Note that for a given τ , problem (3.5) is parameterized by λ . Thus, we can use parametric programming to efficiently compute the smoothed quantile estimate for all λ values. Recall that when λ is sufficiently large, smoothed isotonic quantile regression will reduce to linear quantile regression. We can start from this trivial case with $\lambda = \infty$. We then find the next breakpoint for λ , i.e., the smallest λ such that the current solution/basis remains optimal. λ is then decreased to the next breakpoint, and the linear program is re-solved to optimality from the current basis. This process continues until λ is reduced to 0. This parametric approach was also used in Koenker, Ng, and Portnoy (1994), but they selected the parameter using the Schwarz Information Criterion (Schwarz, 1978). In order to select a λ that performs well for the full quantile path, we repeat this procedure for several prespecified τ levels, and use the average cross validation risk to determine the optimal λ . Figure 3.7 shows an example of the average cross validation risk for all possible λ values.

Incorporation of Demand Concavity

Under certain circumstances, we may have additional domain knowledge other than the Law of Demand. For example, in pricing, it may be reasonable to assume the demand function $D(p, \varepsilon)$ is concave in p for any ε (see Federgruen and Heching, 1999). Demand concavity (or convexity) can be easily incorporated in the isotonic quantile regression model. Recall that the estimated quantile function is a piecewise linear function. The differences between the slopes of consecutive linear segments are given by the vector \mathbf{Kq} . Concavity of the demand function is equivalent to having quantile functions with monotone decreasing slopes, i.e., the differences \mathbf{Kq} are non-negative. Thus, we can add the constraints $\mathbf{Kq} \geq 0$ to (3.2) to guarantee that the estimated quantile function is concave. We have the following linear

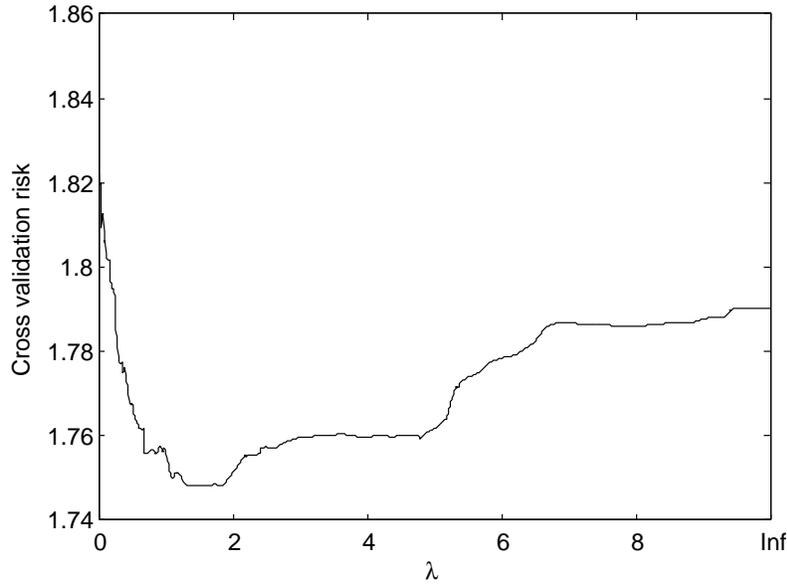


Figure 3.7: Cross validation risk of different smoothing parameters

program

$$\begin{aligned}
 & \min \tau \mathbf{1}' \mathbf{u} + (1 - \tau) \mathbf{1}' \mathbf{v} \\
 & \text{s.t. } \mathbf{q} + \mathbf{u} - \mathbf{v} = \mathbf{d} \\
 & \quad \mathbf{U} \mathbf{q} \geq \mathbf{0} \\
 & \quad \mathbf{K} \mathbf{q} \geq \mathbf{0} \\
 & \quad \mathbf{u}, \mathbf{v} \geq \mathbf{0}
 \end{aligned} \tag{3.6}$$

Similarly, convex demand information can be incorporated by adding the constraints $\mathbf{K} \mathbf{q} \leq \mathbf{0}$ to the linear program in (3.2).

Note that incorporating demand concavity achieves an effect that is similar to smoothing. To see this, consider the Lagrangian dual of problem (3.6) when the concavity constraints $\mathbf{K} \mathbf{q} \geq \mathbf{0}$ are relaxed with Lagrangian multipliers $\mathbf{w} = (w_1, \dots, w_{n-2})$. By duality theory, there exist multipliers $\bar{\mathbf{w}} \geq \mathbf{0}$, such that problem (3.6) is equivalent to the following problem,

$$\begin{aligned}
 & \min \tau \mathbf{1}' \mathbf{u} + (1 - \tau) \mathbf{1}' \mathbf{v} + \bar{\mathbf{w}}' \mathbf{K} \mathbf{q} \\
 & \text{s.t. } \mathbf{q} + \mathbf{u} - \mathbf{v} = \mathbf{d} \\
 & \quad \mathbf{U} \mathbf{q} \geq \mathbf{0} \\
 & \quad \mathbf{u}, \mathbf{v} \geq \mathbf{0}
 \end{aligned}$$

The optimal solution to the above problem also satisfies the constraints $\mathbf{K} \mathbf{q} \geq \mathbf{0}$. Thus, the

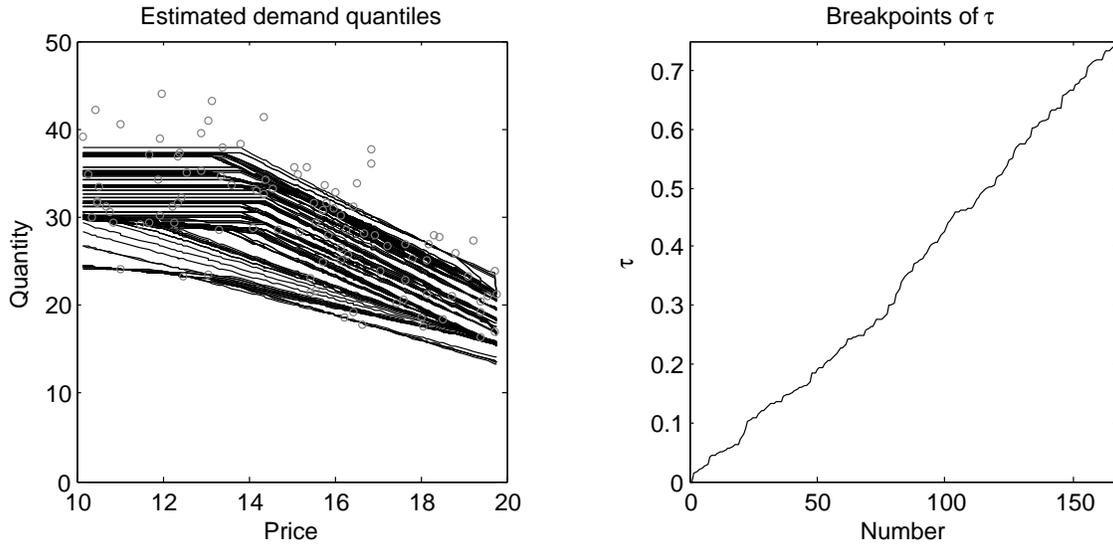


Figure 3.8: Quantile path and breakpoints with concavity constraint

last term in the objective function of the above problem is equal to

$$\sum_{i=1}^{n-2} \bar{w}_i |-h_i q_i + (h_i + h_{i+1}) q_{i+1} - h_{i+1} q_{i+2}|. \quad (3.7)$$

Let us compare (3.7) with the definition of total variation in (3.3). The expression in (3.7) can be viewed as the weighted total variation. Thus, incorporating demand concavity is equivalent to employing a weighted version of smoothing. Figures 3.8 and 3.9 show a numerical example with the concavity constraint.

Kernelization

We have used smoothing to improve the basic data-driven approach. Another widely used regularization method the kernelization method, which is widely used in support vector machines (SVM; see Suykens and Vandewalle, 1999). Assume the demand quantile function is of the form

$$q^\tau(p) = \alpha + \langle \beta, \phi(p) \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\phi(p) = (\phi_1(p), \phi_2(p), \dots)$ is a set of basis functions or feature map, which maps the observations into a higher dimensional space. $\langle \beta, \phi(p) \rangle$ can be viewed as a linear combination of the basis functions. It can also be viewed as a function in the reproducing kernel Hilbert space (RKHS) corresponding to the feature map ϕ . The definition and theoretical foundations of RKHS are beyond the scope of this chapter. Interested readers may refer to the review paper by Wahba (1999).

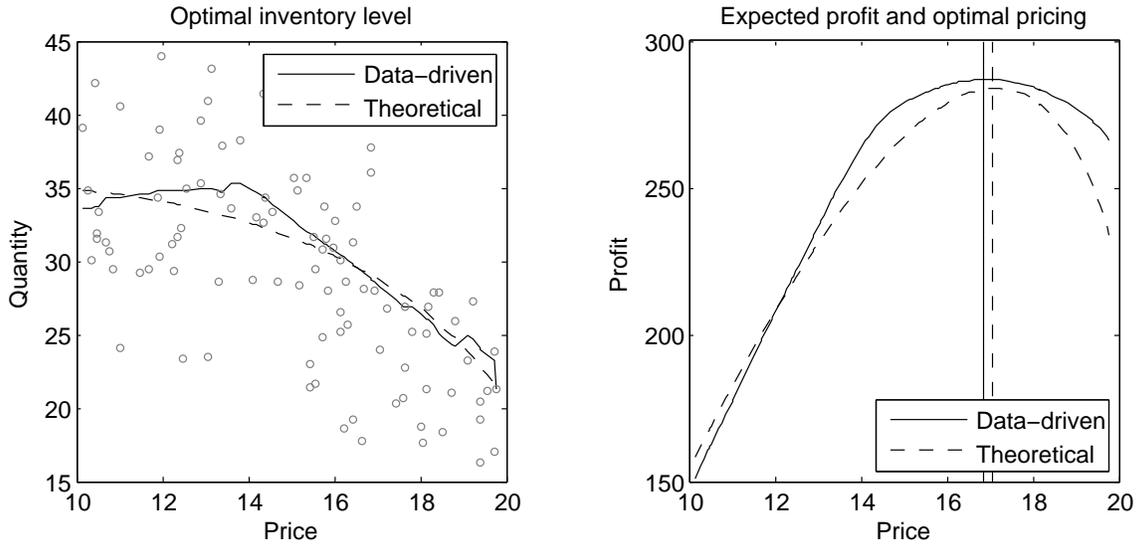


Figure 3.9: Improved data-driven solutions with concavity constraint

Takeuchi et al. (2006) developed a nonparametric quantile regression method using the kernelization method. Consider the following optimization problem which is similar to multivariate linear quantile regression

$$\sum_{i=1}^n \rho^\tau(d_i - \alpha - \langle \boldsymbol{\beta}, \boldsymbol{\phi}(p_i) \rangle) + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2.$$

The first term is the same empirical risk, i.e., the sum of weighted absolute errors, as in linear and isotonic quantile regression. The second term can be viewed as \mathcal{L}_2 regularization of the linear combination of basis functions. It can also be viewed as penalizing the RKHS norm (see Wahba, 1999). Monotone quantile estimates are obtained by constraining the derivative of $q^\tau(p)$. Let $\boldsymbol{\psi} = (\psi_1(p), \psi_2(p), \dots)$ denote the derivative of $\boldsymbol{\phi}$, i.e.,

$$\psi_j(p) = \frac{d}{dp} \phi_j(p).$$

Then, the derivative of $q^\tau(p)$ is given by

$$\frac{d}{dp} q^\tau(p) = \langle \boldsymbol{\beta}, \boldsymbol{\psi}(p) \rangle.$$

The following constraints are added to obtain monotone quantile estimates

$$\langle \boldsymbol{\beta}, \boldsymbol{\psi}(p_i) \rangle \leq 0, \forall i = 1, \dots, n.$$

As in (3.2), define auxiliary variables u_i and v_i to represent positive and negative errors. We have the following quadratic program

$$\begin{aligned} \min_{\alpha, \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}} \quad & \sum_{i=1}^n \tau u_i + (1 - \tau)v_i + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2 \\ \text{s.t.} \quad & \alpha + \langle \boldsymbol{\beta}, \boldsymbol{\phi}(p_i) \rangle + u_i - v_i = d_i, \quad \forall i = 1, \dots, n \\ & \langle \boldsymbol{\beta}, \boldsymbol{\psi}(p_i) \rangle \leq 0, \quad \forall i = 1, \dots, n \\ & u_i, v_i \geq 0, \quad \forall i = 1, \dots, n \end{aligned} \quad (3.8)$$

We do not solve this problem directly. Instead, we solve the dual problem, and the optimal estimate can be recovered from the dual solution.

Consider the Lagrangian of (3.8) with multipliers $\boldsymbol{\mu}$ and $\boldsymbol{\nu} \geq 0$

$$\begin{aligned} & \mathcal{L}(\alpha, \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\nu}) \\ &= \sum_{i=1}^n \tau u_i + (1 - \tau)v_i + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2 \\ & \quad + \sum_{i=1}^n \mu_i (d_i - \alpha - \langle \boldsymbol{\beta}, \boldsymbol{\phi}(p_i) \rangle - u_i + v_i) + \sum_{i=1}^n \nu_i \langle \boldsymbol{\beta}, \boldsymbol{\psi}(p_i) \rangle \\ &= \sum_{i=1}^n (\tau - \mu_i)u_i + (1 - \tau + \mu_i)v_i - \alpha \sum_{i=1}^n \mu_i \\ & \quad + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2 - \left\langle \boldsymbol{\beta}, \sum_{i=1}^n \mu_i \boldsymbol{\phi}(p_i) - \nu_i \boldsymbol{\psi}(p_i) \right\rangle + \sum_{i=1}^n \mu_i Y_i \end{aligned}$$

Minimizing $\mathcal{L}(\alpha, \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\nu})$ with respect to $\alpha, \boldsymbol{\beta}$ and $\mathbf{u}, \mathbf{v} \geq 0$, we obtain the dual of (3.8)

$$\begin{aligned} \min \quad & \frac{1}{2\lambda} \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{bmatrix}^T \begin{bmatrix} \mathbf{K} & -\mathbf{D} \\ -\mathbf{D}^T & \mathbf{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{bmatrix} - \mathbf{d}^T \boldsymbol{\mu} \\ \text{s.t.} \quad & (\tau - 1)\mathbf{1} \leq \boldsymbol{\mu} \leq \tau\mathbf{1} \\ & \mathbf{1}^T \boldsymbol{\mu} = 0 \\ & \boldsymbol{\nu} \geq \mathbf{0} \end{aligned} \quad (3.9)$$

where \mathbf{K} , \mathbf{D} , and \mathbf{H} are $n \times n$ matrices with elements

$$\begin{aligned} K_{ij} &= \langle \boldsymbol{\phi}(p_i), \boldsymbol{\phi}(p_j) \rangle, \\ D_{ij} &= \langle \boldsymbol{\phi}(p_i), \boldsymbol{\psi}(p_j) \rangle, \\ H_{ij} &= \langle \boldsymbol{\psi}(p_i), \boldsymbol{\psi}(p_j) \rangle. \end{aligned}$$

The variable α in (3.8) is the dual variable corresponding to the equality constraint in (3.9). Also, when minimizing $\mathcal{L}(\alpha, \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\nu})$, we obtain the intermediate result that

$$\boldsymbol{\beta} = \frac{1}{\lambda} \sum_{i=1}^n \mu_i \boldsymbol{\phi}(p_i) - \nu_i \boldsymbol{\psi}(p_i).$$

Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\nu}}$ be the optimal solution to (3.9) and let $\hat{\alpha}$ be the optimal dual variable associated with the equality constraint in (3.9). The conditional quantile function can then be recovered as

$$\begin{aligned} & \hat{q}^\tau(p) \\ &= \hat{\alpha} + \left\langle \hat{\boldsymbol{\beta}}, \boldsymbol{\phi}(p) \right\rangle \\ &= \hat{\alpha} + \frac{1}{\lambda} \sum_{i=1}^n \hat{\mu}_i \langle \boldsymbol{\phi}(p_i), \boldsymbol{\phi}(p) \rangle - \hat{\nu}_i \langle \boldsymbol{\psi}(p_i), \boldsymbol{\phi}(p) \rangle \end{aligned} \quad (3.10)$$

From (3.9) and (3.10), we see that the feature map $\boldsymbol{\phi}(p)$ or its derivative $\boldsymbol{\psi}(p)$ does not appear in the optimization problem or the recovered quantile estimate. This suggests that we do not need to specify $\boldsymbol{\phi}(p)$. Instead, we only need to know the inner products $\langle \boldsymbol{\phi}(p), \boldsymbol{\phi}(p') \rangle$, $\langle \boldsymbol{\phi}(p), \boldsymbol{\psi}(p') \rangle$, and $\langle \boldsymbol{\psi}(p), \boldsymbol{\psi}(p') \rangle$ for given p and p' . This can be done using the “kernel trick”. Let $\kappa(p, p')$ be a kernel function such that

$$\kappa(p, p') = \langle \boldsymbol{\phi}(p), \boldsymbol{\phi}(p') \rangle.$$

Taking the first and second partial derivatives of the kernel function, we have

$$\kappa_2(p, p') \triangleq \frac{\partial}{\partial p'} \kappa(p, p') = \langle \boldsymbol{\phi}(p), \boldsymbol{\psi}(p') \rangle,$$

and,

$$\kappa_{12}(p, p') \triangleq \frac{\partial^2}{\partial p \partial p'} \kappa(p, p') = \langle \boldsymbol{\psi}(p), \boldsymbol{\psi}(p') \rangle.$$

Thus, the matrices in (3.9) are given by

$$\begin{aligned} K_{ij} &= \kappa(p_i, p_j), \\ D_{ij} &= \kappa_2(p_i, p_j), \\ H_{ij} &= \kappa_{12}(p_i, p_j), \end{aligned}$$

and the recovered quantile estimate is given by

$$\hat{q}^\tau(p) = \hat{\alpha} + \frac{1}{\lambda} \sum_{i=1}^n \hat{\mu}_i \kappa(p, p_i) - \hat{\nu}_i \kappa_2(p, p_i).$$

Similarly to the isotonic quantile regression case, we need to estimate the full quantile path. Note that the quadratic program in (3.9) is parameterized by τ . We can apply the parametric active set method (PASM) in Best (1996) to efficiently compute all the necessary quantile levels. PASM takes advantage of the fact that the optimal solution to a parametric quadratic program of the form (3.9) is piecewise linear in the parameter τ . The breakpoints of the piecewise linear solution path are τ values such that the primal or dual active sets will change, and can be found from the Karush-Kuhn-Tucker (KKT) conditions.

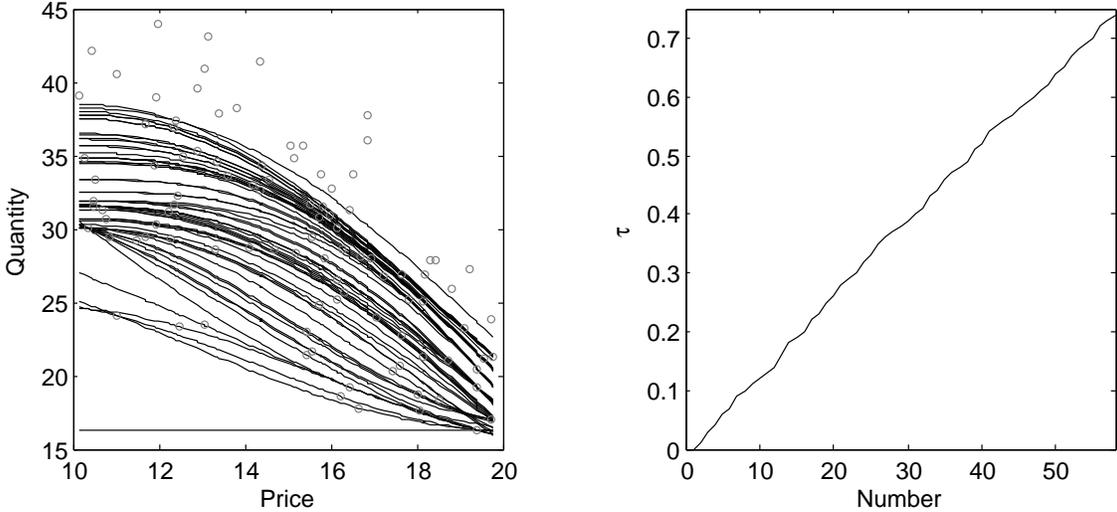


Figure 3.10: Quantile path and breakpoints using kernelization

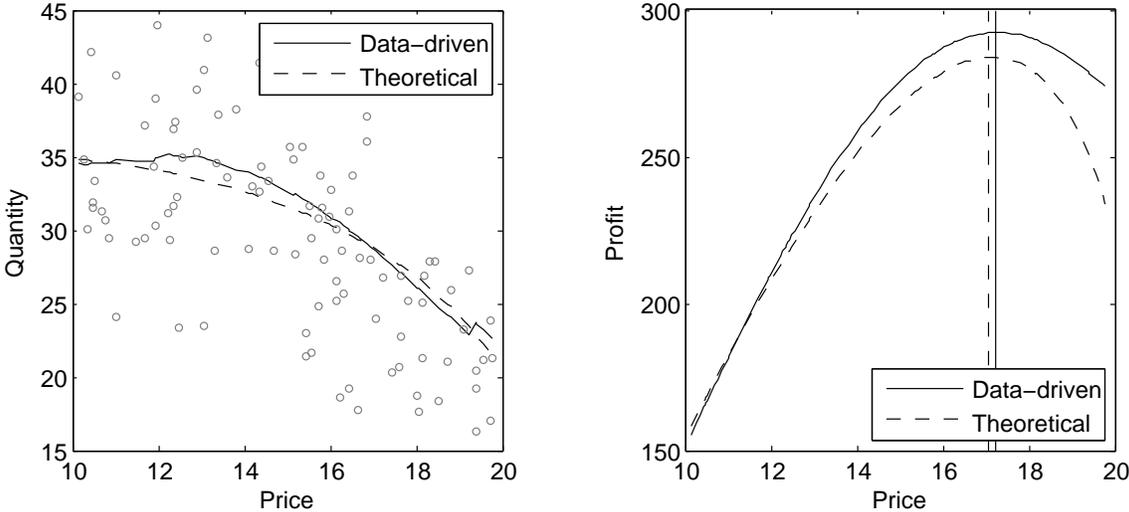


Figure 3.11: Improved data-driven solutions using kernelization

An implementation of PASM was presented by Potschka et al. (2010). Note that since the solution path is piecewise linear, calculation of the expected profit is slightly different than in the isotonic regression case. Figures 3.10 and 3.11 show results for a numerical example of the kernelization method.

For the kernelization method, the smoothing factor λ measures the penalty associated with the model complexity. When λ is sufficiently large, the problem reduces to one-sample unconditional quantile estimation. The performance of the method also depends on the kernel function. In this chapter, we use the Gaussian radial-based function (RBF) kernel

$$\kappa(p, p') = \exp \left\{ -\frac{(p - p')^2}{2\sigma^2} \right\}.$$

The intuition underlying RBF kernels is that if the price difference $|p - p'|$ is small, the difference between the conditional demand distributions of p and p' should also be small. Thus, demand observations with a price p' , which is close to p , should carry more weight when estimating the demand for price p . The parameter σ is called the bandwidth. It controls the degree to which differences between prices affect the weights of the observations. When σ is small, the estimate depends more on local observations, i.e., those with small $|p - p'|$. When σ is large, the estimate depends on the observations more globally. Observations with large $|p - p'|$ may still have substantial impact on the estimate. Selection of λ and σ is extensively studied for SVM. Please refer to Cherkassky and Ma (2004) and the references therein. In this chapter, we select the parameters using cross validation. More details are discussed in Section 3.5.

3.5 Numerical Experiments

We study the performance of the data-driven approach using numerical experiments. Demand data are generated using the additive-multiplicative model

$$D(p, \varepsilon) = \mu(p) + \sigma(p)\varepsilon.$$

For $\mu(p)$ or $\sigma(p)$, we consider four different types of functions, which are constant, concave, linear, and convex, respectively. Prices p_1, \dots, p_n are generated randomly from a uniform distribution on the interval $[p, \bar{p}]$. The random factors $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed. We consider three different distributions for the random factor, namely, exponential, uniform, and normal. The distributions are transformed such that the random factor has mean 0 and variance 1. Given p_i and ε_i , demand d_i is calculated using the demand function $D(p, \varepsilon)$. To avoid negative demand, we then let $d_i = \max\{0, d_i\}$.

Combinations of different functional forms of $\mu(p)$ and $\sigma(p)$ and different distributions of ε create 48 different demand models (including 3 trivial cases where both $\mu(p)$ and $\sigma(p)$ are constant.) For each demand model, we generate data sets with sample sizes $n = 50, 100$, and 200 , respectively. For each sample size, we further generate 10 random replications. For each

Table 3.1: Summary of the numerical experiment results

n	Method	Δp	$ \Delta p $	Δy	$ \Delta y $	$\Delta \phi$
50	I	-4.35	11.38	10.16	13.00	7.76
	S	-0.59	8.18	3.77	7.37	3.68
	K	4.00	8.24	5.93	9.12	4.83
100	I	-4.35	11.70	11.07	12.73	8.32
	S	-1.53	6.42	1.90	4.82	2.04
	K	1.95	7.06	3.13	5.86	3.31
200	I	-5.31	11.47	9.32	10.63	7.12
	S	-0.76	5.52	1.23	3.87	1.56
	K	0.27	5.24	0.42	4.33	1.82

Note: n : sample size; Δp : % price difference; Δy : % inventory difference; $\Delta \phi$: % optimality gap; I: isotonic; S: smoothing; K: kernelization.

sample, we solve the data-driven pricing-newsvendor problem using three different methods, namely, the basic isotonic regression method with no regularization, the improved method with smoothing, and the improved method with kernelization.

As we mentioned in the previous section, the smoothing factor λ and bandwidth σ are selected using cross validation. For the smoothing method, we use parametric programming to efficiently evaluate all possible λ in $[0, \infty)$. For the kernelization method, we apply the bandwidth selection trick described in Takeuchi et al. (2006). We first determine an initial bandwidth σ_0 , equal to the average of the 10-th and 90-th percentiles of the differences between prices, and an initial smoothing factor $\lambda_0 = n^{-2} \log(n)$. Then, combinations of bandwidths $10^k \sigma_0$, $k = -2, -1, 0, 1, 2$, and smoothing factors $10^k \lambda_0$, $k = -2, -1, 0, 1, 2$, are enumerated. We use repeated random sub-sampling cross validation. In five independent replications, we randomly partition the sample into a training set and a validation set with equal sizes. In order to select proper parameters for the full quantile path, we consider representative quantiles with $\tau = 0.2, 0.4$, and 0.6 . The average cross validation risk from different replications and quantile levels is used to select the parameters.

We focus on the relative difference between the data-driven solution and the theoretical optimal solution. Let \hat{p} be the optimal price found by the data-driven approach, and p^* the theoretical optimal price. The relative difference in the pricing decision is defined as

$$\Delta p = \frac{100(\hat{p} - p^*)}{p^*}.$$

Similarly, the relative difference in the inventory decision is defined as

$$\Delta y = \frac{100(\hat{y} - y^*(\hat{p}))}{y^*(\hat{p})},$$

where \hat{y} is the optimal inventory level by the data-driven approach, and $y^*(p)$ is the theoretical optimal inventory level for price p . Note that we do not compare \hat{y} with $y^*(p^*)$, i.e., the

Table 3.2: Summary of the numerical experiment results with demand concavity

n	Method	Δp	$ \Delta p $	Δy	$ \Delta y $	$\Delta\phi$
50	I	4.61	5.53	3.81	8.78	4.95
	S	1.48	3.85	0.76	6.47	2.86
	K	2.09	4.63	0.47	7.69	3.31
	C	0.80	3.10	-0.46	5.77	2.01
100	I	1.26	4.84	1.26	6.27	2.75
	S	-0.17	2.31	-0.81	3.46	0.82
	K	1.71	3.32	0.17	3.87	2.31
	C	-0.20	1.93	-0.62	3.87	0.95
200	I	1.20	4.45	1.72	5.51	2.70
	S	0.26	2.45	-0.31	2.96	1.12
	K	1.23	2.38	-2.75	5.70	2.00
	C	-0.37	2.39	-0.06	2.95	1.26

Note: n : sample size; Δp : % price difference; Δy : % inventory difference; $\Delta\phi$: % optimality gap; I: isotonic; S: smoothing; K: kernelization; C: demand concavity.

theoretical optimal inventory level given the pricing decision is also theoretically optimal. Instead, we compare \hat{y} with the theoretical optimal inventory level for \hat{p} . The optimality gap, i.e., the relative difference in the expected profit, is defined as

$$\Delta\phi = \frac{100(\phi^* - \phi(\hat{p}, \hat{y}))}{\phi^*},$$

where ϕ^* is the theoretical highest expected profit. Note that $\Delta\phi$ is always positive. Table 3.1 summarizes the average relative differences, as well as the average of their absolute values, across all 48 demand models, each with 10 random replications.

From Table 3.1, we see that in general, the data-driven approach is able to find close-to-optimal solutions. Comparing different methods with the same sample size, we see that substantial improvement is achieved by smoothing and kernelization. Comparing different sample sizes, we see that due to the lack of regularization, the basic isotonic approach may not benefit from more observations. On the other hand, for smoothing and kernelization, the performance can be significantly improved when more observations become available.

We also study the effect of incorporating demand concavity. We focus on concave additive demand models, where $D(p, \varepsilon)$ is guaranteed to be concave for given ε . Table 3.2 summarizes the corresponding results. As we mentioned, incorporating demand concavity is equivalent to smoothing with some weighted total variation penalty. From Table 3.2, we see that incorporating demand concavity can also enhance the performance of the data-driven approach, and it indeed achieves an effect similar to that of smoothing.

3.6 Summary and Future Directions

In this chapter, we present a data-driven approach for the pricing-newsvendor problem where the relationship between price and demand is unknown. The approach does not make any parametric assumption on the underlying demand model. Instead, it is based on historical observations and basic domain knowledge, which are available in practice. Parametric programming is applied to efficiently estimate the conditional quantile path of the demand. Smoothing and kernelization are used to improve the estimates and decisions. Additional domain knowledge, such as demand concavity, can also be easily incorporated in the approach. Numerical experiments show that the data-driven approach is able to find close-to-optimal solutions. Smoothing, kernelization, and the incorporation of additional domain knowledge can significantly enhance the performance of the approach. The contribution of this chapter is twofold. First, it presents a new method for the pricing newsvendor problem that does not require model selection, while most of the existing methods require the selection of a parametric demand model. Thus, it improves the applicability of the joint pricing-inventory management models. Second, it applies nonparametric statistical learning methods (i.e., smoothing spline and kernelization) to improve the performance of the data-driven approach. These methods are relatively new in the operations research literature, and we believe they have great potential in solving many other operations research problems.

For future research, we plan to generalize this approach for joint pricing-inventory management for an assortment. In assortment management, the demand for one product may depend on the prices of other related products. For example, if two products are substitutes for each other, then demand for one product will increase in the price of the other. If two products are complementary to each other, then demand for one product will decrease in the price of the other. These monotonicity constraints can be incorporated in the kernelization method. Instead of univariate quantile regression, we will use multivariate quantile regression to estimate the conditional demand distribution which depends on the prices of multiple products. The current approach can also be improved in several dimensions. First, the current smoothing method uses \mathcal{L}_1 regularization, which results in piecewise linear quantile estimates with few breakpoints. We plan to study the effect of \mathcal{L}_2 regularization. Second, we have shown how to incorporate demand concavity in the approach. It is important to consider how other domain knowledge, e.g., price elasticity, can be incorporated. Finally, advanced bandwidth selection methods can be applied to improve the performance of smoothing and kernelization.

Chapter 4

Inventory Management for Perishable Goods with Parameter Uncertainty

4.1 Introduction

Consider the classical newsvendor model in stochastic inventory management. It is typically assumed that the customer demand has a known distribution function $F(\cdot; \theta)$ with parameter θ . (Note that we are abusing the notation θ , as it can denote a vector of parameters. Later in this chapter, we consider the case where the distribution has a location parameter and a scale parameter.) For each θ , we can find the optimal order quantity $y^*(\theta)$. In practice, given historical demand observations $\mathbf{X} = (X_1, \dots, X_n)$, we can first select the form of the distribution function. Then, the parameter of the demand distribution is estimated using a specific criterion, e.g., maximum likelihood. With the estimated parameter $\hat{\theta}$, we find the optimal order quantity as $y^*(\hat{\theta})$. This sequential approach, which we call separated estimation-optimization in this chapter, has been the standard approach in inventory management and many other operations management problems. Liyanage and Shanthikumar (2005) showed that the separation of parameter estimation and order quantity optimization will lead to suboptimality. They proposed an integrated approach called operational statistics. In this approach, instead of having the intermediate step of parameter estimation, one determines the order quantity as a function or statistic $g(\cdot)$ of demand observations directly. For a given demand sample \mathbf{X} , the optimal order quantity can then be determined as $g(\mathbf{X})$. They studied three types of operational statistics based on the sample mean, the empirical distribution, and the sample spacings, respectively. Chu, Shanthikumar, and Shen (2008) found the optimal operational statistic within a general class of functions.

In this chapter, we first study properties of the optimal operational statistic in Chu, Shanthikumar, and Shen (2008). We show that the optimal operational statistic is consistent and guaranteed to outperform the traditional separated estimation-optimization approach. We also find that the benefit achieved by operational statistics is higher when demand variability is larger. We then generalize the operational statistics approach to the risk-averse

newsvendor problem under the conditional value-at-risk (CVaR) criterion. (Note that the CVaR criterion usually leads to overly conservative solution. We chose this criterion for ease of exposition, while the results can be generalized to other coherent risk measures, e.g., a weighted sum of mean and CVaR.) The results in Chu, Shanthikumar, and Shen (2008) can be directly generalized to maximize the expectation of conditional CVaR. In order to model risk-aversion to both demand sampling risk and future demand uncertainty risk, we introduce a new criterion, called total CVaR, and find the optimal operational statistic under this new criterion. We think the total CVaR criterion could be a useful framework for modeling risk-aversion in the presence of parameter uncertainty.

The remainder of the chapter is organized as follows. Section 4.2 reviews the related literature. Section 4.3 introduces the model setting and some background on operational statistics. Section 4.4 presents several properties of operational statistics. Section 4.5 generalizes operational statistics to the risk-averse case under the CVaR criterion. Section 4.6 summarizes the results and discusses directions for future work.

4.2 Literature Review

Inventory management, or more specifically, the newsvendor problem, is widely studied in the operations research literature. We consider a repeated newsvendor setting, i.e., the ordering decision is repeated for multiple periods, but excess inventory at the end of each period is not carried over to the next period. This setting is suitable for perishable products. We focus on the case where the parameters of the demand distribution are unknown. The problem of demand parameter estimation in the newsvendor problem was first studied by Hayes (1969). He proposed minimizing the expected total operating cost (ETOC), which leads to a special case of the operational statistics approach in Liyanage and Shanthikumar (2005) and Chu, Shanthikumar, and Shen (2008). Akcay, Biller, and Tayur (2011) generalized the ETOC framework to more general demand distributions using the Johnson translation system. Ramamurthy, Shanthikumar, and Shen (2012) studied the operational statistics approach when the demand distribution has an unknown shape parameter. Chu, Shanthikumar, and Shen (2009) studied operational statistics for the pricing-newsvendor problem. The focus of this chapter is to study several important properties of operational statistics, and to generalize the operational statistics approach to the risk-averse case.

In the presence of parameter uncertainty, the most popular approach for inventory management is the Bayesian framework. It was first applied to the newsvendor problem by Scarf (1959), and then by Iglehart (1964). Unlike the Bayesian framework, operational statistics is a frequentist approach. Interestingly, the optimal operational statistic found by Chu, Shanthikumar, and Shen (2008) coincides with a non-trivial Bayesian framework. In Section 4.5, we find the optimal operational statistic under the total CVaR criterion, which also has a Bayesian interpretation. We note that this interesting phenomenon, i.e., the optimal frequentist procedure coinciding with a Bayesian framework, is not rare in the Statistics literature (see Bayarri and Berger, 2004, for more details).

Another related approach is the sampling-based approach. In this approach, only historical demand data are available. No parametric model is assumed. Instead, the empirical counterpart of the problem is solved. Levi, Roundy, and Shmoys (2007) studied sampling-based policies for the newsvendor problem and its multi-period version with inventory carry-over. They derived bounds on the performance of the policies. Levi, Perakis, and Uichanco (2012) improved the bound for the newsvendor problem presented in Levi, Roundy, and Shmoys (2007) using spread information. Huh and Rusmevichientong (2009) studied a sampling-based approach for the newsvendor problem with censored demand observations. Huh et al. (2011) studied the same problem using the Kaplan-Meier estimator. Jain et al. (2011) considered the case where demand depends on the inventory level. They developed a new approach called operational objective learning based on kernel smoothing. Unlike the sampling-based approach, the operational statistics approach considers the case where the distribution family is known, e.g., after a distribution function has been selected, and focuses on parameter uncertainty only.

Another related stream of research is on the distribution-free or distributionally-robust newsvendor problem. It was first studied by Scarf (1958) and then by Gallego and Moon (1993). In this problem, the parameters (i.e., mean and variance) of the demand are known, but the distribution is unknown. The order quantity is chosen to minimize the expected cost under the worst-case distribution. Other types of distributional information have also been used in the newsvendor model. Lim, Shanthikumar, and Shen (2006) considered the case where the demand distribution is close to a known nominal distribution. The distance between the actual distribution and the nominal distribution is measured by the Kullback-Leibler divergence, also known as the relative entropy. Wang, Glynn, and Ye (2010) considered a similar problem, except that the nominal distribution is the empirical distribution given by the previous demand observations. Focusing on the worst-case may be overly conservative in practice. An alternative objective is to minimize the worst-case regret. Perakis and Roels (2008) studied the newsvendor model where the order quantity is chosen to minimize the maximum regret of not acting optimally.

Risk-aversion is widely observed in inventory management. The risk-averse newsvendor problem has been studied by Eeckhoudt, Gollier, and Schlesinger (1995) using expected utility functions, and by Chen and Federgruen (2000) using mean-variance analysis. Recently, modeling risk-aversion in inventory management using coherent risk measures (e.g., CVaR) has gained intensive attention. Coherent risk measures have important theoretical and practical value in risk analysis. The newsvendor problem under coherent risk measures has recently been studied by Ahmed, Cakmak, and Shapiro (2007), Gotoh and Takano (2007), and Choi and Ruszczyński (2008). In addition to the standard newsvendor problem, Chen, Xu, and Zhang (2009) considered the pricing newsvendor problem under CVaR, and Choi, Ruszczyński, and Zhao (2011) considered the multi-product newsvendor problem. Yang et al. (2009) studied coordinating contracts when selling to a risk-averse newsvendor under the CVaR criterion. Bertsimas and Thiele (2005) proposed a sampling-based approach for the newsvendor problem. Instead of maximizing the empirical mean profit, they proposed to maximize the empirical CVaR of the profit to improve the robustness of the solutions. We

are the first to consider the CVaR newsvendor problem with parameter uncertainty. We also introduce a new criterion, called total CVaR, which incorporates risk-aversion to both demand sampling risk and future demand uncertainty risk.

4.3 Background on Operational Statistics

In this section, we introduce the model setting and some background on the operational statistics approach. We adopt the same setting for the repeated newsvendor problem as in Chu, Shanthikumar, and Shen (2008). Let s be the unit selling price, and c the unit ordering cost. We assume there is no salvage value, which is consistent with the setting in Chu, Shanthikumar, and Shen (2008). Incorporating salvage value will not change the structure of the problem significantly. The profit from order quantity y and demand X is given by

$$\Phi(y, X) = s \min\{y, X\} - cy.$$

Assume the demand distribution is characterized by a location parameter τ and a scale parameter $\theta > 0$, i.e., $X = \tau + \theta Z$, where Z has a known distribution. Let $F(x; \tau, \theta)$ and $f(x; \tau, \theta)$ denote the cumulative distribution function and the density function of the demand, respectively. With some abuse of notation, we use $F(z)$ and $f(z)$ to denote the distribution function and density function of Z , i.e., $F(z) = F(z; 0, 1)$ and $f(z) = f(z; 0, 1)$. The expected profit from order quantity y is denoted by

$$\phi(y; \tau, \theta) = \int \Phi(y, x) f(x; \tau, \theta) dx.$$

From the well known result for the newsvendor problem, the optimal order quantity

$$y^*(\tau, \theta) = F^{-1}\left(1 - \frac{c}{s}; \tau, \theta\right),$$

where F^{-1} is the inverse of the cumulative distribution function.

Assume the parameters τ and θ are unknown, but demand observations in the previous n periods $\mathbf{X} = (X_1, \dots, X_n)$ are available. In the separated estimation-optimization approach, one first obtains parameter estimates $\hat{\tau}(\mathbf{X})$ and $\hat{\theta}(\mathbf{X})$. Using the estimates, the order quantity can be found as $y^*(\hat{\tau}(\mathbf{X}), \hat{\theta}(\mathbf{X}))$. In the operational statistics approach, one directly determines the order quantity as a function or statistic $g(\mathbf{X})$ of the observations, to maximize the a priori expected profit $\mathbb{E}_{\mathbf{X}}[\phi(g(\mathbf{X}); \tau, \theta)]$. In order to find a uniformly optimal operational statistic for all parameters, the form of the function g must be restricted. Chu, Shanthikumar, and Shen (2008) considered the following class of functions

$$\mathcal{H}_1^e = \{g : \mathbb{R}^n \rightarrow \mathbb{R} : g(\alpha \mathbf{x} - \delta \mathbf{e}) = \alpha g(\mathbf{x}) - \delta, \alpha > 0\}.$$

They showed that the optimal operational statistic in class \mathcal{H}_1^e is given by

$$h^*(\mathbf{x}) = \operatorname{argmax}_y \left\{ \int_{\tau} \int_{\theta} \phi(y; \tau, \theta) \frac{1}{\theta^{n+2}} \prod_{i=1}^n f\left(\frac{x_i - \tau}{\theta}\right) d\theta d\tau \right\}. \quad (4.1)$$

4.4 Properties of Operational Statistics

In this section, we present several desirable properties of the operational statistics approach.

Finding the Optimal Operational Statistic

First, we observe that for a general class of distributions, the optimization problem in (4.1) is relatively easy to solve, i.e., the optimal operational statistics is relatively easy to find. The integral in (4.1) preserves the concavity of $\phi(y; \tau, \theta)$. The optimal operational statistic can then be found from the first order condition. Note that the derivative of the expected profit is given by

$$\phi'(y; \tau, \theta) = s - c - sF(y; \tau, \theta).$$

The derivative is uniformly bounded, i.e.,

$$\sup_y |\phi'(y; \tau, \theta)| \leq \max\{s - c, c\}.$$

Thus, if the distribution satisfies

$$\int_{\tau} \int_{\theta} \frac{1}{\theta^{n+2}} \prod_{i=1}^n f\left(\frac{x_i - \tau}{\theta}\right) d\theta d\tau < \infty, \quad (4.2)$$

then we can take the derivative of the objective function in (4.1) by interchanging the order of the differentiation and the double integral in (4.1).

Proposition 6. *For distributions such that (4.2) holds, the optimal operational statistic can be found by solving*

$$\int_{\tau} \int_{\theta} (s - c - sF(y; \tau, \theta)) \frac{1}{\theta^{n+2}} \prod_{i=1}^n f\left(\frac{x_i - \tau}{\theta}\right) d\theta d\tau = 0. \quad (4.3)$$

For general distributions, equation (4.3) can be solved numerically. For certain distribution types, (4.3) leads to closed-form optimal operational statistic. For example, when the demand is exponentially distributed with mean θ , the optimal operational statistic is given by

$$h^*(\mathbf{X}) = \left((s/c)^{\frac{1}{n+1}} - 1 \right) \sum_{i=1}^n X_i. \quad (4.4)$$

When the demand is uniformly distributed in $[0, \theta]$. The optimal operational statistic is given by

$$h^*(\mathbf{X}) = \begin{cases} \left(\frac{s}{(n+2)c} \right)^{\frac{1}{n+1}} X_{[n]} & \text{for } n \leq \frac{s}{c} - 2, \\ \frac{n+2}{n+1} \left(1 - \frac{c}{s} \right) X_{[n]}, & \text{for } n > \frac{s}{c} - 2, \end{cases} \quad (4.5)$$

where $X_{[n]}$ is the n -th order statistic of the observations. We will use these examples to illustrate some of the results presented later in this chapter.

Asymptotic Optimality

The optimal operational statistic can be viewed as an estimator for the theoretical optimal solution. We will show that the operational statistics approach is asymptotically optimal (i.e., consistent). For ease of exposition, we assume only the scale parameter is unknown, i.e., $X = \theta Z$, where Z has a known distribution. The result can be generalized to the case where both the location and the scale parameters are unknown.

Proposition 7. *The optimal operational statistic converges almost surely to the theoretical optimal solution with known parameter.*

We can verify the result using previous examples given in (4.4) and (4.5) for which the optimal operational statistic is given in closed-form. If the demand is exponentially distributed with mean θ , the optimal order quantity with known parameter is given by

$$y^*(\theta) = \theta \ln \left(\frac{s}{c} \right).$$

By the Strong Law of Large Numbers,

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{a.s.} \theta,$$

where $\xrightarrow{a.s.}$ stands for almost sure convergence. Also, we can show that

$$n(s^{\frac{1}{n+1}} - 1) \rightarrow \ln s.$$

Thus, the optimal operational statistic for the exponential distribution in (4.4) is asymptotically optimal, i.e.,

$$h^*(\mathbf{X}) \xrightarrow{a.s.} y^*(\theta).$$

When demand is uniformly distributed on $[0, \theta]$, the optimal order quantity with a known parameter is given by

$$y^*(\theta) = \left(1 - \frac{c}{s}\right) \theta.$$

We can show that

$$X_{[n]} \xrightarrow{a.s.} \theta.$$

Recall that when n is large,

$$h^*(\mathbf{X}) = \frac{n+2}{n+1} \left(1 - \frac{c}{s}\right) X_{[n]},$$

which is also asymptotically optimal.

Integrated vs. Separated Approach

We have shown that the operational statistics approach is asymptotically optimal. However, the separated estimation-optimization approach is also asymptotically optimal, as long as a consistent estimator for the parameter is used. We will show that the operational statistics approach is superior to the separated estimation-optimization approach, in that it achieves higher a priori expected profit. We assume that the maximum likelihood estimator is used for parameter estimation in the separated estimation-optimization approach. However, the result does not necessarily require using maximum likelihood estimation. A similar result can be shown for the method of moments. Note that we could also compare the operational statistics approach with the sampling-based approach in Levi, Roundy, and Shmoys (2007). However, since the sampling-based approach does not require knowledge of the form of the demand distribution, this comparison will not be fair. Thus, we focus on comparing the operational statistics approach with the separated estimation-optimization approach. The disadvantage of the separated estimation-optimization approach is due to the fact that it sequentially optimizes parameter estimation and the inventory decision. Loss of information occurs when a distribution is selected and fitted in the first stage.

Proposition 8. *The a priori expected profit achieved by the optimal operational statistic is greater than or equal to the one achieved by the separated estimation-optimization approach for all parameters.*

In order to better understand how the operational statistics approach achieves higher a priori expected profit, we illustrate the difference between operational statistics and separated estimation-optimization using a numerical example where the demand is exponentially distributed. We randomly generate one million demand data sets. Each data set has three observations. For each data set, we calculate the optimal operational statistics solutions using the closed-form expressions in (4.4). The optimal separated estimation-optimization solutions are obtained using the maximum likelihood estimators.

In Figure 4.1, we compare the histograms of the operational statistics and separated estimation-optimization solutions. We can see that the separated estimation-optimization approach generates more solutions near the theoretical optimal solution than the operational statistics approach. However, there are also more separated estimation-optimization solutions that are far from the theoretical optimum.

We can measure how close to optimal the solutions are using the optimality gap, i.e., the relative difference between the achieved expected profit and the theoretical optimal expected profit. In Figure 4.2, we compare the histograms of the optimality gaps for the two approaches. We consider four groups of solutions, with optimality gap less than 10%, between 10% and 50%, between 50% and 90%, and greater than 90%, respectively. We see that although the separated estimation-optimization approach generates more solutions in the first group, it also generates more solutions in the last group. On the other hand, the operational statistics approach generates more solutions in the second group. As a result, the operational statistics approach is able to achieve higher a priori expected profit.

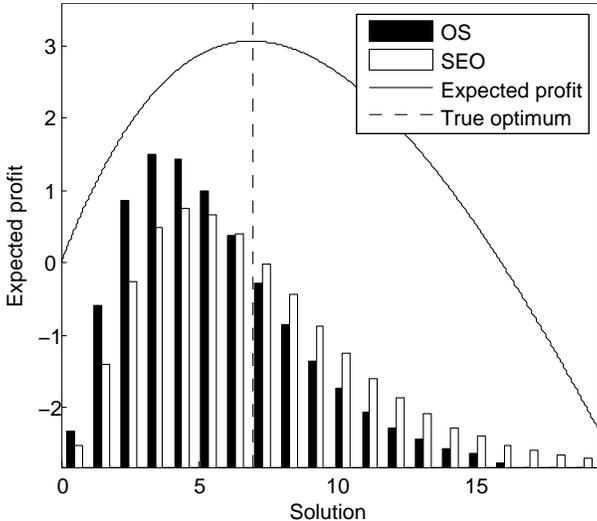


Figure 4.1: Comparison of operational statistics and separated estimation-optimization solutions (exponential distribution)

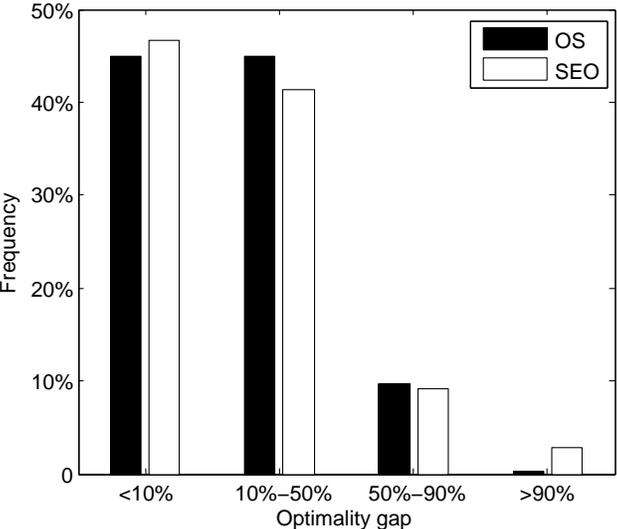


Figure 4.2: Comparison of the optimality gaps for the operational statistics and separated estimation-optimization approaches (exponential distribution)

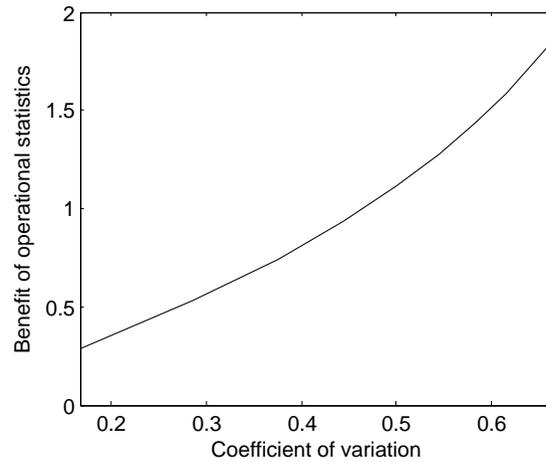


Figure 4.3: Impact of demand coefficient of variation on the benefit of operational statistics (shifted exponential distribution)

The Impact of Demand Variability

We have shown that the operational statistics approach is superior to the separated estimation-optimization approach, in that it achieves higher a priori expected profit. In this subsection, we examine the impact of demand variability on the benefit of using operational statistics. We define the benefit of operational statistics as the relative difference between the a priori expected profits achieved by operational statistics and separated estimation-optimization, respectively. We use a commonly used measure for variability, the coefficient of variation, i.e., the ratio between the standard deviation and the mean.

Proposition 9. *The benefit of operational statistics is larger when the demand variability is higher.*

We illustrate Proposition 9 with a numerical example. The shifted exponential distribution has the probability density function

$$f(x) = \frac{1}{\theta} \exp\left\{-\frac{(x - \tau)}{\theta}\right\},$$

for $x \geq \tau$. The coefficient of variation for the shifted exponential distribution is given by $\theta/(\theta + \tau)$. Although the optimal operational statistic for the shifted exponential distribution does not have a closed-form expression, we can solve for it numerically using Proposition 6. In Figure 4.3, we plot the benefit of operational statistics vs. the demand coefficient of variation. We see that as the coefficient of variation increases, using operational statistics achieves higher benefit.

Note that coefficient of variation is a common measure for risk. Thus, the previous result suggests that using operational statistics achieves larger benefit under higher risk. However,

we have thus far focused on the optimal operational statistic for the risk-neutral case. In the next section, we generalize operational statistics to the risk-averse case.

4.5 Operational Statistics under Risk Aversion

In this section, we generalize operational statistics to the newsvendor problem under risk-aversion. For inventory management, risk-aversion has been modeled using expected utility functions or mean-variance analysis. Recently, modeling risk-aversion using the conditional value-at-risk (CVaR) criterion has received a significant amount of attention. CVaR is an important and widely used risk measure in finance and operations research. For a profit maximization problem, the β -CVaR, denoted by ρ_β , is the mean of the lowest 100β percent of the profit outcomes. For continuous distributions, it is equal to the conditional expectation of profit below the lower 100β -percentile. The parameter $\beta \in (0, 1]$ reflects the degree of risk-aversion. A smaller β represents a higher degree of risk-aversion. If $\beta = 1$, it reduces to the risk-neutral case. CVaR is a coherent risk measure with the following properties:

Convexity: $\rho_\beta[\alpha U + (1 - \alpha)V] \geq \alpha\rho_\beta[U] + (1 - \alpha)\rho_\beta[V]$ (concave for a maximization problem).

Translation Equivariance: $\rho_\beta[V + a] = \rho_\beta[V] + a$, for $a \in \mathbb{R}$.

Positive Homogeneity: $\rho_\beta[\alpha V] = \alpha\rho_\beta[V]$, for $\alpha > 0$.

These properties are essential in solving the risk-averse newsvendor problem and generalizing operational statistics to the risk-averse case.

The CVaR Newsvendor Problem

Consider the newsvendor problem under the CVaR criterion. Let $\rho_\beta[y; \tau, \theta]$ denote the CVaR of profit from order quantity y when the demand has parameters τ and θ . Using the dual representation in Rockafellar and Uryasev (2000),

$$\rho_\beta[y; \tau, \theta] = \rho_\beta[\Phi(y, X)] = \sup_{\eta} \left\{ \eta - \frac{1}{\beta} \int [\eta - \Phi(y, x)]^+ f\left(\frac{x - \tau}{\theta}\right) dx \right\}.$$

In the CVaR newsvendor problem, the optimal order quantity y is chosen to maximize $\rho_\beta[y; \tau, \theta]$. The optimal solution is a variant of the standard risk-neutral newsvendor order quantity. The following result is due to Chen, Xu, and Zhang (2009).

Proposition 10. *In the CVaR newsvendor problem with parameters τ and θ , the β -CVaR of profit is given by*

$$\rho_\beta[y; \tau, \theta] = (s - c)y - \frac{s}{\beta} \int_0^y F\left(\frac{x - \tau}{\theta}\right) dx, \quad (4.6)$$

and the optimal order quantity is given by

$$y^*(\tau, \theta) = \tau + \theta F^{-1} \left(\beta \left(1 - \frac{c}{s} \right) \right). \quad (4.7)$$

Optimal Operational Statistic

Next we solve the CVaR newsvendor problem with unknown parameters using operational statistics. Given demand observations \mathbf{X} , the conditional CVaR using operational statistic $g(\cdot)$ is given by $\rho_\beta[g(\mathbf{X}); \theta]$. Note that the conditional CVaR is a random variable that depends on \mathbf{X} . We can then take the expectation of conditional CVaR over \mathbf{X} , which we call the a priori expected CVaR.

We seek the optimal operational statistic to maximize the a priori expected CVaR $\mathbb{E}_{\mathbf{X}} [\rho_\beta[g(\mathbf{X}); \tau, \theta]]$. Using the results in (4.6) and (4.7), we can show that the optimal operational statistic in (4.1) can be generalized to this case. We can also show that the properties in Propositions 6 through 9 still hold.

Proposition 11. *For the a priori expected CVaR objective, the optimal operational statistic in class \mathcal{H}_1^e is given by*

$$h^*(\mathbf{x}) = \operatorname{argmax} \left\{ \int_{\tau} \int_{\theta} \rho_\beta[y; \tau, \theta] \frac{1}{\theta^{n+2}} \prod_{i=1}^n f \left(\frac{x_i - \tau}{\theta} \right) d\theta d\tau \right\}.$$

The optimal operational statistic under risk-aversion is a variant of that for the risk-neutral case in (4.1). For example, when demand is exponentially distributed with mean θ , the optimal operational statistic under risk-aversion also has a closed-form expression, which is given by

$$h^*(\mathbf{X}) = \left(\left(\frac{s}{(1-\beta)s + \beta c} \right)^{\frac{1}{n+1}} - 1 \right) \sum_{i=1}^n X_i.$$

When the demand is uniformly distributed in the interval $[0, \theta]$, the optimal operational statistic is given by,

$$h^*(\mathbf{X}) = \begin{cases} \left(\frac{s}{(n+2)((1-\beta)s + \beta c)} \right)^{\frac{1}{n+1}} X_{[n]} & \text{for } n \leq \frac{s}{(1-\beta)s + \beta c} - 2, \\ \frac{n+2}{n+1} \beta \left(1 - \frac{c}{s} \right) X_{[n]}, & \text{for } n > \frac{s}{(1-\beta)s + \beta c} - 2. \end{cases}$$

Comparing the optimal operational statistics under risk-aversion with those for the risk-neutral case in (4.4) and (4.5), we see that for the same demand sample \mathbf{X} , the optimal operational statistic under risk-aversion is smaller than the risk-neutral case. Also, we see that the optimal operational statistic is increasing in β , i.e., the order quantity is decreasing in the degree of risk-aversion. This result can be generalized to more general distribution types.

Proposition 12. *For distributions such that*

$$\int_{\tau} \int_{\theta} \frac{1}{\theta^{n+2}} \prod_{i=1}^n f\left(\frac{x_i - \tau}{\theta}\right) d\theta d\tau < \infty,$$

the optimal operational statistic is decreasing in the degree of risk-aversion.

The Total CVaR Criterion

In the previous subsection, we found the optimal operational statistic that maximizes the a priori expected CVaR. Note that the profit $\Phi(g(\mathbf{X}), X_{n+1})$ is a random variable induced by both the demand observations \mathbf{X} and the future demand X_{n+1} . In the a priori expected CVaR objective $\mathbb{E}_{\mathbf{X}}[\rho_{\beta}[g(\mathbf{X}); \theta]]$, the conditional CVaR $\rho_{\beta}[g(\mathbf{X}); \tau, \theta]$ is calculated with respect to X_{n+1} only. The expectation $\mathbb{E}_{\mathbf{X}}[\cdot]$ is taken with respect to \mathbf{X} . Both \mathbf{X} and X_{n+1} contribute to the risk in this problem, but risk-aversion is only incorporated for the uncertainty in X_{n+1} , i.e., the future demand uncertainty. Risk-aversion is not reflected for the uncertainty in \mathbf{X} , i.e., uncertainty in demand sampling.

An alternative objective is the actual CVaR of the random variable $\Phi(g(\mathbf{X}), X_{n+1})$, denoted by $\rho_{\beta}[\Phi(g(\mathbf{X}), X_{n+1})]$. It reflects risk-aversion to both the sampling risk (i.e., risk originating from \mathbf{X}) and the future demand uncertainty risk (i.e., risk originating from X_{n+1}). Thus, we call this objective the total CVaR criterion.

For ease of exposition, we focus on the case where only the scale parameter is unknown, i.e., $X = \theta Z$, where Z has a known distribution. The result can be generalized to the case where both the location and the scale parameters are unknown. Using the dual representation in Rockafellar and Uryasev (2000), the total CVaR of profit is given by,

$$\rho_{\beta}[\Phi(g(\mathbf{X}), X_{n+1})] = \sup_{\eta} \left\{ \eta - \frac{1}{\beta} \int \int [\eta - \Phi(g(\mathbf{x}), x_{n+1})]^+ \prod_{i=1}^{n+1} f(x_i; \theta) dx_{n+1} d\mathbf{x} \right\}.$$

Although the optimal operational statistic in the previous section was obtained using the a priori expected CVaR objective, we can still compare the performance of the operational statistics solution and the separated estimation-optimization solution under the total CVaR criterion. We randomly generate one million demand data sets. Each data set consists of demand observations in the three periods immediately before an inventory decision is made plus the demand observed in the subsequent period. Using the historical observations, we calculate the optimal order quantities using operational statistics and separated estimation-optimization. We then calculate the realized profit in the next period using the order quantity and the realized demand. The total CVaR is then estimated using the empirical CVaR of the realized profits from all the demand samples. In Figure 4.4, we plot the total CVaR obtained using operational statistics and separated estimation-optimization. We can see that, although the operational statistics solution is derived for the a priori expected CVaR objective, it still achieves higher total CVaR than the separated estimation-optimization solution.

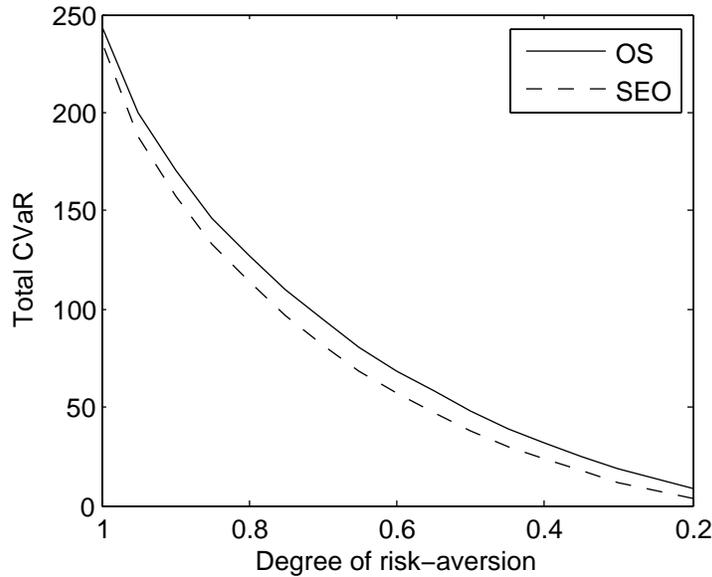


Figure 4.4: Total CVaR achieved by operational statistics and separated estimation-optimization

Next we show how to find the optimal operational statistic for the total CVaR criterion. As in the risk-neutral case, we consider functions within the class,

$$\mathcal{H}_1 = \{g : \mathbb{R}^n \rightarrow \mathbb{R} : g(\alpha \mathbf{x}) = \alpha g(\mathbf{x}), \alpha \geq 0\}.$$

Proposition 13. *For the total CVaR criterion, the optimal operational statistic in class \mathcal{H}_1 is given by*

$$h^*(\mathbf{x}) = \underset{y}{\operatorname{argmax}} H(\mathbf{x}, y), \tag{4.8}$$

where

$$H(\mathbf{x}, y) = \sup_{\eta} \left\{ \int_0^{\infty} \int \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(y, x_{n+1})}{\theta} \right]^+ \right] \frac{1}{\theta^{n+2}} \prod_{i=1}^{n+1} f\left(\frac{x_i}{\theta}\right) dx_{n+1} d\theta \right\}.$$

Recall that the optimal operational statistic for the risk-neutral case in (4.1) coincides with a Bayesian framework. The optimal operational statistic under the total CVaR criterion also has a Bayesian interpretation. Using Jeffrey’s non-informative prior for the scale parameter, the term

$$\frac{1}{\theta^{n+1}} \prod_{i=1}^n f\left(\frac{X_i}{\theta}\right)$$

can be viewed as the posterior distribution of the random parameter Θ . Thus

$$\int_0^\infty \left(\int \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\phi(y, x_{n+1})}{\theta} \right]^+ \right] \frac{1}{\theta} f\left(\frac{x_{n+1}}{\theta}\right) dx_{n+1} \right) \frac{1}{\theta^{n+1}} \prod_{i=1}^n f\left(\frac{x_i}{\theta}\right) d\theta$$

can be viewed as the posterior mean of

$$\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(y, X_{n+1})}{\Theta} \right]^+.$$

The objective function in (4.8) is thus the CVaR of the normalized profit $\frac{\phi(y, X_{n+1})}{\Theta}$ with regard to the posterior distribution of parameter Θ .

4.6 Summary and Future Directions

In this chapter, we study the operational statistics approach in Chu, Shanthikumar, and Shen (2008) for perishable inventory management with parameter uncertainty. We find that the operational statistics approach is consistent and guaranteed to outperform the separated estimation-optimization approach. We find that the benefit of using operational statistics is greater when demand variability is higher. We then study operational statistics for the risk-averse newsvendor problem under the conditional value-at-risk (CVaR) criterion. We find that the results in Chu, Shanthikumar, and Shen (2008) for the risk-neutral case can be generalized to the objective of maximizing the expected CVaR. In order to model risk-aversion to both demand sampling risk and future demand uncertainty risk, we introduce a new criterion called the total CVaR, and find the optimal operational statistic under this criterion.

A challenging problem for future research is to derive operational statistics for multi-period inventory management problems. An interesting question is whether the well-studied base-stock policy will still be optimal. The operational statistics framework can also be applied to other problems in operations research, such as pricing and revenue management.

Appendix A

Proofs

Proof of Lemma 1. First, note that the minimization problem in (2.1) is decomposable in $i \in \mathcal{I}$. For each $i \in \mathcal{I}$, the optimal solution is straightforward:

$$h_i(\mathbf{x}, S) = d_i \min\{c_{ij} : j \in S, x_j = 1\}.$$

For any $S \subset T \subset J$, let us evaluate the decrease in cost after adding $j \in J \setminus T$ to the set. If $x_j = 0$, then adding j will not affect the assignment or the cost, and thus condition (2.4) holds trivially. So we only need to consider the case where $x_j = 1$. Let $j(S)$ denote the facility to which customer i is assigned given disruption scenario S , and similarly $j(T)$. Since $S \subset T$, it is easy to see that $c_{i,j(T)} \leq c_{i,j(S)}$. There are three cases depending on the value of c_{ij} :

1. $c_{i,j} \leq c_{i,j(T)}$

In this case, having the additional facility j will reduce the cost for both disruption scenarios S and T . For disruption scenario S , the amount of reduction is given by,

$$h_i(\mathbf{x}, S) - h_i(\mathbf{x}, S \cup \{j\}) = c_{i,j(S)} - c_{i,j}.$$

Similarly, for disruption scenario T , the amount of reduction is given by,

$$h_i(\mathbf{x}, T) - h_i(\mathbf{x}, T \cup \{j\}) = c_{i,j(T)} - c_{i,j}.$$

Since $c_{i,j(T)} \leq c_{i,j(S)}$, we have

$$h_i(\mathbf{x}, S) - h_i(\mathbf{x}, S \cup \{j\}) \geq h_i(\mathbf{x}, T) - h_i(\mathbf{x}, T \cup \{j\}).$$

2. $c_{i,j(T)} \leq c_{i,j} \leq c_{i,j(S)}$

In this case, having the additional facility j will reduce the cost for only disruption scenario S . Thus,

$$h_i(\mathbf{x}, S) - h_i(\mathbf{x}, S \cup \{j\}) \geq h_i(\mathbf{x}, T) - h_i(\mathbf{x}, T \cup \{j\}).$$

$$3. c_{i,j(T)} \leq c_{i,j(S)} \leq c_{i,j}$$

In this case, having the additional facility j will not reduce the cost for either disruption scenario. Thus,

$$h_i(\mathbf{x}, S) - h_i(\mathbf{x}, S \cup \{j\}) = h_i(\mathbf{x}, T) - h_i(\mathbf{x}, T \cup \{j\}).$$

The condition of increasing differences holds for all three cases. Thus, $h_i(\mathbf{x}, S)$ is supermodular in S for all $i \in \mathcal{I}$. For all customers, the total cost $h(\mathbf{x}, S) = \sum_i h_i(\mathbf{x}, S)$ is the sum of supermodular functions. Thus $h(\mathbf{x}, S)$ is also supermodular in S . \square

Proof of Lemma 2. We prove the result by confirming that the solution given in Lemma 2 is both primal feasible and dual feasible for the inner problem given in (2.3). It is easy to see that the solution is primal feasible, and has objective value

$$\phi = \sum_{s=0}^J (q_{s+1} - q_s) h(\mathbf{x}, \boldsymbol{\xi}^s).$$

The dual problem with dual variables $\boldsymbol{\lambda}$ is given by,

$$\begin{aligned} \min \quad & \sum_{j \in \mathcal{J}} (1 - q_j) \lambda_j \\ \text{s.t.} \quad & \sum_{j \in S} \lambda_j \geq h(\mathbf{x}, S), \quad \forall S \subseteq \mathcal{J}, 0 \in S \end{aligned}$$

Consider solution

$$\lambda_j = h(\mathbf{x}, \boldsymbol{\xi}^j) - h(\mathbf{x}, \boldsymbol{\xi}^{j-1}), \quad \forall j = 1, \dots, J,$$

and

$$\lambda_0 = h(\mathbf{x}, \boldsymbol{\xi}^0).$$

We will show this solution is dual feasible, i.e.,

$$\sum_{j \in S} \lambda_j \geq h(\mathbf{x}, S), \quad \forall S \subseteq \mathcal{J}.$$

For any $S \subseteq \mathcal{J}$, without loss of generality, assume $S = \{j_0, j_1, \dots, j_n\}$, where $0 = j_0 < j_1 < \dots < j_n$. Define disruption scenarios $\boldsymbol{\zeta}^0, \boldsymbol{\zeta}^1, \dots, \boldsymbol{\zeta}^n$, where $\zeta_j^s = I(j \in S \cap j \leq s)$. It is easy to see that $\boldsymbol{\zeta}^s \leq \boldsymbol{\xi}^s$ for all $s = 0, 1, \dots, J$. By property (2.4) of supermodular functions,

$$h(\mathbf{x}, \boldsymbol{\xi}^{j_k}) - h(\mathbf{x}, \boldsymbol{\xi}^{j_{k-1}}) \geq h(\mathbf{x}, \boldsymbol{\zeta}^{j_k}) - h(\mathbf{x}, \boldsymbol{\zeta}^{j_{k-1}}).$$

Also notice that $\boldsymbol{\xi}^{j_n} = \boldsymbol{\zeta}^{j_n}$ is the disruption scenario where all locations in S are online, and that $\boldsymbol{\xi}^{j_0} = \boldsymbol{\zeta}^{j_0}$ is the scenario where only location 0 is online. Thus,

$$\begin{aligned} \sum_{k=1}^n \lambda_{j_k} &= \sum_{k=1}^n [h(\mathbf{x}, \boldsymbol{\xi}^{j_k}) - h(\mathbf{x}, \boldsymbol{\xi}^{j_{k-1}})] \\ &\geq \sum_{k=1}^n [h(\mathbf{x}, \boldsymbol{\zeta}^{j_k}) - h(\mathbf{x}, \boldsymbol{\zeta}^{j_{k-1}})] = h(\mathbf{x}, S) - \lambda_0. \end{aligned}$$

Thus, the solution λ_j , $j = 0, 1, \dots, J$, is dual feasible. The corresponding dual objective value

$$\psi = \sum_{j=1}^J (1 - q_j) [h(x, \boldsymbol{\xi}^j) - h(x, \boldsymbol{\xi}^{j-1})] + h(x, \boldsymbol{\xi}^0) = \phi.$$

Thus, by duality, the solution given in Lemma 2 is optimal for the inner problem in (2.3), i.e., it is the worst-case distribution. \square

Proof of Lemma 4. Consider the Lagrangian of problem (2.6) where the capacity constraint is relaxed with multipliers $\boldsymbol{\mu} = (\mu_0, \dots, \mu_J)$:

$$L(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu}) = \sum_{j \in \mathcal{J}} x_j B_j \xi_j \mu_j + \min_{\mathbf{y}} \left\{ \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_i (c_{ij} - \mu_j) y_{ij} \left| \begin{array}{l} \sum_{j \in \mathcal{J}} y_{ij} = 1, \quad \forall i \in \mathcal{I} \\ y_{ij} \geq 0, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J} \end{array} \right. \right\}$$

We will first show that $L(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu})$ is supermodular in $(\boldsymbol{\xi}, \boldsymbol{\mu})$. By Theorem 2.3.4 in Simchi-Levi et al. (2004), the first term in $L(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu})$ is supermodular in $(\boldsymbol{\xi}, \boldsymbol{\mu})$. The second term is equal to $\sum_{i \in \mathcal{I}} d_i \min\{c_{ij} - \mu_j\}$. For each i , $\min\{c_{ij} - \mu_j\}$ is also supermodular in $\boldsymbol{\mu}$. Thus, $L(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu})$ is supermodular in $(\boldsymbol{\xi}, \boldsymbol{\mu})$. By strong duality, $h(\mathbf{x}, \boldsymbol{\xi}) = \max_{\boldsymbol{\mu} \leq 0} L(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu})$. By Proposition 2.3.5 in Simchi-Levi et al. (2004), $h^C(\mathbf{x}, \boldsymbol{\xi})$ is supermodular in $\boldsymbol{\xi}$. \square

Proof of Lemma 5. The proof is similar to the proof for Lemma 1. Without loss of generality, we consider the restricted set of nodes $\mathcal{V}(\mathbf{x}) = \{j \in \mathcal{V} : x_j = 1\}$, i.e., the set of opened facilities. Notice that the hub allocation problem is separable by O-D pairs. For O-D pair (i, i') , the cost is given by

$$h_{ii'}(\mathbf{x}, S) = d_{ii'} \min\{c_{ii'jj'} : j, j' \in \mathcal{V}(\mathbf{x})\}$$

For any $S \subset T \subset \mathcal{V}(\mathbf{x})$, consider the decrease in cost after adding $j \in \mathcal{V}(\mathbf{x}) \setminus T$ to the set. It is easy to see that the decrease for S is always greater than or equal to the decrease for T . Thus, $h(\mathbf{x}, S)$ is supermodular in S . \square

Proof of Proposition 7. The proof is based on the observation by Chu, Shanthikumar, and Shen (2008) that the optimal operational statistic coincides with a Bayesian framework. For the scale parameter, the Jeffrey's non-informative prior is given by $1/\theta$. Thus, the second term in the integrand in (4.1), i.e.,

$$\frac{1}{\theta^{n+1}} \prod_{i=1}^n f\left(\frac{X_i}{\theta}\right),$$

can be regarded as the posterior distribution (without normalization) of Θ after observing demand sample $\mathbf{X} = (X_1, \dots, X_n)$. Chu, Shanthikumar, and Shen (2008) showed that this posterior distribution is proper. The objective function in (4.1) can then be regarded as the posterior mean of the normalized profit

$$\frac{\phi(y; \Theta)}{\Theta}.$$

Let θ_0 denote the unknown true parameter, and let $\Theta|\mathbf{X}$ denote the posterior distribution of the random parameter Θ . For simplicity, let

$$g(y, \theta) = \frac{\phi(y; \theta)}{\theta}.$$

Let $M_n(y)$ denote the posterior mean, i.e., $M_n(y) = \mathbb{E}_{\Theta|\mathbf{X}}[g(y, \Theta)]$, and $M(y)$ the objective under the true parameter, i.e., $M(y) = g(y, \theta_0)$. Let \hat{y}_n denote the optimal operational statistic, i.e., $\hat{y}_n = \operatorname{argmax} M_n(y)$, and y^* the theoretical optimal solution with known parameter, i.e., $y^* = \operatorname{argmax} M(y)$. We need to show that $\hat{y}_n \xrightarrow{a.s.} y^*$, where $\xrightarrow{a.s.}$ denotes almost sure convergence.

Ibragimov and Has'Minskii (1981) showed that the posterior distribution $\Theta|\mathbf{X}$ is consistent, i.e., for any $\epsilon > 0$,

$$\mathbb{P}_{\Theta|\mathbf{X}}(\theta_0 - \epsilon, \theta_0 + \epsilon) \xrightarrow{a.s.} 1.$$

The result requires certain regularity conditions on the likelihood function, which hold in general for almost all parametric distributions. We do not list the conditions here. Interested readers may refer to Ibragimov and Has'Minskii (1981) for more details. The consistency of posterior distribution implies the point-wise convergence of $M_n(y)$ to $M(y)$, i.e., for any given $y > 0$,

$$M_n(y) \xrightarrow{a.s.} M(y).$$

Next we need to show that $M_n(y)$ is stochastically Lipschitz continuous in y , i.e.,

$$\sup_{y_1, y_2} |M_n(y_1) - M_n(y_2)| < K|y_1 - y_2|,$$

where K is a random variable that depends on \mathbf{X} and is almost surely bounded. The expected profit

$$\phi(y; \theta) = s \int_0^y \bar{F}\left(\frac{x}{\theta}\right) dx - cy.$$

Thus,

$$|\phi(y_1) - \phi(y_2)| = \left| s \int_{y_2}^{y_1} \bar{F}\left(\frac{x}{\theta}\right) dx - c(y_1 - y_2) \right| \leq (s + c) |y_1 - y_2|,$$

and,

$$\begin{aligned} |M_n(y_1) - M_n(y_2)| &\leq \int |\phi(y_1; \theta) - \phi(y_2; \theta)| \frac{1}{\theta^{n+2}} \prod_{i=1}^n f(x_i; \theta) d\theta \\ &\leq (s + c) \int \frac{1}{\theta^{n+2}} \prod_{i=1}^n f(x_i; \theta) d\theta |y_1 - y_2|. \end{aligned}$$

Notice that

$$\int \frac{1}{\theta^{n+2}} \prod_{i=1}^n f(x_i; \theta) d\theta = \mathbb{E}_{\Theta|\mathbf{X}} [\Theta^{-1}].$$

By continuity and consistency of the posterior distribution, $\mathbb{E}_{\Theta|\mathbf{X}}[\Theta^{-1}]$ is almost surely bounded. Thus $M_n(y)$ is stochastically Lipschitz continuous. Point-wise convergence and Lipschitz continuity then imply that $M_n(y)$ converges uniformly to $M(y)$, i.e.,

$$\sup_y M_n(y) \xrightarrow{a.s.} M(y).$$

Using uniform convergence, we can show $M_n(\hat{y}_n) \xrightarrow{a.s.} M(y^*)$. By the definition of \hat{y}_n and y^* , we have $M_n(\hat{y}_n) \geq M_n(y^*)$ and $M(y^*) \geq M(\hat{y}_n)$. Thus,

$$\begin{aligned} |M_n(\hat{y}_n) - M(y^*)| &\leq \max\{|M_n(y^*) - M(y^*)|, |M_n(\hat{y}_n) - M(\hat{y}_n)|\} \\ &\leq \sup_y |M_n(y) - M(y)| \xrightarrow{a.s.} 0, \end{aligned}$$

and,

$$|M(\hat{y}_n) - M(y^*)| \leq |M(\hat{y}_n) - M_n(\hat{y}_n)| + |M_n(\hat{y}_n) - M(y^*)| \xrightarrow{a.s.} 0$$

Thus,

$$M(\hat{y}_n) \xrightarrow{a.s.} M(y^*).$$

We have also shown that the objective function $M_n(y)$ is strictly concave. This guarantees the convergence of their maxima, $\hat{y}_n \xrightarrow{a.s.} y^*$. \square

Proof of Proposition 8. We prove the result by showing that the separated estimation-optimization approach using MLE is actually a special case of operational statistics in class \mathcal{H}_1^e . Recall that the demand is given by $X = \tau + \theta Z$. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be n i.i.d. demand observations. The MLE for parameters τ and θ is given by

$$(\tau^*(\mathbf{X}), \theta^*(\mathbf{X})) = \operatorname{argmax}_{\tau, \theta} \left\{ \frac{1}{\theta^n} \prod_{i=1}^n f\left(\frac{X_i - \tau}{\theta}\right) \right\}.$$

The optimal order quantity in the separated estimation-optimization approach is then given by

$$\hat{y}(\mathbf{X}) = y^*(\tau^*(\mathbf{X}), \theta^*(\mathbf{X})) = \tau^*(\mathbf{X}) + \theta^*(\mathbf{X})F^{-1}\left(\beta\left(1 - \frac{c}{s}\right)\right).$$

Now consider another demand sample $\tilde{\mathbf{X}} = \alpha\mathbf{X} - \delta$, where $\alpha > 0$. The MLE based on these observations is given by

$$(\tau^*(\tilde{\mathbf{X}}), \theta^*(\tilde{\mathbf{X}})) = \operatorname{argmax}_{\tau, \theta} \left\{ \frac{1}{\theta^n} \prod_{i=1}^n f\left(\frac{\alpha X_i - \delta - \tau}{\theta}\right) \right\}.$$

It is easy to see that

$$(\tau^*(\alpha\mathbf{X} - \delta), \theta^*(\alpha\mathbf{X} - \delta)) = (\alpha\tau^*(\mathbf{X}) - \delta, \alpha\theta^*(\mathbf{X})).$$

The corresponding optimal order quantity is given by

$$\hat{y}(\alpha \mathbf{X} - \delta) = \alpha \tau^*(\mathbf{X}) - \delta + \alpha \theta^*(\mathbf{X}) F^{-1} \left(\beta \left(1 - \frac{c}{s} \right) \right) = \alpha \hat{y}(\mathbf{X}) - \delta.$$

Thus $\hat{y}(\mathbf{X}) \in \mathcal{H}_1^e$. Since the optimal operational statistic $h^*(\mathbf{X})$ achieves the highest a priori expected profit in class \mathcal{H}_1^e , it must achieve a higher a priori expected profit than $\hat{y}(\mathbf{X})$, which is a special case of operational statistics in class \mathcal{H}_1^e . \square

Proof of Proposition 9. Recall that the demand is from a location-scale family, i.e., $X = \tau + \theta Z$. It is easy to see that for any function $g \in \mathcal{H}_1^e$,

$$\mathbb{E}[\phi(g(\mathbf{X}); \tau, \theta)] = \theta \mathbb{E}[\phi(g(\mathbf{Z}); 0, 1)] + (s - c)\tau.$$

As we have shown, both the optimal operational statistic and the optimal solution from the separated estimation-optimization approach using MLE are in class \mathcal{H}_1^e . Let $h^*(\mathbf{X})$ denote the optimal operational statistic, and $\hat{y}(\mathbf{X})$ the optimal solution from the separated estimation-optimization approach. Then the relative benefit of using operational statistics is given by

$$\begin{aligned} & \frac{\mathbb{E}[\phi(h^*(\mathbf{X}); \tau, \theta)] - \mathbb{E}[\phi(\hat{y}(\mathbf{X}); \tau, \theta)]}{\mathbb{E}[\phi(\hat{y}(\mathbf{X}); \tau, \theta)]} \\ &= \frac{\theta (\mathbb{E}[\phi(h^*(\mathbf{Z}); 0, 1)] - \mathbb{E}[\phi(\hat{y}(\mathbf{Z}); 0, 1)])}{\theta \mathbb{E}[\phi(\hat{y}(\mathbf{Z}); 0, 1)] + (s - c)\tau} \\ &= \frac{\mathbb{E}[\phi(h^*(\mathbf{Z}); 0, 1)] - \mathbb{E}[\phi(\hat{y}(\mathbf{Z}); 0, 1)]}{\mathbb{E}[\phi(\hat{y}(\mathbf{Z}); 0, 1)] + (s - c)\frac{\tau}{\theta}} \end{aligned}$$

Thus, the relative benefit of using operational statistics increases in θ/τ . Without loss of generality, we can assume $\mathbb{E}[Z] = 0$ and $\sqrt{V[Z]} = 1$. Thus $\mathbb{E}[X] = \tau$ and $\sqrt{V[X]} = \theta$, and the coefficient of variation of demand is equal to θ/τ . This completes the proof. \square

Proof of Proposition 10. Using the dual representation of CVaR,

$$y^*(\tau, \theta) = \operatorname{argmax}_y \left\{ \sup_{\eta} G(\eta, y) \right\},$$

where

$$\begin{aligned} G(\eta, y) &= \eta - \frac{1}{\beta} \int [\eta - \Phi(y, x)]^+ f(x; \tau, \theta) dx \\ &= \eta - \frac{1}{\beta} \int [\eta - s \min\{y, x\} + cy]^+ f(x; \tau, \theta) dx \\ &= \eta - \frac{1}{\beta} \left[\int_0^y [\eta - sx + cy]^+ f(x; \tau, \theta) dx + [\eta - sy + cy]^+ \int_y^\infty f(x; \tau, \theta) dx \right] \end{aligned} \tag{A.1}$$

The function G given in (A.1) is concave. For a given y , it is piecewise differentiable in η . Consider the following three cases:

1. For $\eta \leq -cy$, $G(\eta, y) = \eta$, $\partial G/\partial \eta = 1$.

2. For $-cy < \eta \leq (s-c)y$,

$$\begin{aligned} G(\eta, y) &= \eta - \frac{1}{\beta} \int_0^y [\eta - sx + cy]^+ f(x; \tau, \theta) dx \\ &= \eta - \frac{1}{\beta} \int_0^{\frac{\eta+cy}{s}} (\eta - sx + cy) f(x; \tau, \theta) dx. \end{aligned}$$

Thus,

$$\frac{\partial G}{\partial \eta} = 1 - \frac{1}{\beta} F\left(\frac{\eta + cy}{s}; \tau, \theta\right).$$

In particular,

$$\left. \frac{\partial G}{\partial \eta} \right|_{\eta=-cy} = 1, \quad \left. \frac{\partial G}{\partial \eta} \right|_{\eta=(s-c)y} = 1 - \frac{1}{\beta} F(y; \tau, \theta).$$

3. For $(s-c)y < \eta$,

$$G(\eta, y) = \eta - \frac{1}{\beta} \int_0^y (\eta - sx + cy) f(x; \tau, \theta) dx - \frac{1}{\beta} (\eta - (s-c)y) [1 - F(y; \tau, \theta)].$$

Thus,

$$\frac{\partial G}{\partial \eta} = 1 - \frac{1}{\beta} < 0.$$

Let $\eta^*(y)$ be the optimal η that maximizes $G(\eta, y)$ for a given y . Based on the three cases given above, we have $\eta^*(y) \in (-cy, (s-c)y]$.

For the second case, we have the stationary point $\hat{\eta}(y) = sF^{-1}(\beta; \tau, \theta) - cy$. Observe that the stationary point may not be feasible. If $y \geq F^{-1}(\beta; \tau, \theta)$, then $\hat{\eta}(y) \in (-cy, (s-c)y]$. Thus, $\eta^*(y) = \hat{\eta}(y)$, and the objective function,

$$G(\eta^*(y), y) = sF^{-1}(\beta; \tau, \theta) - cy - \frac{1}{\beta} \int_0^{F^{-1}(\beta; \tau, \theta)} s(F^{-1}(\beta; \tau, \theta) - x) f(x; \tau, \theta) dx.$$

Taking the derivative of the above expression w.r.t. y , we obtain

$$\frac{d}{dy} G(\eta^*(y), y) = -c < 0.$$

Thus, $F^{-1}(\beta; \tau, \theta)$ is optimal in this region.

If $y \leq F^{-1}(\beta; \tau, \theta)$, the $\hat{\eta} \geq (s-c)y$, and $\partial G/\partial \eta \geq 0$. Thus, $\eta^*(y) = (s-c)y$, and the CVaR of profit can be expressed as

$$\rho_\beta[y; \tau, \theta] = G(\eta^*(y), y) = (s-c)y - \frac{1}{\beta} \int_0^y s(y-x) f(x; \tau, \theta) dx.$$

Taking the derivative of CVaR w.r.t. y , we have

$$\frac{d}{dy}G(\eta^*(y), y) = s - c - \frac{s}{\beta}F(y; \tau, \theta).$$

In particular,

$$\left. \frac{d}{dy}G(\eta^*(y), y) \right|_{y=F^{-1}(\beta; \tau, \theta)} = -c < 0.$$

Thus,

$$y^*(\tau, \theta) = F^{-1}\left(\beta\left(1 - \frac{c}{s}\right); \tau, \theta\right).$$

□

Proof of Proposition 13. First, we will show that a uniform optimal operational statistic exists for all parameters. Since $\mathbf{X} = \theta\mathbf{Z}$ and $X_{n+1} = \theta Z_{n+1}$, for $g \in \mathcal{H}_1$, we have

$$\Phi(g(\mathbf{X}), X_{n+1}) = \theta\Phi(g(\mathbf{Z}), Z_{n+1}).$$

By the positive homogeneity of coherent risk measures, we have

$$\rho_\beta[\Phi(g(\mathbf{X}), X_{n+1})] = \theta\rho_\beta[\Phi(g(\mathbf{Z}), Z_{n+1})].$$

Thus, for any $g_1, g_2 \in \mathcal{H}_1$, either

$$\rho_\beta[\Phi(g_1(\mathbf{X}), X_{n+1})] \geq \rho_\beta[\Phi(g_2(\mathbf{X}), X_{n+1})],$$

or

$$\rho_\beta[\Phi(g_1(\mathbf{X}), X_{n+1})] \leq \rho_\beta[\Phi(g_2(\mathbf{X}), X_{n+1})],$$

for all θ . In other words, if g_1 is better than g_2 for one parameter, then g_1 is also better than g_2 for all parameters.

Since the optimality of operational statistics is uniform for all parameters, we can focus on one specific parameter, e.g., $\theta = 1$. Using the dual representation of CVaR, i.e.,

$$\rho_\beta[\Phi(g(\mathbf{Z}), Z_{n+1})] = \sup_\eta \left\{ \int \int \left[\eta - \frac{1}{\beta} [\eta - \Phi(g(\mathbf{z}), z_{n+1})]^+ \right] \prod_{i=1}^{n+1} f(z_i) dz_{n+1} d\mathbf{z} \right\},$$

We need to choose g and η to maximize the objective function,

$$\int \int \left[\eta - \frac{1}{\beta} [\eta - \Phi(g(\mathbf{z}), z_{n+1})]^+ \right] \prod_{i=1}^{n+1} f(z_i) dz_{n+1} d\mathbf{z}$$

In order to find an operational statistic within the class \mathcal{H}_1 , we change the objective function into a path integral over

$$\{\mathbf{z} = r\mathbf{w} : \|\mathbf{w}\|_2 = 1, r \geq 0\}.$$

The path integral objective is given by

$$\int_{\mathbf{w}: \|\mathbf{w}\|_2=1} \int_0^\infty G(r, \mathbf{w}) r^{n-1} \prod_{i=1}^n f(rw_i) dr d\mathbf{w},$$

where

$$G(r, \mathbf{w}) = \int \left[\eta - \frac{1}{\beta} [\eta - \Phi(g(r\mathbf{w}), z_{n+1})]^+ \right] f(z_{n+1}) dz_{n+1}$$

Applying the change of variable $\sigma = \frac{1}{r}$, the path integral objective becomes

$$\int_{\mathbf{w}: \|\mathbf{w}\|_2=1} \int_0^\infty G\left(\frac{1}{\sigma}, \mathbf{w}\right) f(z_{n+1}) \frac{1}{\sigma^{n+1}} \prod_{i=1}^n f\left(\frac{w_i}{\sigma}\right) dz_{n+1} d\sigma d\mathbf{w}.$$

Applying another change of variable $w_{n+1} = \sigma z_{n+1}$, we obtain

$$G\left(\frac{1}{\sigma}, \mathbf{w}\right) = \int \left[\eta - \frac{1}{\beta} \left[\eta - \Phi\left(g\left(\frac{\mathbf{w}}{\sigma}, \frac{w_{n+1}}{\sigma}\right)\right)^+ \right] \right] \frac{1}{\sigma} f\left(\frac{w_{n+1}}{\sigma}\right) dw_{n+1}.$$

By the properties of the profit function $\Phi(y, X)$ and the class \mathcal{H}_1 ,

$$\Phi\left(g\left(\frac{\mathbf{w}}{\sigma}, \frac{w_{n+1}}{\sigma}\right), \frac{w_{n+1}}{\sigma}\right) = \frac{\Phi(g(\mathbf{w}), w_{n+1})}{\sigma}.$$

Thus, the path integral objective becomes

$$\int_{\mathbf{w}: \|\mathbf{w}\|_2=1} \int_{\sigma>0} \int_{z_{n+1}} \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(g(\mathbf{w}), w_{n+1})}{\sigma} \right]^+ \right] \frac{1}{\sigma^{n+2}} \prod_{i=1}^{n+1} f\left(\frac{w_i}{\sigma}\right) dw_{n+1} dr d\mathbf{w}.$$

Using pointwise maximization of the path integral objective for a given demand sample \mathbf{w} , we will have exactly the same result as $h^*(\cdot)$ in (4.8).

Finally, we need to show that the pointwise solution in (4.8) indeed belongs to the class \mathcal{H}_1 , i.e., it is positive homogeneous of degree one. Consider $\mathbf{x} = \theta\mathbf{w}$, for some $\theta > 0$. The operational statistic is given by,

$$\begin{aligned} & h^*(\theta\mathbf{w}) \\ &= \operatorname{argmax}_y \left\{ \sup_{\eta} \left\{ \int_0^\infty \int \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(y, x_{n+1})}{\sigma} \right]^+ \right] \frac{1}{\sigma^{n+2}} f\left(\frac{x_{n+1}}{\sigma}\right) \prod_{i=1}^n f\left(\frac{\theta w_i}{\sigma}\right) dx_{n+1} d\sigma \right\} \right\}. \end{aligned}$$

Applying change of variable $\sigma' = \frac{1}{\theta}\sigma$, $y' = \frac{1}{\theta}y$, and $w_{n+1} = \frac{1}{\theta}x_{n+1}$, we have

$$\begin{aligned} & h^*(\theta\mathbf{w}) \\ &= \operatorname{argmax}_y \left\{ \sup_{\eta} \left\{ \int_0^\infty \int \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(y, \theta w_{n+1})}{\theta\sigma'} \right]^+ \right] \frac{1}{\theta^n} \frac{1}{(\sigma')^{n+2}} \prod_{i=1}^{n+1} f\left(\frac{w_i}{\sigma'}\right) dw_{n+1} d\sigma' \right\} \right\} \\ &= \theta \operatorname{argmax}_{y'} \left\{ \sup_{\eta} \left\{ \int_0^\infty \int \left[\eta - \frac{1}{\beta} \left[\eta - \frac{\Phi(y', w_{n+1})}{\sigma'} \right]^+ \right] \frac{1}{(\sigma')^{n+2}} \prod_{i=1}^{n+1} f\left(\frac{w_i}{\sigma'}\right) dw_{n+1} d\sigma' \right\} \right\} \\ &= \theta h^*(\mathbf{w}) \end{aligned}$$

Thus, $g^*(\cdot)$ is in class \mathcal{H}_1 . This completes the proof.

□

Appendix B

Data and Supplementary Results

B.1 Severe Weather Hazard Data for the Supply Chain Design Example

In Section 2.4, we presented a supply chain design example motivated by a real-world situation. The marginal disruption probability is estimated using the severe weather hazard probability data provided by the NOAA Storm Prediction Center (SPC). The original data we obtain contains the probability that at least one significant hazard (e.g., tornadoes, windstorms, hails, etc.) occurs within 25 miles of any node on a 50-mile grid in one calendar year. For a given geographic coordinate, the probability is found by interpolation on this grid. Given the yearly probability p_Y , we first convert it to an approximate quarterly probability p_Q , using $p_Q = 1 - (1 - p_Y)^{25}$. Notice that this is the probability that a hazard will occur within 25 miles of the facility. We assume that when a hazard occurs, there will be a 50% chance that the facility will actually be disrupted and remain closed for the entire quarter. The final estimated disruption probability is shown in Table B.1. Notice that we do not assume the worst-case distribution given in Lemma 2, or the disruption propagation effect that we used to illustrate Lemma 2. We only use the severe weather hazard data to estimate the marginal disruption probabilities.

B.2 Selected Reliable Facility Location Results Using the 88-Node Data Set

The numerical results in Table 2.5 are based on the 49-node data set in Daskin (1995). We also compare the two models using the 88-node data set in Daskin (1995), which includes Washington, D.C., the state capitals and the 50 largest cities in the contiguous United States minus duplicates. Both data sets are available at http://coral.ie.lehigh.edu/~larry/wp-content/datasets/RPMP/RPMP_data.zip by courtesy of Professor Larry Snyder. As shown in Table B.2, we have observations that are similar to those from the 49-node data

Table B.1: Estimated marginal disruption probability for the supply chain design example.

Location	q	Location	q	Location	q
CA	1.45	WA	0.79	UT	1.28
NY	3.73	MD	6.68	NE	12.34
TX	7.95	MN	8.57	NM	1.59
FL	7.49	LA	10.14	ME	0.93
PA	6.96	AL	10.80	NV	0.83
IL	11.32	KY	8.78	NH	2.08
OH	7.08	AZ	1.24	ID	0.88
MI	6.99	SC	8.26	RI	2.01
NJ	5.43	CO	8.41	MT	0.92
NC	7.72	CT	3.34	SD	5.99
GA	9.93	OK	13.02	DE	5.42
VA	7.02	OR	0.97	ND	6.65
MA	1.76	IA	12.88	DC	6.91
IN	10.62	MS	13.43	VT	1.55
MO	9.47	KS	11.42	WY	7.71
WI	8.74	AR	11.20		
TN	10.31	WV	3.45		

Note: q : marginal disruption probability

set. The main difference is that the regret of the traditional model assuming independent disruptions is now much lower. The reason is that the transportation cost in the 88-node data set is much higher. As a result, more facilities are opened, and the regret from ignoring correlation is smaller. Nonetheless, the regret and relative regret from the traditional model is still much higher than that of the robust model.

Table B.2: Selected results for the 88-node data set.

α	ω	θ	n_R	z_R	Δz_R	$\% \Delta z_R$	CPU $_R$	GAP $_R$	n_I	z_I	Δz_I	$\% \Delta z_I$	CPU $_I$	GAP $_I$
0.1	20000	200	9	12.13	0.00	0.00	27.57	0.00	9	12.10	0.00	0.00	595.28	0.03
0.2	20000	200	9	12.19	0.00	0.00	19.00	0.00	9	12.13	0.00	0.00	2251.65	0.06
0.3	20000	200	9	12.24	0.00	0.00	27.53	0.00	9	12.17	0.00	0.00	7244.31	0.21
0.1	40000	200	9	12.14	0.00	0.00	26.42	0.00	9	12.10	0.00	0.00	602.19	0.03
0.2	40000	200	9	12.19	0.00	0.00	24.60	0.00	9	12.13	0.00	0.00	2286.13	0.06
0.3	40000	200	9	12.25	0.00	0.00	29.15	0.00	9	12.17	0.00	0.00	7319.25	0.21
0.1	80000	200	9	12.14	0.00	0.00	24.86	0.00	9	12.10	0.00	0.00	595.25	0.03
0.2	80000	200	9	12.20	0.00	0.00	27.75	0.00	9	12.13	0.00	0.00	2274.91	0.06
0.3	80000	200	9	12.26	0.00	0.00	23.51	0.00	9	12.17	0.00	0.00	7318.49	0.21
0.1	20000	400	9	13.07	0.00	0.00	18.33	0.00	9	12.23	0.00	0.00	7449.17	0.81
0.2	20000	400	9	13.59	0.00	0.03	16.35	0.00	9	12.34	0.02	0.13	7368.53	2.16
0.3	20000	400	9	14.09	0.02	0.17	10.67	0.00	9	12.44	0.00	0.01	7596.53	3.49
0.1	40000	400	9	13.59	0.00	0.00	21.15	0.00	9	12.23	0.00	0.00	7581.20	0.81
0.2	40000	400	9	14.37	0.00	0.03	14.39	0.00	9	12.34	0.02	0.13	7479.55	2.16
0.3	40000	400	9	15.10	0.05	0.34	13.77	0.00	9	12.44	0.20	1.59	7219.25	3.57
0.1	80000	400	9	14.60	0.03	0.23	27.99	0.00	9	12.23	0.20	1.60	7549.96	0.81
0.2	80000	400	9	15.78	0.15	0.97	18.97	0.00	9	12.34	0.21	1.72	7428.50	2.16
0.3	80000	400	9	16.95	0.29	1.70	14.97	0.00	9	12.44	0.20	1.59	7671.36	3.49
0.1	20000	800	9	19.93	0.16	0.82	16.22	0.00	9	12.48	0.20	1.59	7428.52	2.87
0.2	20000	800	9	23.79	0.34	1.44	15.64	0.00	9	12.72	0.22	1.75	7342.71	5.31
0.3	20000	800	9	27.63	0.54	1.96	13.44	0.00	9	12.97	0.21	1.63	7313.85	7.86
0.1	40000	800	9	26.37	0.56	2.13	16.56	0.00	9	12.48	0.20	1.59	7449.96	2.87
0.2	40000	800	9	33.44	0.94	2.81	17.58	0.00	9	12.72	0.22	1.75	7374.47	5.31
0.3	40000	800	9	40.50	1.34	3.30	18.81	0.00	9	12.97	0.21	1.63	7374.53	7.86
0.1	80000	800	9	39.24	1.36	3.46	19.35	0.00	9	12.48	0.20	1.59	7414.98	2.87
0.2	80000	800	9	52.75	2.14	4.05	17.19	0.00	9	12.72	0.22	1.75	7318.18	5.31
0.3	80000	800	9	66.25	2.93	4.43	17.82	0.00	9	12.97	0.21	1.63	7324.58	7.86
Average			9	21.14	0.40	1.90	19.98	0.00	9	12.40	0.10	0.82	6080.49	2.54

Note: R : robust model, I : independent model, n : number of facilities, z : expected cost ($\times 10^5$), Δz : regret ($\times 10^5$), $\% \Delta z$: relative regret (%), CPU: computation time (s), GAP: optimality gap at termination (%).

Bibliography

- Aboolian, Robert, Tingting Cui, and Zuo-Jun Max Shen (2013). “An efficient approach for solving reliable facility location models”. *INFORMS Journal on Computing* 25.4, pp. 720–729.
- Agrawal, Shipra, Yichuan Ding, Amin Saberi, and Yinyu Ye (2010). “Correlation robust stochastic optimization”. In: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '10. Austin, Texas, pp. 1087–1096.
- (2012). “Price of correlations in stochastic optimization”. *Operations Research* 60.1, p-p. 150–162.
- Ahmed, Shabbir, Ulas Cakmak, and Alexander Shapiro (2007). “Coherent risk measures in inventory problems”. *European Journal of Operational Research* 182.1, pp. 226–238.
- Akcay, Alp, Bahar Biller, and Sridhar Tayur (2011). “Improved inventory targets in the presence of limited historical demand data”. *Manufacturing & Service Operations Management* 13.3, pp. 297–309.
- An, Yu, Yu Zhang, and Bo Zeng (2011). “The reliable hub-and-spoke design problem: Models and algorithms”. Working paper. University of South Florida, Tampa, FL.
- An, Yu, Bo Zeng, Yu Zhang, and Long Zhao (2012). “Reliable p -median facility location problem: Two-stage robust models and algorithms”. Working paper. University of South Florida, Tampa, FL.
- Arcelus, F.J., Satyendra Kumar, and G. Srinivasan (2012). “Risk tolerance and a retailer’s pricing and ordering policies within a newsvendor framework”. *Omega* 40.2, pp. 188–198.
- Aydin, Goker and Evan L. Porteus (2008). “Joint inventory and pricing decisions for an assortment”. *Operations Research* 56.5, pp. 1247–1255.
- Barlow, Richard E., David J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972). *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley Series in Probability and Mathematical Statistics. John Wiley.
- Bayarri, M. J. and James O. Berger (2004). “The interplay of Bayesian and frequentist analysis”. *Statistical Science* 19.1, pp. 58–80.
- Berman, Oded, Dmitry Krass, and Mozart B. C. Menezes (2007). “Facility reliability issues in network p -median problems: Strategic centralization and co-location effects”. *Operations Research* 55.2, pp. 332–350.

- Berman, Oded, Dmitry Krass, and Mozart B.C. Menezes (2013). “Location and reliability problems on a line: Impact of objectives and correlated failures on optimal location patterns”. *Omega* 41.4, pp. 766–779.
- Bertsimas, Dimitris, David B. Brown, and Constantine Caramanis (2011). “Theory and applications of robust optimization”. *SIAM Review* 53.3, pp. 464–501.
- Bertsimas, Dimitris, Karthik Natarajan, and Chung-Piaw Teo (2004). “Probabilistic combinatorial optimization: moments, semidefinite programming, and asymptotic bounds”. *SIAM Journal on Optimization* 15.1, pp. 185–209.
- Bertsimas, Dimitris and Aurélie Thiele (2005). *A data-driven approach to newsvendor problems*. Tech. rep. Technical report, Massachusetts Institute of Technology.
- Bertsimas, Dimitris and John N. Tsitsiklis (1997). *Introduction to Linear Optimization*. Athena Scientific.
- Besbes, Omar and Assaf Zeevi (2009). “Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms”. *Operations Research* 57.6, pp. 1407–1420.
- (2012). “Blind network revenue management”. *Operations Research* 60.6, pp. 1537–1550.
- Best, Michael J. (1996). “An algorithm for the solution of the parametric quadratic programming problem”. In: *Applied Mathematics and Parallel Computing*. Ed. by Herbert Fischer, Bruno Riedäller, and Stefan Schäffler. Physica-Verlag HD, pp. 57–76.
- Burnetas, Apostolos N. and Craig E. Smith (2000). “Adaptive ordering and pricing for perishable products”. *Operations Research* 48.3, pp. 436–443.
- Chan, Lap Mui Ann, Zuo-Jun Max Shen, David Simchi-Levi, and Julie L Swann (2004). “Coordination of pricing and inventory decisions: A survey and classification”. In: *Handbook of Quantitative Supply Chain Analysis: Modeling in the e-Business Era*. Ed. by David Simchi-Levi, S. David Wu, and Zuo-Jun Max Shen. Kluwer Academic Publishers. Chap. 9, pp. 335–392.
- Chen, Fangruo and Awi Federgruen (2000). “Mean-variance analysis of basic inventory models”. Technical report, Columbia University.
- Chen, Qi, Xiaopeng Li, and Yanfeng Ouyang (2011). “Joint inventory-location problem under the risk of probabilistic facility disruptions”. *Transportation Research Part B: Methodological* 45.7, pp. 991–1003.
- Chen, Xin and David Simchi-Levi (2012). “Pricing and inventory management”. In: *The Oxford Handbook of Pricing Management*. Ed. by Ozalp Ozer and Robert Phillips. Oxford University Press. Chap. 30, pp. 784–822.
- Chen, Youhua (Frank), Minghui Xu, and Zhe George Zhang (2009). “Technical Note – A risk-averse newsvendor model under the CVaR criterion”. *Operations Research* 57.4, pp. 1040–1044.
- Cherkassky, Vladimir and Yunqian Ma (2004). “Practical selection of SVM parameters and noise estimation for SVM regression”. *Neural Networks* 17.1, pp. 113–126.
- Choi, Sungyong and Andrzej Ruszczyński (2008). “A risk-averse newsvendor with law invariant coherent measures of risk”. *Operations Research Letters* 36.1, pp. 77–82.

- Choi, Sungyong, Andrzej Ruszczyński, and Yao Zhao (2011). “A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk”. *Operations Research* 59.2, pp. 346–364.
- Chu, Leon Yang, J. George Shanthikumar, and Zuo-Jun Max Shen (2008). “Solving operational statistics via a Bayesian analysis”. *Operations Research Letters* 36.1, pp. 110–116.
- (2009). “Pricing and revenue management with operational statistics”. Working paper, University of California, Berkeley.
- Church, Richard L. and Maria P. Scaparra (2007). “Protecting critical assets: The r -interdiction median problem with fortification”. *Geographical Analysis* 39.2, pp. 129–146.
- Church, Richard L., Maria P. Scaparra, and Richard S. Middleton (2004). “Identifying critical infrastructure: The median and covering facility interdiction problems”. *Annals of the Association of American Geographers* 94.3, pp. 491–502.
- Cryer, Jonathan D., Tim Robertson, F. T. Wright, and Robert J. Casady (1972). “Monotone median regression”. *The Annals of Mathematical Statistics* 43.5, pp. 1459–1469.
- Cui, Tingting, Yanfeng Ouyang, and Zuo-Jun Max Shen (2010). “Reliable facility location design under the risk of disruptions”. *Operations Research* 58.4, pp. 998–1011.
- Daskin, Mark S. (1995). *Network and Discrete Location*. John Wiley & Sons, Inc.
- Edmonds, Jack (1971). “Matroids and the greedy algorithm”. *Mathematical Programming* 1.1, pp. 127–136.
- Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger (1995). “The risk-averse (and prudent) newsboy”. *Management Science* 41.5, pp. 786–794.
- Elmaghraby, Wedad and Pinar Keskinocak (2003). “Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions”. *Management Science* 49.10, pp. 1287–1309.
- Federgruen, Awi and Aliza Heching (1999). “Combined pricing and inventory control under uncertainty”. *Operations Research* 47.3, pp. 454–475.
- Feng, Qi, Sirong Luo, and Dan Zhang (2014). “Dynamic inventory-pricing control under backorder: Demand estimation and policy optimization”. *Manufacturing & Service Operations Management* 16.1, pp. 149–160.
- Gallego, Guillermo and Ilkyeong Moon (1993). “The distribution free newsboy problem: Review and extensions”. *The Journal of the Operational Research Society* 44.8, pp. 825–834.
- Gallego, Guillermo and Garrett van Ryzin (1994). “Optimal dynamic pricing of inventories with stochastic demand over finite horizons”. *Management Science* 40.8, pp. 999–1020.
- Gotoh, Jun-Ya and Yuichi Takano (2007). “Newsvendor solutions via conditional value-at-risk minimization”. *European Journal of Operational Research* 179.1, pp. 80–96.
- Hayes, Robert H. (1969). “Statistical estimation problems in inventory control”. *Management Science* 15.11, pp. 686–701.
- Huh, Woonghee Tim and Paat Rusmevichientong (2009). “A nonparametric asymptotic analysis of inventory planning with censored demand”. *Mathematics of Operations Research* 34.1, pp. 103–123.

- Huh, Woonghee Tim, Retsef Levi, Paat Rusmevichientong, and James B. Orlin (2011). “Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator”. *Operations Research* 59.4, pp. 929–941.
- Ibragimov, I. A. and R. Z. Has’Minskii (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.
- Iglehart, Donald L. (1964). “The dynamic inventory problem with unknown demand distribution”. *Management Science* 10.3, pp. 429–440.
- Jain, Ankit, Andrew E. B. Lim, J. George Shanthikumar, and Lian Yu (2011). “Objective operational learning with application to newsvendor problem”. Working paper. University of California, Berkeley.
- Karlin, Samuel and Charles R. Carr (1962). “Prices and optimal inventory policy”. In: *Studies in Applied Probability and Management Science*. Ed. by Kenneth Arrow, Samuel Karlin, and Herbert Scarf. Stanford University Press, Stanford, CA, pp. 159–172.
- Kocabiyıkoğlu, Ayşe and Ioana Popescu (2011). “An elasticity approach to the newsvendor with price-sensitive demand”. *Operations Research* 59.2, pp. 301–312.
- Koenker, Roger and Gilbert Bassett Jr. (1978). “Regression quantiles”. *Econometrica* 46.1, pp. 33–50.
- Koenker, Roger, Pin Ng, and Stephen Portnoy (1994). “Quantile smoothing splines”. *Biometrika* 81.4, pp. 673–680.
- Levi, Retsef, Gerogia Perakis, and Joline Uichanco (2012). “The data-driven newsvendor problem: New bounds and insights”. Working paper, Massachusetts Institute of Technology.
- Levi, Retsef, Robin O. Roundy, and David B. Shmoys (2007). “Provably near-optimal sampling-based policies for stochastic inventory control models”. *Mathematics of Operations Research* 32.4, pp. 821–839.
- Li, Qingwei, Bo Zeng, and Alex Savachkin (2013). “Reliable facility location design under disruptions”. *Computers & Operations Research* 40.4, pp. 901–909.
- Li, Xiaopeng and Yanfeng Ouyang (2010). “A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions”. *Transportation Research Part B: Methodological* 44.4, pp. 535–548.
- (2011). “Reliable sensor deployment for network traffic surveillance”. *Transportation Research Part B: Methodological* 45.1, pp. 218–231.
- (2012). “Reliable traffic sensor deployment under probabilistic disruptions and generalized surveillance effectiveness measures”. *Operations Research* 60.5, pp. 1183–1198.
- Li, Xiaopeng, Yanfeng Ouyang, and Fan Peng (2013). “A supporting station model for reliable infrastructure location design under interdependent disruptions”. *Transportation Research Part E: Logistics and Transportation Review*. Forthcoming.
- Liang, Yong, Zuo-Jun Max Shen, and Ye Xu (2013). “Flexible hub location model for air transportation networks”. Working paper. University of California, Berkeley, CA.
- Liberatore, Federico, Maria P. Scaparra, and Mark S. Daskin (2012). “Hedging against disruptions with ripple effects in location analysis”. *Omega* 40.1, pp. 21–30.

- Lim, Andrew E. B., J. George Shanthikumar, and Zuo-Jun Max Shen (2006). “Model uncertainty, robust optimization, and learning”. In: *TutORials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*. INFORMS. Chap. 3, pp. 66–94.
- Lim, Michael K., Mark S. Daskin, Achal Bassamboo, and Sunil Chopra (2010). “A facility reliability problem: Formulation, properties, and algorithm”. *Naval Research Logistics* 57.1, pp. 58–70.
- Lim, Michael K., Achal Bassamboo, Sunil Chopra, and Mark S. Daskin (2013). “Facility location decisions with random disruptions and imperfect estimation”. *Manufacturing & Service Operations Management* 15.2, pp. 239–249.
- Liyanage, Liwan H. and J. George Shanthikumar (2005). “A practical inventory control policy using operational statistics”. *Operations Research Letters* 33.4, pp. 341–348.
- Losada, Chaya, M. Paola Scaparra, Richard L. Church, and Mark S. Daskin (2012). “The stochastic interdiction median problem with disruption intensity levels”. *Annals of Operations Research* 201.1, pp. 345–365.
- Lu, Ye and David Simchi-Levi (2013). “On the unimodality of the profit function of the pricing newsvendor”. *Production and Operations Management* 22.3, pp. 615–625.
- Magnanti, Thomas L. and Richard T. Wong (1981). “Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria”. *Operations Research* 29.3, pp. 464–484.
- Mak, Ho-Yin and Zuo-Jun Max Shen (2012). “Risk diversification and risk pooling in supply chain design”. *IIE Transactions* 44.8, pp. 603–621.
- Mills, Edwin S. (1959). “Uncertainty and price theory”. *The Quarterly Journal of Economics* 73.1, pp. 116–130.
- Murray, Chase C., Abhijit Gosavi, and Debabrata Talukdar (2012). “The multi-product price-setting newsvendor with resource capacity constraints”. *International Journal of Production Economics* 138.1, pp. 148–158.
- Peng, Peng, Lawrence V. Snyder, Andrew Lim, and Zuli Liu (2011). “Reliable logistics networks design with facility disruptions”. *Transportation Research Part B: Methodological* 45.8, pp. 1190–1211.
- Perakis, Georgia and Guillaume Roels (2008). “Regret in the newsvendor model with partial information”. *Operations Research* 56.1, pp. 188–203.
- Petruzzi, N. C. and M. Dada (1999). “Pricing and the newsvendor problem: A review with extensions”. *Operations Research* 47.2, pp. 183–194.
- Portnoy, Stephen (1991). “Asymptotic behavior of the number of regression quantile break-points”. *SIAM Journal on Scientific and Statistical Computing* 12.4, pp. 867–883.
- Potschka, Andreas, Christian Kirches, Hans Georg Bock, and Johannes P. Schlöder (2010). “Reliable solution of convex quadratic programs with parametric active set methods”. Technical Report, Heidelberg University.
- Qi, Lian and Zuo-Jun Max Shen (2007). “A supply chain design model with unreliable supply”. *Naval Research Logistics* 54.8, pp. 829–844.

- Qi, Lian, Zuo-Jun Max Shen, and Lawrence V. Snyder (2010). “The effect of supply disruptions on supply chain design decisions”. *Transportation Science* 44.2, pp. 274–289.
- Ramamurthy, Vivek, J. George Shanthikumar, and Zuo-Jun Max Shen (2012). “Inventory policy with parametric demand: Operational statistics, linear correction, and regression”. *Production and Operations Management* 21.2, pp. 291–308.
- Raz, Gal and Evan L. Porteus (2006). “A fractiles perspective to the joint price/quantity newsvendor model”. *Management Science* 52.11, pp. 1764–1777.
- Rockafellar, R. Tyrrell and Stanislav Uryasev (2000). “Optimization of conditional value-at-risk”. *Journal of Risk* 2.3, pp. 21–42.
- Roels, Guillaume (2013). “Risk premiums in the price-setting newsvendor model”. Working paper, University of California, Los Angeles.
- Salinger, Michael and Miguel Ampudia (2011). “Simple economics of the price-setting newsvendor problem”. *Management Science* 57.11, pp. 1996–1998.
- Scaparra, Maria P. and Richard L. Church (2008a). “A bilevel mixed-integer program for critical infrastructure protection planning”. *Computers & Operations Research* 35.6, pp. 1905–1923.
- (2008b). “An exact solution approach for the interdiction median problem with fortification”. *European Journal of Operational Research* 189.1, pp. 76–92.
- Scarf, Herbert (1958). “A min-max solution of an inventory problem”. In: *Studies in The Mathematical Theory of Inventory and Production*. Ed. by Kenneth J. Arrow, Samuel Karlin, and Herbert Scarf. Stanford University Press. Chap. 12, pp. 201–209.
- (1959). “Bayes solutions of the statistical inventory problem”. *The Annals of Mathematical Statistics* 30.2, pp. 490–508.
- Schwarz, Gideon (1978). “Estimating the dimension of a model”. *The Annals of Statistics* 6.2, pp. 461–464.
- Shen, Zuo-Jun Max, Roger Lezhou Zhan, and Jiawei Zhang (2011). “The reliable facility location problem: Formulations, heuristics, and approximation algorithms”. *INFORMS Journal on Computing* 23.3, pp. 470–482.
- Snyder, Lawrence V. and Mark S. Daskin (2005). “Reliability models for facility location: The expected failure cost case”. *Transportation Science* 39.3, pp. 400–416.
- (2006). “Stochastic p -robust location problems”. *IIE Transactions* 38.11, pp. 971–985.
- Snyder, Lawrence V., Zumbul Atan, Peng Peng, Ying Rong, Amanda J. Schmitt, and Burcu Sinoysal (2012). “OR/MS models for supply chain disruptions: A review”. Working paper. Lehigh University, Bethlehem, PA.
- Suykens, Johan A.K. and Joos Vandewalle (1999). “Least squares support vector machine classifiers”. *Neural Processing Letters* 9.3, pp. 293–300.
- Takeuchi, Ichiro, Quoc V. Le, Timothy D. Sears, and Alexander J. Smola (2006). “Nonparametric quantile estimation”. *Journal of Machine Learning Research* 7, pp. 1231–1264.
- Tang, Christopher S. (2006). “Perspectives in supply chain risk management”. *International Journal of Production Economics* 103.2, pp. 451–488.
- Wahba, Grace (1999). “Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV”. In: *Advances in Kernel Methods-Support Vector Learning*. Ed. by

- Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola. MIT Press. Chap. 6, pp. 69–87.
- Wang, Zizhuo, Shiming Deng, and Yinyu Ye (2011). “Close the gaps: A learning-while-doing algorithm for a class of single-product revenue management problems”. Working paper. Stanford University.
- Wang, Zizhuo, Peter W. Glynn, and Yinyu Ye (2010). “Likelihood robust optimization for data-driven newsvendor problems”. Working paper, Stanford University.
- Whitin, Thomas M. (1955). “Inventory control and price theory”. *Management Science* 2.1, pp. 61–68.
- Xu, Minghui, Youhua (Frank) Chen, and Xiaolin Xu (2010). “The effect of demand uncertainty in a price-setting newsvendor model”. *European Journal of Operational Research* 207.2, pp. 946–957.
- Xu, Minghui and Ye Lu (2013). “The effect of supply uncertainty in price-setting newsvendor models”. *European Journal of Operational Research* 227.3, pp. 423–433.
- Xu, Xiaolin, Xiaoqiang Cai, and Youhua (Frank) Chen (2011). “Unimodality of price-setting newsvendor’s objective function with multiplicative demand and its applications”. *International Journal of Production Economics* 133.2, pp. 653–661.
- Yang, Lei, Minghui Xu, Gang Yu, and Hanqin Zhang (2009). “Supply chain coordination with CVaR criterion”. *Asia-Pacific Journal of Operational Research* 26.01, pp. 135–160.
- Yang, Shilei, Chunming Victor Shi, and Xuan Zhao (2011). “Optimal ordering and pricing decisions for a target oriented newsvendor”. *Omega* 39.1, pp. 110–115.
- Yano, Candace A. and Stephen M. Gilbert (2004). “Coordinated pricing and production/procurement decisions: A review”. In: *Managing Business Interfaces*. Ed. by Amiya K. Chakravarty and Jehoshua Eliashberg. Vol. 16. International Series in Quantitative Marketing. Springer US, pp. 65–103.
- Yao, Li, Youhua (Frank) Chen, and Houmin Yan (2006). “The newsvendor problem with pricing: Extensions”. *International Journal of Management Science and Engineering Management* 1.1, pp. 3–16.
- Young, Leslie (1978). “Price, inventory and the structure of uncertain demand”. *New Zealand Operations Research* 6.2, pp. 157–177.