PRICING WITH INFORMATIVE DELAY ANNOUNCEMENTS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

MIDDLE EAST TECHNICAL UNIVERSITY

BY

SIRMA KARAKAYA

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER SCIENCE

IN

INDUSTRIAL ENGINEERING

JUNE, 2017

Approval of the Thesis:

PRICING WITH INFORMATIVE DELAY ANNOUNCEMENTS

submitted by **Sırma Karakaya** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver ______________

Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Murat Köksalan ______________
Head of Department, **Industrial Engineering Dep., METU**

Assoc. Prof. Dr. Seçil Savaşaneril Tüfekçi ______________
Supervisor, **Industrial Engineering Dep., METU**

Prof. Dr. Yasemin Serin ______________
Co-Supervisor, **Industrial Engineering., METU**

Examining Committee Members:

Prof. Dr. Haldun Süral ______________
Industrial Engineering Dept., METU

Assoc. Prof. Dr. Seçil Savaşaneril Tüfekçi ______________
Industrial Engineering, METU

Prof. Dr. Yasemin Serin ______________
Industrial Engineering Dept., METU

Asst. Prof. Dr. Benhür Satır ______________
Industrial Engineering Dept., Çankaya University

Asst. Prof. Dr. Emre Nadar ______________
Industrial Engineering Dept., Bilkent University

Date: 23.06.2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name :

Signature :

ABSTRACT

PRICING WITH INFORMATIVE DELAY ANNOUNCEMENTS

Karakaya,Sırma

M.S, Department of Industrial Engineering
Supervisor :Assoc. Prof. Dr. Seçil Savaşaneril
Co-Supervisor: Prof. Dr.Yasemin Serin

June 2017, 80 Pages

In many industries price and delay (lead-time) information affect customer behavior. According to the provided delay information and/or price, customers choose either to stay or balk. In this thesis, we study the effects of these two decision variables on the profitability. For the pricing problem, we study two schemes: static pricing and dynamic pricing. For the delay announcement problem, we consider three schemes: no information sharing, partial information sharing, and full information sharing. We model the system as a stochastic discrete-time Markovian queue, and model the pricing problems through Markov decision process. We compare several delay information and pricing schemes and through analytical results and numerical study identify the conditions that make specific delay information or pricing schemes more preferable. We also analyze the impact of the information asymmetry on the preference of the provider. We consider pricing schemes with varying levels of flexibility. Our findings show that even if the information asymmetry favors the service provider, the delay information scheme might severely curb the benefit of the pricing flexibility. Furthermore, results show that traffic intensity, customer-sensitivity level and price quotes have a significant impact on selecting the delay information scheme.

Keywords: Delay announcement, Dynamic Pricing, Stochastic Discrete-Time Markovian queue, Markov Decision Process, Information Asymmetry

ÖZ

GECİKME SÜRELERİ BİLDİRİMİ ALTINDA FİYATLANDIRMA

Karakaya,Sırma

Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi :Doç. Dr. Seçil Savaşaneril
Ortak Tez Yöneticisi: Prof. Dr.Yasemin Serin

Haziran 2017, 80 Sayfa

Birçok endüstri alanında fiyat ve gecikme süresi (teslim tarihi) bilgileri müşteri davranışını etkilemektedir. Müşteriye verilen bilgiye göre müşteriler sistemde kalmaya veya sistemden ayrılmaya karar vermektedirler. Bu tezde, bu iki karar değişkeninin kar üzerine etkileri araştırılmaktadır. Fiyatlandırma problemi için statik fiyatlandırma ve dinamik fiyatlandırma olmak üzere iki senaryo, ilan edilen gecikme süresi için de üç farklı senaryo üzerinde çalışılmıştır. Müşterilere hiç gecikme bilgisi verilmemesi durumu, kısmi bilgi verilmesi durumu ve kesin bilgi verilmesi durumu. Sistem kesikli Markov kuyruğu şeklinde modellenmiş, fiyatlandırma problemi ise Markov karar verme süreci çerçevesinde ele alınmıştır. Gecikme süresi ve fiyatlandırma senaryoları karşılaştırılmış, analitik sonuçlar ve sayısal çalışmalarla hangi koşullarda hangi senaryoların daha tercih edilebilir olduğu belirlenmiştir. Ayrıca, bilgi asimetrisisin servis sağlayıcısının tercihleri üzerine etkileri de incelenmiştir. Fiyatlandırma problemi değişik fiyat esneklikleri altında çalışılmıştır. Bulgular, bilgi asimetrisinin servis sağlayıcısına fayda sağlasa bile, gecikme süresi senaryolarının fiyat esnekliklerinin yararlarını sınırlandığını göstermektedir. Ayrıca, trafik yoğunluğu, müşteri hassasiyet seviyesi ve fiyat atamaları da gecikme süresi senaryo seçimini etkilemektedir.

Anahtar Kelimeler: Gecikme süresi, Dinamik Fiyatlandırma, Stokastik kesikli Markov kuyruğu, Markov Karar Verme Süreci, Bilgi Asimetrisi

To My Family

ACKNOWLEDGEMENTS

I wish to express my deepest appreciations to all people who helped me throughout this study starting from my supervisor Assoc. Prof. Dr. Seçil Savaşaneril. I would like to thank her for her endless support, patience and encouragement throughout this study. Her willingness to motivate me contributed greatly to this study. I would also like to thank my co-supervisor Prof. Dr. Yasemin Serin for her continuous helps to solve problems encountered during my study. Without their guidance and support, I would not complete this thesis.

I would also like to thank to my examining committee members Prof. Dr. Haldun Süral, Asst. Prof. Dr. Benhür Satır, Asst. Prof. Dr. Emre Nadar for their valuable suggestions and comments.

My special thanks g o to my dear parents Sevil and Yalçın Karakaya for their endless love, support, and understanding throughout this study and my life.

Last but not least, I would like to express my deepest gratitude to my dear sister Burcu Karakaya who always shared her own thesis writing experiences with me and motivated me whenever I was feeling down.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Price and delay information are the two factors that determine customer behavior and thus the system performance. Today, most firms accept the importance of these factors and use them as a weapon to gain the competitive advantage in today's market conditions (Zhao et al. (2011)). A survey conducted by Performance Management Group shows that top 110 performers in five major manufacturing sectors do not only focus on the cost but also on the speed of the delivery time to attract more customers (Boyacı and Ray (2003)). Uzun and Poturak (2014) claim that in online shopping the first three factors that affect customers' satisfaction is the convenience of trust, price and the quality of products. Corporate Executive Board's multiple surveys consisting of more than 7000 customers and interviews show that the long-term relationship between a firm and a customer depends on more than 40 variables, including price and trust to the firms. Furthermore, for items costing more than $50, customers start to investigate other alternative firms, which simply indicate that, if firms cannot offer a competitive price they may lose their customers. (Spenner and Freeman (2012)).

Although lead time and price quotation is important for the firms, there is still uncertainty about what should be the level of information revealed through quoted lead times and how should be a product or service priced. Often due to work congestion, quoted lead time and realized lead time differ from each other and this difference may negatively affect the credibility of the firm. While a shorter lead time attracts more customers, due to firms' capacity constraints it might not always be possible to adhere to the announced lead time. A longer lead time, on the other hand, might cause losing the customers and thus losing profit in a competitive market. The lead time and price quotation problem becomes more complex in the existence of heterogeneous customers. Customers may be either sensitive to delivery time or price or be sensitive to both. Firms should be well aware of the customer sensitivities to increase their chance to attract more customers.

Companies' may prefer static quotations or dynamic quotations (Zhang (2015)). Static quotations correspond to the case in which the lead time and price are pre-determined and fixed. For instance, some firms choose to offer price/lead time menus to attract both lead time sensitive and price sensitive customers. The basic idea is to offer longer lead times with lower prices and shorter lead times with higher prices. Dell, Roadway Express, Consolidated Freightways and Amazon are some firms that offer menus. Some firms prefer to promise a uniform lead-time guarantee to its customers. For instance, Domino Pizza guarantees 30 minute delivery, Federal Express, Ameritrade, an online trading market, promise to provide the service/product within a guaranteed time and accept to pay penalties if they cannot meet their promises (Boyacı and Ray (2003) and So (2000)).

Dynamic quotations correspond to announcements that might change dynamically over time, depending on some system state. Each arriving customer can possibly be announced a different lead time and price. For instance, call centers choose either to share the number of customers in the queue or directly provide the estimated waiting time based on the current number of customers. In many queues, such as banks', hotels', and amusement parks' queues, customer waiting positions are represented by ticket numbers where customers estimate their waiting time based on the number on the ticket. In a similar fashion, many hotels, airways, and telecommunication companies prefer to price their services dynamically to maximize their capacity utilization.

Both quotation modes have advantages and disadvantages. Based on the market conditions, customer sensitivities, firms' operational constraints one approach may outperform the other one. Under static quotation mode, since all customers are announced the same lead time and price, firms do not even have to keep track of detailed information about the system state. However, some customers might consider the announced lead time and/or price high and choose to leave the system. Firms do not have much flexibility to attract these customers. Under dynamic quotation mode, firms have to make some investments to gather detailed information on real-time state of the system. While under dynamic quotation costly investments may be required and operational complexities might increase, lead-time announcements become more precise, which might be preferable by the customers.

For the pricing problem, static pricing might be preferable since it does not require the update of the prices. Also, static pricing might increase the credibility of the firms because it ensures that all the arriving customers are proposed the same price. However, static pricing does not take demand fluctuations, market conditions and future available costs into consideration. Firms might not always estimate the right price; announced price may be considered high by the customers and cause customer lose, or price may be lower than customers' willingness to pay and firms might not incur the highest possible profits. Dynamic pricing provides adjusting prices based on time and fluctuating demand. It allows maximizing profits with each customer. However, customers might realize that they pay higher prices than the others and this might cause customer dissatisfaction. Firms should provide the trust that they fairly quote prices.

In real market conditions, we do not only observe static lead time with static pricing or dynamic lead time with dynamic pricing. Some firms prefer to adopt mixed strategies such as dynamic lead time with fixed price. For instance, McDonald's follow a fixed price/variable lead time policy; lead time varies according to demand rates. (Webster 2002)). Flight ticket sells can also be considered as an example of mixed strategy. For a given departure date and time (fixed lead time), different prices are proposed to the customers. Mixed approaches might be better for the firms since they are able to choose the most preferred scheme for lead time and the most preferred pricing scheme.

The existence of too many alternatives makes the decision of lead time and price announcement mode more difficult. The important point is to well analyze the conditions and understand which policies are better under which circumstances. The selection of these policies is important both in manufacturing and service systems. While lead time quotation is valid for manufacturing context, delay information is a valid term for service systems. In this thesis, we use delay information and lead time quotation interchangeably, since all the models provided in this thesis are applicable in both systems.

This study investigates how much information regarding the delay should be revealed to the customers and how a service provider should quote prices. The amount of information revealed to the customers affect the customer's purchasing decision, and to maximize profit the service provider must take customer response into consideration. For both price and delay announcement we define quotation schemes with various levels of "informativeness".

When announcing the delays, if customers are informed about the details regarding the customers waiting in the system, this corresponds to an informative scheme. In contrast, service provider may prefer to reveal minimal information regarding the delays, and this would correspond to a less informative delay announcement scheme. If the price is quoted based on real-time status of the system (e.g. congestion in the system), this corresponds to a more informative price quotation scheme. In contrast, if prices are static with respect to the system status, this corresponds to a less informative scheme.

To understand the impact of quotation schemes with various levels of informativeness on profit, and find out about the interplay between price and delay announcements, we specifically define the following 2 schemes for pricing: (1) static pricing, and (2) dynamic pricing, and 3 schemes for delay announcement: (1) No information, (2) Partial information, (3) Full information. Overall, the profit of service provider is compared under 6 schemes. Furthermore, we assume that (1) the service provider might have perfect information regarding the system status, but may be selective in revealing the information to the customers (asymmetric information case), or (2) might have exactly the same information level that the customers have, and reveals whatever information he has (symmetric information). In total, we study 12 scenarios.

We compare these scenarios and through analytical results and numerical analysis identify the conditions that make specific delay information or pricing scheme preferable. We also analyze the impact of information asymmetry on the preference of the service provider. As provided in Chapter 6, we consider pricing schemes with varying levels of price flexibility and assume customers are differently sensitive to the precision of the delay announcement.

The outline of the study is as follows. In Chapter 2, we first provide the related literature. In Chapter 3, we introduce modeling assumptions for the service system and explain service provider's and customers problem, seperately. Models under symmetric information and under asymmetric information cases are described in Chapter 4 and Chapter 5, respectively. In Chapter 6, findings upon computational analysis are discussed. In Chapter 7, general remarks are made and future research directions are discussed.

CHAPTER 2

LITERATURE REVIEW

We study different levels of delay information sharing and different pricing schemes, and the optimal policy in the presence of delay- and price-sensitive customers. Our research relates to three main streams of past research: (i) those on delay information sharing schemes, (ii) those on pricing schemes and (iii) those that jointly address delay and price quotation decisions under a given information and pricing scheme.

2.1 Delay Information Sharing Schemes

Delay information sharing might be viewed within supply chain information sharing. In the absence of information asymmetry, chain may become more coordinated. Papers by Güllü (1997), Gavirneni et al. (1999), Cachon and Fisher (2000), Smichi-Levi and Zhao (2003), Tian and Pan (2008) analyze the effect of downstream demand information sharing with the upstream, on upstream profits, or on total chain cost. Swaminathan et al. (1995), Chen and Yu (2005), Dobson and Pinker (2006), Choi et al. (2008), Bendre and Nielsen (2013) on the other hand, consider settings where upstream information such as capacity, quality or location in the transportation stage is shared with the downstream. Chen and Yu (2005) investigate the benefit to the downstream. Dobson and Pinker (2006), and Choi et al. (2008), investigate the benefit to the upstream. Swaminathan et al. (1995) and Bendre and Nielsen (2012) consider the benefit to the overall system. In our paper, we consider a setting where the upstream shares wait time information with the downstream. The upstream decides on how much information to reveal to the downstream, and how the revelation scheme affects the profitability of the upstream.

There exists a vein of past literature on service systems that incorporates the effect of delay announcements on customer behavior. Guo and Zipkin (2007) study three schemes of delay information sharing: no information partial information and full information. The system is modeled as a single-server M/M/1 queue. Each customer is assumed to have his own utility function which equals a reward for having the service minus a waiting cost. Waiting cost depends on a customer specific function. Service provider announces each arriving customer a delay and based on utility function customers choose either to enter or leave the system.

Under no information all customers receive the same information. Under partial information, the provider tells the customer, the occupancy of the system as an arrival occurs. The customer either chooses to stay or balk after computing his utility function. Under full information, the provider tells the customers the exact waiting time. The authors compare scenarios in terms of throughput and utility of customers. Comparison of no information and partial information shows that more information improves at least one party but it is not clear whether the provider or the customers receive the benefit. With more information, average utility always increases but based on the cost scale distribution, the throughput may or may not increase. The same observation is also valid for the comparison between no information and full information scenarios. Comparison between partial and full information shows that if the cumulative distribution function of customer-type parameter is power distribution then more information increases throughput and the average utility. The authors claim that the shape of cost distribution may change the observations and the results are sensitive to modeling assumptions.

Guo and Zipkin (2009) extend the study to the case where the service time has a phase-type distribution. The authors propose two different queue models with different types of delay information at the existence of balking behavior. The first model is called partial information and the other one is called phase information. Both models are studied as M/M/1 and for both models FCFS discipline is assumed. In this paper, also customers are assumed to choose to stay or balk based on their own utility function. Under the first model each arriving customer learns the system occupancy in terms of intervals. Therefore, customers are announced an approximation for the range of the current length. Under the second model, the service provider informs each customer about the remaining number of the phases in the system. It is assumed that for each customer's service time K random number of phases exist. The number of phases is geometrically distributed and those phases have independent and identical exponential distributions. The statistic of remaining number of phases behaves as a continuous-time Markov chain. By deriving and solving balance equations the system performance measures are observed. Guo and Zipkin (2009) compare the two models in terms of customer and server utility. The authors claim that for both models announcing delay information more accurately may or may not be beneficial for the server and/or the customers. If customers are heterogeneous, the server benefits more from providing accurate delay information. If information benefits the server, the ratio of the average utility obtained from more information is greater than the ratio of their idle probabilities.

Whitt (1999) considers a call center where customers are announced delays under different announcement schemes. The author proposes two models which both are modeled as birth-and-death (BD) stochastic process. For both models FCFS service discipline is adopted. The first model represents the case which no delay information is provided to customers. Customers just learn whether they can immediately have service or not. The author regards this model as the traditional BD to describe performance when customers have opportunity to balk or renege from the system. It is assumed that, the customer waits until his delay threshold, and then reneges if service has not yet been provided. The second model represents the case where servers inform customers about anticipated delays or provide state information so that arriving customers compute an expected waiting time. In the second model, all reneging is replaced by balking. Authors explain this observation as because customers are aware of their expectations they always respond to additional information. When all servers are busy instead of reneging state-dependent balking.is observed. Once customers join the queue, customers are much more likely to remain until they begin service. Thus, reneging is less observed when customers understand/observe that the remaining time to wait is steadily declining.

As an alternative to the second model, Whitt (1999) introduces a BD model to represent state-dependent balking instead of time-dependent reneging. The author also deals with a variation of the second model where reneging is still observed. A general BD model representing state-dependent balking is defined and state-dependent reneging is also included. Whitt (1999) makes a stochastic comparison of models. The author claims that the first model produces higher throughput at the expense of having some customers wait without eventually receiving service. With common parameter tuples, the probability of receiving service of an arriving customer is very close for the two models. Also, the two models do not differ when all servers are busy. As the load increases the difference is more emphasized. Under heavy load, for both models low proportion of customers are served and the delays experienced by these customers are not large.

Jouini et al. (2011) also analyze a call center with impatient customers. It is assumed that each arriving customer has different patience time, and it is exponentially distributed. The authors propose three models and this assumption is valid for all of them. The first model corresponds to the perfect setting where arriving customers are told the exact waiting time. Each arriving customer compares the announced lead time with his willingness to wait, and if it is less than his willingness to wait then he chooses to wait, if not he balks. Since the information provided is certain for this case no reneging behavior is observed.

In the second model, the customers are not informed about delay. The authors claim that as uncertainty increases, the proportion of customers who are unwilling to wait also increases compared to the first model. The customers either choose to balk immediately either after they learn that all the servers are busy or after they learn that no information will be provided to them. If a customer does not balk then he either waits until his service ends or until the end of his willingness to wait which simply corresponds to the reneging behavior. It is assumed that once a service is provided to a customer he cannot renege. The authors suggest that this scenario can be modeled as M/M/s+M queuing system with balking. Under the third model, customers are announced a delay ensuring a coverage. Under this model, customers do not balk from the system due to uncertainty but as a function of announced delay. Customers may or may not update their patience value according to announced delay. The two cases are analyzed as they represent two different customer behaviors. Jouini et al. (2011) claim that as the announcement coverage increases reneging behavior decreases but balking behavior increases and more coverage may not always result in the benefit of the service provider. Also, a different version of the third model is introduced. Instead of announcing delay information with certain coverage a mean delay is announced.

The authors conclude that announcements with higher reliability are more important when the arriving customer's impatience is high, when system is smaller and when system congestion is high. Comparison of the second and the third model shows that under the second model, less balking but more reneging behavior is observed. As system size increases the performances of the two models become similar, and the benefit of more information decreases. Comparison between different versions of the third model show that if customers less frequently update their patience rate according to delay information, or if pooling increases, the performance of both models become similar.

Allon et al. (2011a) consider a service system where both the service provider and the customers are strategic in their actions. The service provider is strategic in the way he provides information. He may choose to give truthful information or totally mislead customers about the delay which the authors define it as intentional vagueness. The customers are strategic in deciding whether to enter the system or balk and in the way they interpret the provided information. The authors model the game between the provider and the customers as Markov-perfect Bayesian Nash equilibrium (MPBNE). Their findings indicate that information improves the outcome for all the players and even though the information provided to customers is nonverifiable the profits of the firm and the average utility of the customers increase.

Allon and Bassamboo (2011b) approach to delay announcement problem from a different perspective and investigate the timing of the delay announcement. Customers are either immediately or after some time informed about the announcement. The effect of postponing the announcement of delay is analyzed in terms of profits and the utilities for the firm and the customers. The authors suggest that delaying the announcement improves the profit but in practice, this approach may be very costly. Furthermore, they suggest that under certain settings, providing delay information immediately may be better for the firm because immediately providing delay information may create credibility for the firm. An increase of the credibility of the firm improves not only the profits but also the customers' overall utility. However, such an approach may also hurt a firm's credibility if the firm is more sophisticated in its strategies.

Service (or make-to-order manufacturing) systems where the server (manufacturer) provides varying levels of system observability can be considered in the same stream and are relevant to our study.

Larsen (1998) studies a MTO manufacturer, where customers are either announced the steady-state or system state dependent delay information under a static pricing scheme. Both models are proposed within M/M/1 framework. When a customer arrives to the system, he decides to enter the system if his reward minus the sum of the price charge of his job and the expected waiting cost is positive. Otherwise, he immediately balks from the system. Assuming rewards are uniformly distributed, Larsen (1998) investigates how the total welfare and/or total profit changes with the reward. The author explains that changing the bounds of the reward means changing the transparency of the customers market.

The author shows that if customers are announced the steady-state delay and if the market becomes more transparent, it not certain whether it is good or bad for the job shop, it depends on the parameters. For welfare optimization more transparency is often not beneficial. If customers are announced a system state dependent delay, for profit optimization again the transparency seems to be beneficial for the job shop but for the welfare optimization it is not certain whether transparency is better or not. For both scenarios, the optimal price for profit optimization is greater than the price for welfare optimization. When the range of uniform reward distribution decreases, the welfare contribution also decreases and the profit contribution either increases or decreases. Comparison of scenarios shows that, except for a special case, more information is always better.

Ding et al. (2014) consider an observable queue with a single server where customers are not directly announced a waiting time, but provided a ticket number which represents waiting positions. Customers may always observe the system state and may renege at any time. The tendency to renege is assumed to be dynamically dependent on difference between the ticket number and the number being served. The authors develop a procedure to be able to calculate the percentage of reneging customers approximately.

Hassin (1986) addresses the question whether it is socially optimal to suppress information on the queue. If suppression is made, then customers decide on joining the queue on the basis of the known distribution of waiting times. Findings show that it is never socially optimal to suppress queue length information.

Hassin (2007) extends the previous study by introducing the uncertainty on system parameters: service rate, the service quality, or the waiting conditions. The author questions whether the server should withhold these parameters from customers or inform them about the realized parameters of random variables. Additionally he investigates whether the server should set a priori or a posteriori price for the service. For uncertain service rate, μ, informing customers and adopting dynamic pricing yield either a greater or equal profit compared to profit when customers are informed and quoted a single price. Giving no information yields lower profit compared to the case where information is provided. As the uncertainty of μ increases, the difference between the profits obtained under informative and uninformative cases increases, and the server's objective differs from the social welfare. For uncertain waiting cost, C, if waiting cost is high, informing customers is better. However, based on the C value informing customer may not always increase the profits with respect to uninformative case. For uncertain value (quality) of service, R, if C is close to 0 and the server is restricted to a single price, he is motivated to conceal the realized value of R from the customers. Under informative case, and in the existence of two prices, profits increase as the uncertainty increases. If there is a single price, profits initially decrease as a function of variance and there is a profit increase only for larger variance.

Shone et al. (2013) study an M/M/1 queue where the queue is either observable or not by the customers. There are two types of customers, selfish and altruistic. The models are compared in terms of system properties and performance measures under two different types of optimal customer behaviors which the authors call selfishly optimal and social optimal. Findings indicate that revealing or suppressing the information on the queue length may or may not affect the joining rate, and it depends on the input parameters and the type of the customer.

Dobson and Pinker (2006) investigates the effect of level of delay information on a firm. The authors introduce a firm which has possibility to provide customers the estimates of lead time based on the number of customers in the queue. This firm is compared with the one that provides the same lead time to all its customers, based on the long run lead time averages. These two levels of information correspond to two different models. For both scenarios it is assumed that customers decide to make a request based on the announced lead time. If quoted lead time is higher than their willingness to wait, they depart from the system, if not then they submit their request. Customers have different willingness to wait. The scenarios are compared in terms of the firms' throughput.

In the first scenario because all the customers are quoted the lead time based on past averages, there is a single request submission and thus the system is regarded as an M/M/1 queue. In the second scenario, the firm announces customers the state-dependent lead time which is provided based on the number of requests in the queue. The system can be characterized as a birth-death process. The authors show that announcing lead time based on the queue length may increase throughput and decrease the waiting time of customers. However, the benefit of sharing lead time information is dependent on the shape of the demand curve. They add that there are some cases which the throughput decreases as a result of increased information sharing. However, they suggest that in many cases providing state-dependent lead time information is better than the information provided based on average values. It can increase a firm's throughput while reducing customer waiting times and the variance in the waiting time. Furthermore, a firm's throughput tends to increase with information sharing when customer lead time expectations are heterogeneous, and tends to decrease when expectations are homogeneous.

In this thesis, we study three different levels of delay information sharing, which simply are no information, partial information and full information. We model different information levels under discrete-time Markov chain.

2.2 Pricing Schemes

Gayon et al. (2009) study different pricing strategies for a production-inventory system where the supply is capacitated and the demand is fluctuating. The demand is assumed to be dependent on both the environment and the quoted price. The paper proposes three different pricing schemes; static pricing, which all the customers are assigned a pre-determined price, environment dependent pricing, which a price for each environment is assigned and dynamic pricing which the price changes based on both the current environment and the stock level. The objective is to find an optimal replenishment under each pricing policy and decide which pricing policy is better under which conditions.

The authors claim that when demand is stationary, static pricing is nearly optimal and the benefit of dynamic pricing over static pricing is relatively small. If demand rate is fluctuating, on the other hand, then dynamic pricing significantly improves profits with respect to static pricing. However, because changing prices may be costly and may cause negative customer reactions, environment dependent pricing might be preferred. It is as effective as dynamic pricing when demand environment fluctuates.
Cachon and Feldman (2010) investigate whether a firm should change its prices dynamically to shifts in demand or adopt static pricing. The authors study three different pricing schemes, namely static pricing, dynamic pricing and a scheme that

combines static and dynamic pricing which they call constrained dynamic pricing. It corresponds to charging a list price and if necessary, marking down the price but never marking up. Findings show that in the presence of strategic customers, a firm may be better off with static pricing than dynamic pricing if the customers' valuation for the product is highly variable and constrained dynamic pricing may be better than both static and dynamic pricing.

Koenig and Meissner (2009) compare dynamic and list pricing for a firm selling multiple products. The authors claim that adopting dynamic pricing may be better with decreasing capacity when demand is held constant. If the capacity consumption is uniform then again choosing dynamic pricing might be better. However, if changing the prices is costly or impractical, list pricing might be preferred.

In this thesis, we quantify the benefit of dynamic pricing over static pricing in the presence of various delay information sharing schemes.

2.3 Joint Delay and Price Quotation

In this section, we mention the studies that jointly consider delay (or lead-time) announcement and pricing policies. Our work is distinct in that these studies do not address the scheme selection problem, but rather determine the optimal policy under a given scheme.

So and Song (1998) propose a framework to analyze pricing, delivery time guarantee and capacity expansion in service systems. All customers are quoted the same price and lead-time. Such a scheme corresponds to one of the information and pricing scheme couples in our study which we call static pricing no information (SPNI) scenario. We develop further more flexible schemes as well. As in our case, So and Song (1998)'s model is applicable to both service and MTO manufacturing systems. The authors' findings indicate that if the demand is low, then quoting a smaller delivery time and a lower price is optimal. If operating costs are high then announcing smaller delivery time but higher price is optimal. If the firm chooses to have high desired service level, then it should set a larger delivery time quote but a lower price. Finally, if firm has high unit operating cost, it becomes more critical to decide the guarantee the optimal delivery time and the corresponding price.

So (2000) extends the previous study to a monopolistic versus a competitive setting. The author indicates that if firms are homogenous than the optimal policies are the same with the monopolistic market structure. If firms, however, are heterogeneous then based on the market and firms' characteristics, the optimal policies change. While a high capacity firm competes with shorter time guarantee, a firm with lower unit operating costs competes with lower price. If the market is time-sensitive, firms compete less on price but on delivery time. That is why, in such a market prices are higher but delivery-times are shorter. If the market is price sensitive, prices are low but delivery times are longer. The author claims that the firms should be aware of the market sensitivity and should decide on their actions based on the market structure.

Palaka et al. (1998) investigates lead-time setting, capacity utilization and pricing decisions in the existence of lead-time sensitive customers. The authors model the firm's operations as M/M/1 queue and assume that demand is a linear function of price and lead-time. They investigate the effects of changing parameters such as unit waiting cost, price and lead-time sensitivities, and etc. on a firm's decisions. Their findings show that as lead-time sensitivity increases the optimal price lowers. An increase in unit waiting cost results in lower optimal arrival rate and lead-time announcement. If unit lateness penalty is low, then firm chooses to increase price in order to decrease arrival rate. If penalty is high, firm chooses to announce high lead-times to avoid penalties. In order to compensate high quoted lead-times, it drops the prices.

Boyacı and Ray (2003), analyze the delivery-time and price quotation policies for a firm that sells two substitutable products. One of the products has a shorter delivery time, and the other one has a faster delivery time. The paper aims to find a delivery-time for the faster product and the optimal prices for both of them. The authors consider three cases; the firm is either constrained in capacity for none, or one or both products. Under the assumption that the firm has no capacity constraint, an increase in the guaranteed lead time for the faster product results in a linear decrease, and either a linear increase or decrease in optimal prices for the fast and slow products, respectively. If the firm has capacity cost for faster product, the optimal price is higher for the slower product and lower for the faster product if the market is time sensitive compared to no capacity constraint case. Under this scenario, if the firm has capacity cost for the faster product, it should reduce time differentiation and reduce/increase price differentiation if the marginal cost is low/high. If the firm has capacity costs for both products, the optimal price for the faster product is higher. For the slower product it is also higher for time sensitive market compared to the case when there is no capacity cost. Furthermore, again compared with no capacity cost case, the demand for the faster product is higher and lower for the slower product.

Ray and Jewkes (2004) model an operating system of a firm and its customers where customers are either more sensitive to delivery time or the market price. The market price is assumed to be dependent on the length of delivery time. The objective of the firm is to maximize its profits by selecting an optimal delivery time with the assumption that reducing delivery time requires investment and the firm must satisfy a pre-specified service level. The authors conclude that the behavior of firms change according to the sensitivity of customers. Furthermore, if price is dependent on the delivery-time then firms must give more importance to the selection of optimal delivery time.

Zhao et al. (2011) investigates price and lead time quotation modes for a firm in either service or MTO manufacturing industries. They propose two models which they call uniform quotation mode (UQM) and differentiated quotation mode (DQM). Under UQM, the firm offers a single lead time and price. Under DQM the firm offers a menu of lead times and prices for customers to choose from. Customers are assumed to be either lead time sensitive (LS) or price sensitive (PS). The authors analyze which modes are better under which conditions. Their findings indicate that if LS customers do not value the product more than PS customers, adopting UQM is better than DQM, independently from any operational costs. If LS customers value the product more, however, then the firm should check delay costs of customers. If delay cost of LS customers is higher than the delay cost of PS customers, then DQM should not be adopted. Apart from the valuation of LS customers to the product, the benefit of quotation mode is dependent on both the firm's operational properties and customer characteristics. If PS customers have positive utilities in UQM and LS customers have positive utilities in DQM, then a firm should not adopt DQM.

Afeche and Pavlin (2016) also investigate how should a firm design a price/lead time menu in the existence of time sensitive customers. The aim is to maximize revenues under the optimal price/lead time menu. The authors model a service or MTO provider as an M/M/1 model and analyze the necessary and sufficient conditions in terms of capacity, the market size and structure of delay-cost distribution. Their finding indicate that for fixed capacity as arrival rate increases revenue also increases but the service provided to the patient customers slows down. From these customers the gained revenue decreases. When the capacity is low, higher lead times should be offered to the customers and neither pooling nor strategic delay is optimal. If the market size is sufficiently large, the arrival rate is smaller than the market size then providing shorter lead time to the impatient and higher lead time to the patient customers is optimal. Furthermore, at low arrival rates pooling is not optimal since the capacity is enough to provide all the customers relatively shorter lead time.

Ata and Olsen (2008), Çelik and Maglaras (2008), Feng et al. (2010), and Hafizoğlu et al. (2016) consider dynamic pricing and lead-time quotation modes in the existence of either delay sensitive or price sensitive customers. In the paper of Ata and Olsen (2009) customers are assumed to be homogenous and have nonlinear disutility for delay. The firm is monopolist. The dynamic policies are investigated for convex, concave and convex-concave lead time cost functions and proved that these are asymptomatically optimal. Hafızoğlu et al. (2016) consider joint dynamic lead time and price quotation with two customer classes. The first class is contract customers whose orders are always satisfied and fulfilled based on a contract price and lead time based on a pre-determined time horizon. The second class is spot customers who are dynamically quoted a price and lead time. The dynamic quotation problem is modeled as an infinite Markov decision process.

2.4. Positioning our study in the literature

This thesis study is closest to papers of Guo and Zipkin (2007) and Dobson and Pinker (2006). Along with those studies, we compare several delay information schemes in the presence of delay-sensitive customers, to identify the conditions under which each delay information scheme is beneficial to the service provider. However, we enrich the delay information sharing problem by extending it towards several dimensions. First, we model the customer utility by incorporating the sensitivity of the customers to the delay as well as to the precision of the quoted delay. We study the problem also under information asymmetry, which implies that the service provider may not reveal all the information to the customers. Finally, we study the interplay between the pricing schemes and the delay information schemes. We consider dynamic pricing and static pricing schemes, where dynamic pricing corresponds to a more flexible scheme. However, it turns out that depending on the delay information scheme and the information asymmetry the benefit of the flexibility might be slashed. Furthermore, our results show that pricing scheme has a significant influence on selecting the delay information scheme.

CHAPTER 3

THE MODELING FRAMEWORK

In this chapter, in §3.1 the modeling assumptions and the framework are introduced. Based on the framework, the main problem is introduced both from the service provider's and the customers' point of view. While for the provider, the problem is to choose the right lead time and price quotation scheme and thus maximizing the profits, for the customers the problem is to choose either to enter the system and have the service or immediately balk from the system. The detailed analysis of the problem for the service provider and for the customers can be found in §3.1.1 and §3.1.2, respectively. In §3.2, the studied scenarios are introduced and briefly explained.

Table 1: Notation for Chapter 3.

λ	probability of customer arrival in a period
μ	probability of service completion in a period
r	customer valuation of the service (service reward), $r>0$
p	price quoted by the service provider
w_1	the announced delay (expected waiting time) to the arriving customer
$c_1(w_1)$	a non-decreasing function of w_1
w_2	the announced precision delay (variance of the waiting time) to the arriving customer
$c_2(w_2)$	a non-decreasing function of w_2
Θ_1	sensitivity factor indicating the sensitivity of the customer to the precision of the announced delay (a continuous, strictly positive random variable)
$F(x)$	cumulative distribution function of Θ_1
Θ_2	sensitivity factor indicating the sensitivity of the customer to the precision of the announced delay (a continuous, strictly positive random variable)
$H(p,w_1,w_2)$	probability that an arriving customer, who is quoted p, w_1 and w_2, purchases the service

3.1 Description of the Model

We model the system as a discrete-time stochastic process. Every period a customer arrival or a service completion may take place. At most one customer arrives per period with probability λ. In every period, a service is completed with probability μ if there is a customer in service. There are no limits on waiting room capacity, and a single server is assumed. Given this description, the system can be said to evolve following a discrete-time version of an M/M/1 queue and can be modeled through a discrete-time Markov chain.

3.1.1 The Service Provider's Problem

The service provider provides service to the arriving customers following a first-come-first-served (FCFS) policy. Upon arriving, a customer may find other customers waiting in the queue. The arriving customer is quoted a price and announced how much she will wait in total (queue+her own service). This announced duration might involve some error and service provider also informs the customer on the precision of the announcement. If the customer accepts the quotes, she joins the queue, otherwise she leaves the system (balks).

The objective of the service provider is to maximize his profit per period, where profit is defined as revenue minus waiting cost. It is assumed that for each time unit a customer is waiting in the system a unit waiting cost is incurred. Given a specific pricing and delay announcement scheme, the decision to be made is what price to quote to the arriving customers. We assume service provider always makes truthful delay announcements. The truthfulness is ensured by quoting the "expected value of the waiting time" and quoting the precision of the announcement, where we use the "variance of the waiting time" as a proxy for the error/precision of the announcement. As we will see in the following sections, under a given scenario, given that the service provider always provides truthful announcements, he does not make any decisions regarding the announcement of the expected value of delay and the precision of the announced delay. The only decisions to be made are the amount of information to be revealed with the customers and the price quote.

3.1.2 Customer's problem

Each customer receiving service gains a utility from the service, but the utility may be trimmed due to the price to be paid or the delay faced. If the "net" utility gained is negative, the customer leaves the system without entering. The utility function of a customer is as follows:

$$U(p, \Theta_1, w_1, \Theta_2, w_2) := r - \Theta_1 c_1(w_1) - \Theta_2\, c_2(w_2) - p \tag{1}$$

The service reward, r, is assumed same for all customers. Customers differ in their valuation of waiting time and have different sensitivities to precision of the announced waiting time. The sensitivities are expressed by the customer-type parameters Θ_1, Θ_2 respectively. Actually, these parameters make the utility function customer specific. For the sake of simplicity, we assume Θ_2=1.

Under the no information, partial information and full information delay announcement schemes, customers use the service provider's announcements on expected delay and the precision of the delay to evaluate their utility. Let the probability that an arriving customer, who is quoted p, w_1 and w_2, purchases the service be denoted with $H(p,w_1,w_2)$. Given that Θ_2=1, and that $c_1(w_1)$ is a positive non-decreasing function of w_1 and $c_2(w_2)$ is a positive non-decreasing function of w_2, $H(p,w_1,w_2)$ is obtained as follows:

$$\begin{aligned} H(p,w_1,w_2) &= \mathrm{P}\ (\mathrm{U}(.)>0) \\ &= \mathrm{P}\ (r-\Theta_1\, c_1(w_1)-c_2(w_2)-p>0) \\ &= \mathrm{P}\ \left(\Theta_1<\frac{r-p-c_2(w_2)}{c_1(w_1)}\right) \end{aligned}$$

Note that whenever $H(p,w_1,w_2) \in (0,1)$, $H(.)$ is a monotone decreasing function of w_1 and w_2.

3.2. Analysis of the Scenarios

In Figure 1, we define each delay information scheme and each pricing scheme.

Symmetry scheme→		The service provider has the same level of information with the customers (symmetric scheme)	The service provider has perfect information (FI) and chooses the amount of information to reveal to the customers (asymmetric scheme)
Pricing scheme ↓	**Information scheme** ↓		
Static price (SP) Customers are quoted the same price	**No information (NI): Same delay and same precision announced to all customers**	SPNIs	SPNIa
	Partial Information (PI): Delay and precision based on the number of customers in the system	SPPIs	SPPIa
	Perfect/full information (FI): Delay and precision based on exact workload amount	SPFIs	SPFIa
Dynamic Price (DP) Price quoted based on system status	**NI**	DPNIs	DPNIa
	PI	DPPIs	DPPIa
	FI	DPFIs	DPFIa

Figure 1: Detailed explanation of scenario abbreviations

For the pricing, static scheme corresponds to the setting where all arriving customers are quoted the same price. In dynamic scheme on the other hand, each arriving customer may be quoted a different price depending on the status of the system.

For delay announcements, no information scheme corresponds to a setting where the current status of the system is ignored and all customers are announced the same, steady-state, expected delay. The precision of the announcement is measured with the "variance" of the delay and all similarly customers are quoted the steady state variance of the delay.

Partial information scheme corresponds to a setting where the service provider is more informative regarding the delays. The announcements are made based on the number of customers waiting in the system. Conditioning delay announcements on the number of customers still leaves some uncertainty regarding the actual delay faced by the customers. Now, the precision of the delay is also conditioned on the number of customers in the system. If there are a few number of customers waiting in the system, the precision of expected delay would be high (variance of delay would be low), whereas as the number of customers in the system increase the precision would be low (variance would be high).

Full information scheme corresponds to a setting where the service provider gives perfect information regarding the delays. The delay announcements are now made based on the actual workload in the system. This means there is no uncertainty regarding the actual delay, whatever is announced to the customer is exactly the realized delay (the realized delay is announced to the customer).

We analyze each of these scenarios under the symmetric information and asymmetric information cases. Under symmetric information the service provider announces the delays in the most informative way, but may have limited information on the system status. Since he shares all the information he has, the service provider and the customers have the same level of information. Specifically, we consider the cases where the service provider has no information on the system status, can only observe the number of customers in the system, and can fully observe the remaining workload and any arriving customer's workload. Under asymmetric information, the service provider is fully informed regarding the status of the system, in that he can fully observe the remaining workload and any arriving customer's workload. However, he may announce the delays to the customers based on either no information, or partial information, or full information scheme. Although the service provider has always full information customers might have less or the same level of information with the customers.

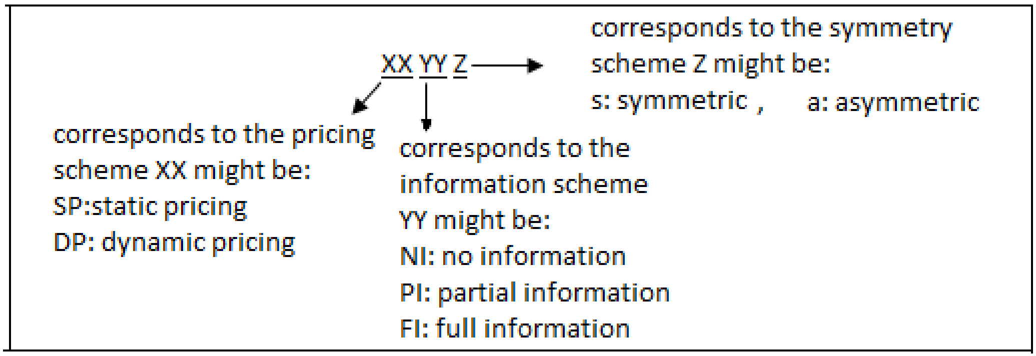


Figure 2: Explanation of scenario abbreviation reading

In Figure 2, we explain how to read a scenario explanation. For a scenario abbreviation, Part XX shown in Figure 2 always corresponds to the pricing scheme. It might be either dynamic pricing or static pricing. While under static pricing all customers are quoted a fixed price, under dynamic pricing based on the service provider's knowledge about the system state, prices are quoted dynamically. Notice that if the server provider has partial information then price quotations are made based on the number of customers in the system, if he has full information, however, then price quotations might be done based on perfect information. YY always corresponds to the information scheme provided to the customers. It might be either, NI, PI or FI. Z corresponds to the information symmetry. Z might be either "s" or "a" which show symmetric or asymmetric cases, respectively. Notice that if information is symmetric then it can be considered that the service provider has YY amount of information; either he has no information (NI), or partial information (PI) or full information (FI). If information is asymmetric, then the service provider has full information, independently from the information level provided to customers.

We specify the decisions of the service provider and the response of the customers under each of the 12 scenarios. In Chapter 4, we first study the scenarios under symmetric information (6 scenarios) and then in Chapter 5, we move on to the case of asymmetric information (6 scenarios).

CHAPTER 4

SYMMETRIC INFORMATION

Under symmetric information the service provider is assumed to have no, partial and full information regarding system status, respectively. Provider makes delay announcements considering the information regarding the system status, in the most informative way. We describe the models constructed under static pricing, with no, partial and full information delay announcement schemes in detail. We denote these scenarios with SPNIs, SPPIs, SPFIs, respectively. Then we model the dynamic pricing counterparts, which we denote with DPNIs, DPPIs, and DPFIs.

Table 2: Additional Notation for Chapter 4.

X_n	System state defining the number of customers in the system in period n
π_i	Steady-state probability of being in state i
ρ	Traffic intensity /server utilization, $\rho = \frac{\lambda}{\mu}$
W	The total waiting time in the system of a customer arriving in steady-state
Y_i	Service time of i^{th} customer in the system
N	A random variable denoting the number of customers in the system including the arriving customer in steady-state
E[N]	Steady-state expected number of customers in the system (including the arriving customer)
Π_i	Profit obtained per period in the long run under scenario i
w_1^i	Expected waiting time when the arriving customer finds i customers in the system
w_2^i	Variance of waiting time when the arriving customer finds i customers in the system
W_i	A random variable denoting waiting time for i customers in the system (including the arriving customer)

Table 2 continued

$w_1^{(wo,j)}$	Expected waiting time when the system workload is wo and arriving customer brings j amount of workload
$w_2^{(wo,j)}$	Variance of waiting time when the system workload is wo and arriving customer brings j amount of workload
WO_n	Total workload of the system in period n
Y_n	Workload of arriving customer in period n
b_j	Probability that the arriving customer brings j amount of workload
$\lambda_{wo,j}(p)$	Probability that an arrival with j amount of workload enters given that the quoted price is p, at state wo.
$\lambda_{wo}(p)$	Effective arrival probability to the system at state wo given that the quoted price is p

4.1. Static pricing under no information (SPNIs)

No information symmetric scheme scenario corresponds to the case where the information on system state is neither known by the service provider nor communicated with arriving customers Each side has the same level of information. All arriving customers are quoted the same price *p*, and are announced the same expected waiting time in the system, w_1, and the same variance of the waiting time in the system, w_2. Announced expected waiting time and the variance of the waiting time are not based on real-time information, but on long-term equilibrium. Based on this information, customers choose to join or leave the system. Because all customers are given the same information, the probability of customers join the system becomes the same.

Before determining w_1 and w_2, we derive expected waiting time and the variance of waiting time for a general discrete-time single server queue. Suppose the customers arrive to the system with probability λ and service completion occurs with probability with μ, where $\lambda<\mu$. (For the sake of simplicity, we initially treat the case where the arrival rate λ is not affected by the announcements. Then, we will incorporate the effect of w_1 and w_2). We define the state of the discrete-time Markov chain as the number of people in the system at period n and denote it with X_n. Then $X_n \in \{0, 1, \ldots\}$. Let π_i denote the steady-state probability that there are i customers in the system.

Lemma 1: For the discrete-time queue with parameters λ and μ, in steady-state,

$\pi_0 = \frac{(\mu-\lambda)}{\mu(1-\lambda)}$ and $\pi_i = \frac{(\mu-\lambda)}{\mu(1-\lambda)}\left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i$ where i=1,2,...

Proof. For this chain, the transition probability function is as follows:

$$P(X_{n+1} = k | X_n = i) = \begin{cases} (1-\lambda) + \lambda\mu & \text{for } i = i = 0 \\ (1-\lambda)(1-\mu) + \lambda\mu & \text{for } i \neq 0 \text{ and } k = i \\ (1-\lambda)\mu & \text{for } i \neq 0 \text{ and } k = i-1 \\ (1-\mu)\lambda & \text{for } \forall\, i \text{ and } k = i+1 \\ 0 & \text{otherwise} \end{cases}$$

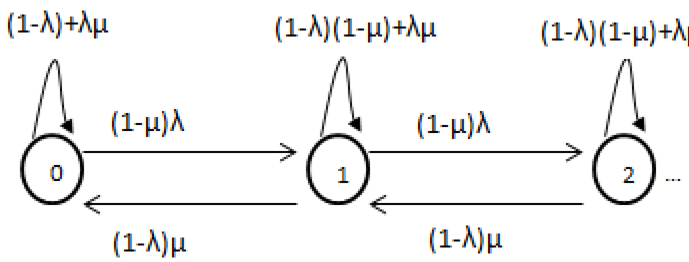


Figure 3: Transition diagram of discrete-time queue.

Then flow balance equations are:

$$\pi_0 = [((1-\lambda) + \lambda\mu)]\, \pi_0 + [(1-\lambda)\mu]\, \pi_1$$

$$\pi_1 = [(1-\mu)\lambda]\, \pi_0 + [(1-\lambda)(1-\mu) + \lambda\mu]\, \pi_1 + [(1-\lambda)\mu]\, \pi_2$$

...

$$\pi_i = [(1-\mu)\lambda]\, \pi_{i-1} + [(1-\lambda)(1-\mu) + \lambda\mu]\, \pi_i + [(1-\lambda)\mu]\, \pi_{i+1}$$

...

which yield

$\pi_i = \left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right] \pi_{i-1}$, $i >= 1$, or $\quad \pi_i = \rho^i \pi_0$, where $\rho = \frac{\lambda(1-\mu)}{\mu(1-\lambda)}$

Since steady-state probabilities add up to 1: $\sum_{i=0}^{\infty} \pi_i = 1, \pi_0 + \rho\, \pi_0 + \rho^2\, \pi_0 + ... = 1$.
Then $\pi_0 = 1 - \rho$, and $\pi_i = (1-\rho)\, \rho^i$, $i = 1, 2, \ldots$

In order to system to reach the steady-state the following should hold $\left[\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right] < 1$,

which is ensured under $\mu > \lambda$.

□

Proposition 1. For the discrete-time queue with parameters λ and μ, in steady-state expected waiting time in system for a customer is $\frac{1-\lambda}{\mu-\lambda}$ on expectation, and the variance of the waiting time in the system is $\frac{(1-\mu)(1-\lambda)}{(\mu-\lambda)^2}$.

Proof. Note that the service time for each customer is a geometric random variable with parameter μ, and the expected service time for a customer is $\frac{1}{\mu}$. An arriving customer who finds i customers in the system should wait $\frac{i}{\mu}$ time units plus her own service time $\frac{1}{\mu}$. We define:

W: the total waiting time in the system of a customer, who has arrived in steady-state,

Y_i: service time of i^{th} customer in the system i=1,2,… Note Y_i's are iid. Then,

$$W = \sum_{i=1}^{N} Y_i$$

where N is the random variable denoting the number of customers in the system in steady-state together with the arriving customer. Note N-1 ~geo (1-ρ). Then E[N] is the steady-state expected number of customers including the arriving customer,:

$$\mathrm{E[N]} = \sum_{i=0}^{\infty} (i+1)\pi_i,$$

$$\mathrm{E[N]} = \pi_0 \sum_{i=1}^{\infty} i\, \rho^{i-1} \quad .$$

Notice that $\sum_{i=1}^{\infty} i\rho^{i-1}$ is the derivative of $\sum_{i=0}^{\infty} \rho^{i}$ with respect to ρ where $\rho < 1$. Therefore, $\sum_{i=0}^{\infty} \rho^{i} = \frac{1}{1-\rho}$ and $\sum_{i=1}^{\infty} i\, \rho^{i-1} = \frac{d(\frac{1}{1-\rho})}{dp} = \frac{1}{(1-\rho)^2}$.

Expected value of W can be obtained from Wald's equality a

$$E[W] = E[Y]E[N] = \frac{1}{\mu}\pi_0 \frac{1}{(1-\rho)^2} = 1 + \frac{1-\mu}{\mu-\lambda}$$

The variance of the waiting time, W is

$$\mathrm{Var}(W) = \mathrm{Var}\,(\textstyle\sum_{i=1}^{N} Y_i) = \mathrm{Var}\,(\mathrm{E}[W|N])) + \mathrm{E}\,[\mathrm{Var}(W|N)] \qquad (2)$$

$$\mathrm{E}[W|N=n] = \mathrm{E}\,(\textstyle\sum_{i=1}^{n} Y_i\,) = \mathrm{E}\,(Y_1 + Y_2 + Y_3 + \ldots) = \frac{n}{\mu}$$

$$\mathrm{Var}\,(\mathrm{E}\,[W|N]) = \mathrm{Var}\,(\frac{N}{\mu}\,) = \frac{1}{\mu^2} Var(N) = \left(\frac{1}{\mu^2}\right)\frac{\rho}{(1-\rho)^2}$$

For the second term in Equation (2)

$$\mathrm{Var}\,(W|N=n) = \mathrm{Var}\,(\textstyle\sum_{i=1}^{n} Y_i\,) = n\frac{1-\mu}{\mu^2}$$

$$\mathrm{E}[\mathrm{Var}\,(W|N=n)] = \mathrm{E}[N\frac{1-\mu}{\mu^2}] = \left(\frac{1}{1-\rho}\right)\left(\frac{1-\mu}{\mu^2}\right)$$

Thus, $$\mathrm{Var}(W) = \frac{1}{\mu^2}\,\frac{\rho}{(1-\rho)^2} + \frac{1}{1-\rho}\,\frac{1-\mu}{\mu^2} = \frac{\rho+(1-\mu)(1-\rho)}{(1-\rho)^2\mu^2} = \frac{(1-\mu)(1-\lambda)}{(\mu-\lambda)^2}$$

□

According to quoted expected waiting time and the variance of the waiting time, customers either accept to purchase the service or they immediately balk from the system. Because not all arriving customers directly enter the system, the effective probability of customer arrival to the system is now:

$\lambda_e = \lambda H(p, w_1, w_2)$

Therefore, λ in the expressions shown in the proof of Proposition 1 should be replaced by λ_e. The steady-state expected waiting time and the variance of the waiting time under λ_e are:

$w_1(\lambda_e) = \frac{(1-\lambda_e)}{(\mu-\lambda_e)}$ and $w_2(\lambda_e) = \frac{(1-\lambda_e)(1-\mu)}{(\mu-\lambda_e)^2}$

Remark. One might ask whether expected waiting time for an arriving customer and the variance of waiting time can be obtained given that the service provider does not have any information regarding the system status. To obtain those values, a discrete-time Markov chain (MC) with the number of customers in the system as the state, is defined. Note that in steady-state, it is not necessary to keep track of real-time information to be able to announce w_1 and w_2 and to determine p. As long as the relationship between λ_e and w_1 and w_2, is known the provider is able to determine the optimal static price; the assumption that the service provider needs no information on system status does not violate the validity of the analysis.

Proposition 2. Under SPNIs scenario, given Θ_1~U[a,b], Θ_2=1,$c_1(w_1)$, $c_2(w_2)$ for a given p there exists a unique equilibrium arrival rate and it is $0 \leq \lambda_e \leq \lambda$.

Proof. As λ_e increases, the expected waiting time in the system, w_1, and the variance of the waiting time in the system, w_2, increase. Note that $c_1(w_1)$ and $c_2(w_2)$ are monotone non-decreasing functions of w_1 and w_2, respectively. Therefore $c_1(w_1)$ and $c_2(w_2)$ will also remain the same or increase. The structure of $H(p,w_1(\lambda_e), w_2(\lambda_e))$ implies that an increase in $c_1(w_1)$ and $c_2(w_2)$ results in the decrease of $H(p,w1(\lambda_e), w_2(\lambda_e))$. We know that the effective probability of arrival is $\lambda_e = \lambda H(p,w_1(\lambda_e), w_2(\lambda_e))$. Because $H(p,w_1(\lambda_e), w_2(\lambda_e))$ is continuous in [0,1], the domain and range of $\lambda H(p,w_1(\lambda_e), w_2(\lambda_e))$ is [0, λ]. Fixed Point Theorem suggests that for a continuous function there exists a fixed point which satisfies x=f(x). Since $H(.)$ is non-decreasing in [0,1], there exists a fixed point that satisfies $\lambda_e = \lambda H(p,w_1(\lambda_e), w_2(\lambda_e))$ which is unique, which we denote with λ_e.

□

The objective of the service provider is to set the price, p^*, that maximizes the profit obtained per period in the long run. We assume there is a waiting cost for the service provider per period. As long as there are customers waiting in the system, there is a penalty of customer dissatisfaction and waiting cost might be considered as this penalty. Notice that for the service provider it is not always better to accept all the customers because as customers wait in the system the service provider has to pay penalty. Let waiting cost is denoted with h.

Then the expected waiting cost is expressed as: $\sum_{i=0}^{\infty} hi\pi_i$ =h $\frac{\rho}{1-\rho}$ = $h\frac{\lambda(1-\mu)}{\mu-\lambda}$. Let Π_{SPNI}(p) denote the profit obtained per period in the long run. Then

$$\Pi_{SPNI} = \max_{p}\{ \lambda_e\, p \,-h\, \frac{\lambda_e(1-\mu)}{\mu-\lambda_e}\}$$

Proposition 3a. Let $\underline{p}$ =$\max_{p}\{p\colon \lambda_e = \lambda\}$ and $\overline{p}$= $\min_{p}\{p\colon \lambda_e = 0\}$. Then $\underline{p}$ is a lower bound on optimal price; p* and $\bar{p}$ is an upper bound on p*.

Proof. Note that H(p, w1(λ_e), w2(λ_e)) is a decreasing function of p. H(p,w_1(λ_e), w_2(λ_e)) is continuous in [0,1]. If the service provider quotes a price which is too low, due to the structure of H(p, w_1((λ_e), w_2(λ_e)) the probability of customers that join the system will be high. The highest value of H(p,w_1(λ_e), w_2(λ_e)) is 1 and there exists a p value, which ensures H(p, w_1(λ_e),w_2(λ_e)) equals to 1 and thus λ_e=λ. The maximum value of all price values, that yield H(.)=1, say p, must be a lower bound on the optimal price, since any price lower than p will decrease the revenue. Similar reasoning reveals, the minimum value of all price values that yield H(.)=0 must be an upper bound on the optimal price.

□

Proposition 3b. The equilibrium effective arrival rate, λ_e, is non-increasing in p.

Proof. Since H is non-increasing in p, λ_e, is non-increasing in p.

□

Example 1.

Suppose that for all arriving customers the value of the service is the same and equals to 10. The probability of a service completion in a system is; μ=0.75, arrival probability of a customer during a period is; λ=0.1. The service provider quotes a price;

p=3. Assuming Θ_1 is uniformly distributed between 1 and 2, Θ_2=1 and there is no waiting cost, we obtain H(p,w_1,w_2) as:

$$H(p, w_1, w_2) = \begin{cases} 0, & \dfrac{7 - \dfrac{(0.75)(1-\lambda_e)}{(0.75-\lambda_e)^2}}{1 + \dfrac{1-\lambda_e}{0.75-\lambda_e}} - 1 \leq 0 \\ \dfrac{7 - \dfrac{(1-\ 0.75)(1-\lambda_e)}{(0.75-\lambda_e)^2}}{1 + \dfrac{1-\lambda_e}{0.75-\lambda_e}} - 1, & 0 < \dfrac{7 - \dfrac{(0.75)(1-\lambda_e)}{(0.75-\lambda_e)^2}}{1 + \dfrac{1-\lambda_e}{0.75-\lambda_e}} - 1 < 1 \\ 1, & \dfrac{7 - \dfrac{(0.75)(1-\lambda_e)}{(07.5-\lambda_e)^2}}{1 + \dfrac{1-\lambda_e}{0.75-\lambda_e}} - 1 \geq 1 \end{cases}$$

If we solve $\lambda_e = \lambda H(p, w_1, w_2)$ under p=3, we obtain $\lambda_e=\lambda$=0.1. This value shows us that quoted price is too low compared to the reservation price and all arriving customers accept to purchase the product. If the quoted price increases, effective arrival rate first remains at value 0.1 and then decreases. Figure 4 shows the effective probability of arrival with respect to price

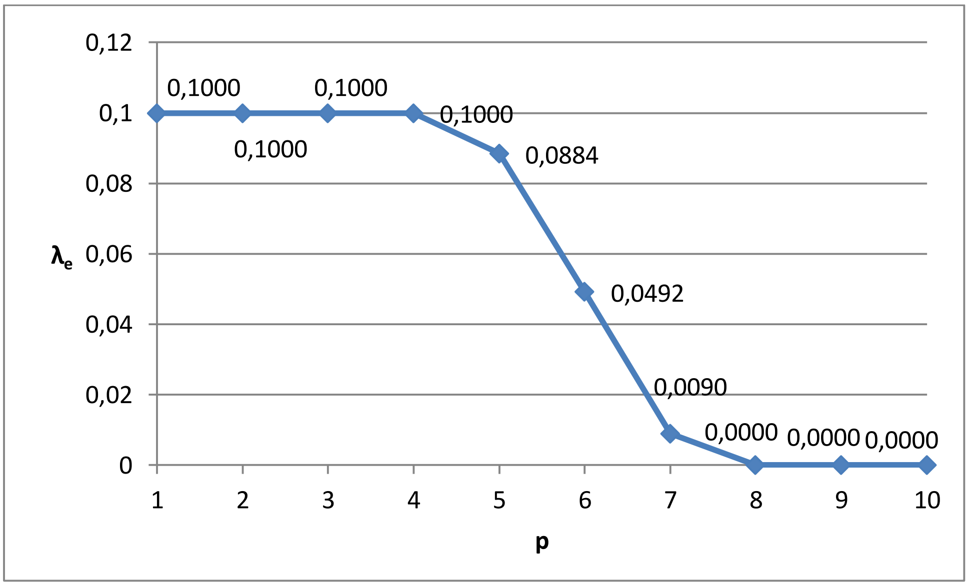


Figure 4: λ_e values with respect to price from 1 to 10.

In accordance with Proposition 3b, we observe that λ_e is a decreasing function of p. The service provider might choose to quote low prices to the arriving customers. As he quotes low prices the probability of customers to join the system is higher, but at the same time the profit that can be obtained is low. If the service provider chooses to quote higher prices the probability of customers to join the system is lower and although the quoted price is high, profits will possibly be low. Therefore, the service provider should decide an optimal price to maximize his profits.

Figure 5 illustrates the relationship between p and the profit. As shown below, up to some p value ($\underline{p}$ of Proposition 3) increasing p results in a linear increase in the profit (since λ_e=λ for p$\in$[0,$\underline{p}$]). Increasing p beyond $\underline{p}$ may or may not increase the profit. In the figure, we observe that an increase beyond p, decreases the profit. Therefore, quoting customers a low price or a too high price is a disadvantage for the service provider.

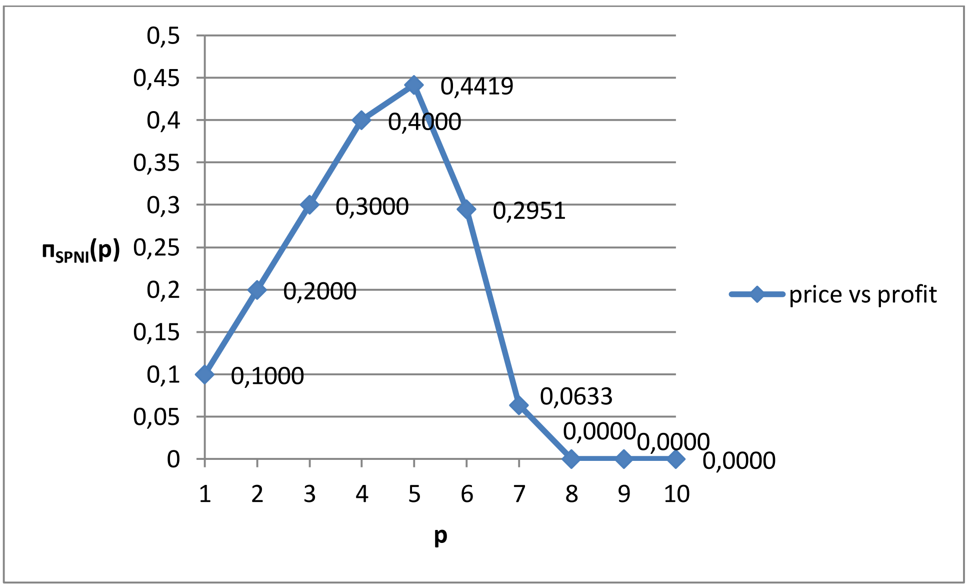


Figure 5: Profit as a function of the quoted price.

4.2 Static pricing under partial information (SPPIs)

Under this scenario, the service provider only knows about the number of customers in the system and each arriving customer is announced a delay (expected waiting time) and the precision of the delay (variance of the waiting time) based on the number of customers in the system upon arrival. Like SPNIs case, we assume all customers are quoted the same price *p*, but expected waiting time and the variance of waiting time are customer specific, i.e., an arrival who sees *i* customers waiting is quoted w_1^i and w_2^i. A customer accepts price *p* if her utility is greater than or equal to 0. The probability that a customer arriving to the state *i*, quoted *p*, purchases the service is $H(p, w_1^i, w_2^i)$. Notice that for each arriving customer, the probability of purchasing the service may now be different depending on the system status. Again, we define the system state X_n as the number of people in the system in period n. According to the quoted price and delay information, arrival to state *i* occurs with probability λ_e^i (p)= $\lambda H(p, w_1^i, w_2^i)$ where *i*=0,1,2,... . We obtain w_1^i and w_2^i as follows.

Let W_i be a random variable denoting time for waiting for *i* customers in the system including the arriving customer:

$$W_i = \sum_{j=1}^{i+1} Y_j$$

The announced delay w_1^i for an arriving customer who finds *i* customers in the system is:

$$w_1^i = E[W_i] = \frac{i+1}{\mu}$$

The precision w_2^i of the announced delay who finds *i* customers in the system is:

$$w_2^i = \text{Var}(W_i) = (i+1)\frac{1-\mu}{\mu^2}$$

Now, the effective arrival rate is different at each state; a customer arrives with probability λ but effective probability of arrival at state *i* is:

$\lambda_e^i = \lambda H(p, w_1^i, w_2^i)$ (3)

As the number of customers in the system increases, the expected waiting time in the system increases and the precision for the announced waiting time decreases (variance increases). Thus the probability of accepting to purchase the service decreases as the number of customers waiting in the system increases. So, we expect smaller effective arrival rates as the number of customers waiting in the system increases.

Proposition 4: There exists an $\overline{N}$, such that arrival rate to state $i \geq \overline{N}$ is 0.

Proof: As the number of people in the system increases, effective arrival rate decreases. This is because *H(p,* w_1^i, w_2^i*),* is monotone decreasing in *i*. Due to the structure of *H(.)* there exists a state at which effective arrival probability is 0. Notice that when the number of customers in the system increases, the variance of the waiting time in the system increases and the probability of a customer to purchase the service decreases. Similarly, as the number of customers in the system increases, the service provider quotes longer waiting time (w_1^i increases as the number of customers in the system increases) and the probability of customers to purchase the service decreases.

□

As the arriving customers are quoted a price, p, w_1^i and w_2^i, they either choose to enter the system or immediately balk from the system. The state transition diagram of the corresponding DTMC is given in Figure 6.

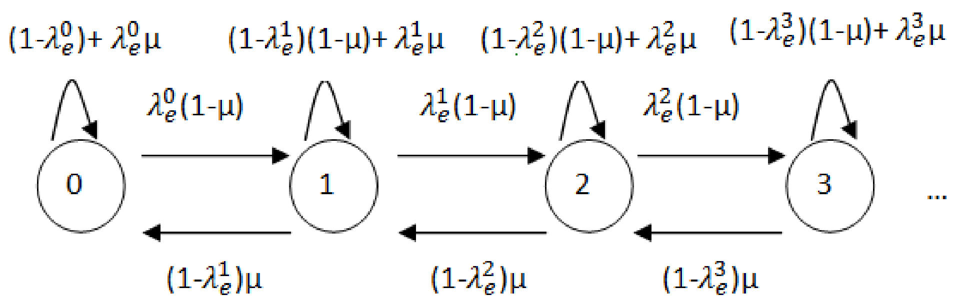


Figure 6: Transition diagram for SPPIs scenario.

The transition probability function is expressed as follows:

$$P(X_{n+1} = k|X_n = i) = \begin{cases} (1-\lambda_e^i) + \lambda_e^i \mu & \text{for } i = k = 0 \\ (1-\lambda_e^i)(1-\mu) + \lambda_e^i \mu & \text{for } i \neq 0 \text{ and } k = i \\ (1-\lambda_e^i)\mu & \text{for } i \neq 0 \text{ and } k = i-1 \\ (1-\mu)\lambda_e^i & \text{for } k = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Note that in (3), the effective arrival probability λ_e^i is a function of *i* as well as the quoted price *p*. The objective of the service provider is to maximize his profit. Therefore, he should select the price *p* such that the profit obtained per period is maximized.

$$\Pi_{SPPI} = \max_p \{\textstyle\sum_i \pi_i \; \lambda_e^i p - \; ih) \}$$

where π_i's depend on the quoted price.

Example 2.

Suppose that all the arriving customers has a constant reservation price; *r*= 10, the probability of a service completion in a system is; μ=0.75, arrival probability of a customer is; λ=0.1, Θ_1~U[1,2] and Θ_2=1. Assuming there is no waiting cost, we analyze the effect of quoting different prices on the effective arrival rate for the customer, who finds 2 customers in the system. Figure 7 shows λ_e^2 values with respect to *p* values with price set {1,2,…,10}.

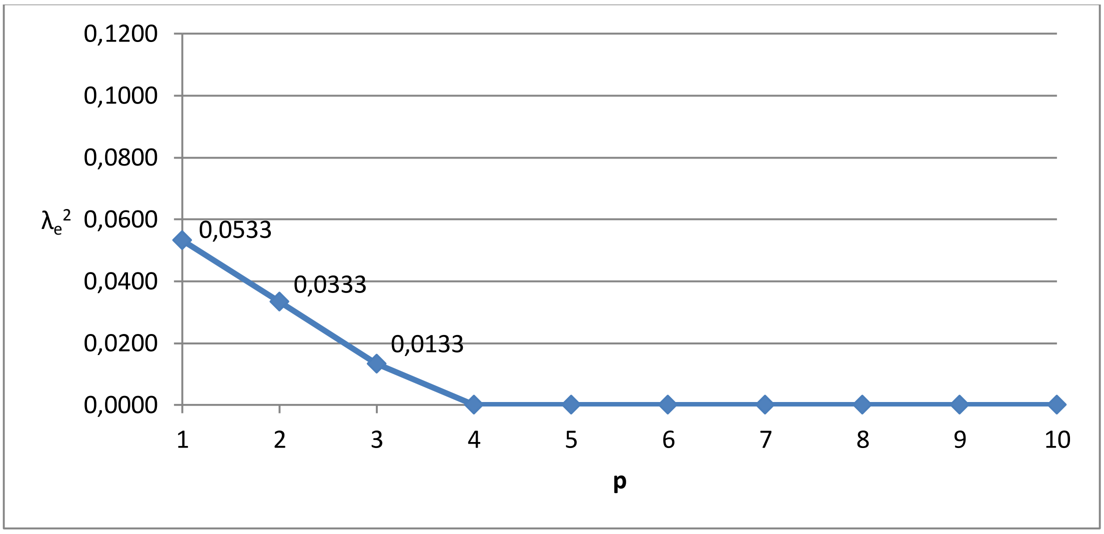


Figure 7: λ_e^2 values in response to p.

From Figure 7 we observe that as price increases the effective arrival rate decreases. For high prices effective arrival rate becomes 0 which corresponds to the case that no customers accept to purchase the service.

Profit per period in response to quoted prices may be observed in Figure 8. Notice that, as for SPNIs scenario, quoting price too low or too high may cause the service provider lose profit.

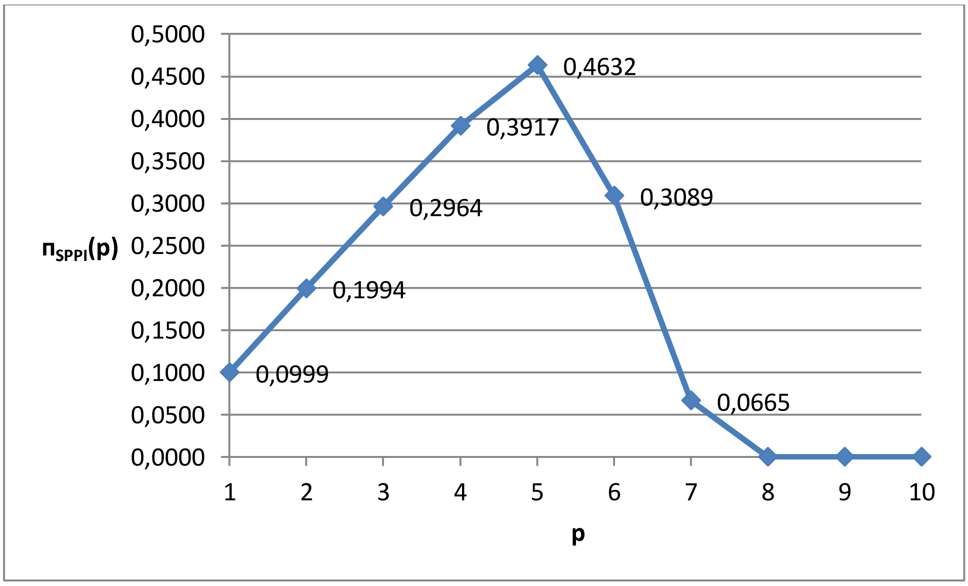


Figure 8: Profit as a function of the quoted price.

4.3 Static pricing under full information (SPFIs)

This scenario corresponds to the case where the service provider has full information regarding the state of the system. Provider is able to observe the total system workload. Since the information scheme is symmetric this information is fully shared with the arriving customer. To facilitate the analysis, we treat each arriving customer as an arriving workload to the system. In the corresponding Markov chain we define WO_n as the total workload of the system in period n. Each customer arrives to the system and server completes a generic job with probability μ in a period. This is equivalent to saying that, the workload brought by each customer has geometric distribution with parameter μ while server completes the job with probability 1 in each period. Depending on both the total workload of the system (WO_n) and the workload of the arriving customers in period n (Y_n) the service provider is able to observe the total workload and he announces the exact waiting time. Notice that previously we defined Y_i as the time for waiting time for just the service of i^{th} customer in the system. Actually, Y_i and Y_n are independent and identically distributed random variables. Under SPFIs, an arriving customer, who finds a workload of *wo* and brings a workload of j will be announced $w_1^{(wo,j)} = wo+j$. The expected waiting time is simply the current workload in the system. Since there does not exist any uncertainty inherent in this announcement, $w_2^{(wo,j)}$ will be 0. Then we may define the probability of a customer to purchase the service as

$H(p, w_1^{(wo,j)}, w_2^{(wo,j)})$.

We observe that the probability of customers to purchase the service decreases as the total workload of the system increases. In order to maximize the steady-state expected profit per period, the service provider must determine a static price to quote to all arriving customers. In order to find the price, the effect of a quoted price on the steady state probabilities should be analyzed. We determine the effective arrival rate, at each state, where state is defined as the total workload in the system. Let:

WO_n: total workload of the system at period n. $WO_n \in \{0,1, 2,..\}$

b_j: probability that the arriving customer brings workload j, $b_j = P(Y_n=j) = (1-\mu)^{j-1}\mu$, j=1,2,3..,

$\lambda_{wo,j}(p)$: the probability that an arrival with j amount of workload enters if the quoted price is p, at state wo=0,1,.. j=1, 2,…

$$\lambda_{wo,j}(p) = \lambda b_j H(p,wo+j,0)$$

$\lambda_{wo}(p)$: effective arrival probability to the system at state is wo given that the quoted price is p,

$$\lambda_{wo}(p) = \sum_{j=1}^{\infty} \lambda_{wo,j}(p)$$

We may think $\lambda_{wo,j}$'s as the effective arrival probability of j amount of workload. We may introduce the transition probabilities as follows:

$$P(WO_{n+1} = wo_2 | WO_n = wo_1)$$

$$= \begin{cases} \lambda(p) b_{wo_2 - wo_1 + 1} H(p, wo_2 + 1\ ,0) & \text{for } wo_2 \geq wo_1, \text{wo}_1 = 1,.. \text{or } wo_2 = 0, wo_2 > 0 \\ (1 - \lambda_{wo2}(p)) + \lambda b_1 H(p, 1\ ,0) & \text{for } wo_1 = wo_2 = 0 \\ (1 - \lambda_{wo2}(p)) & \text{for } wo_1 \neq 0, wo_2 = wo_1 - 1 \\ 0 & \text{otherwise} \end{cases}$$

In the following, we show part of the transition diagram (in fact in each state infinitely many number of arcs should leave)

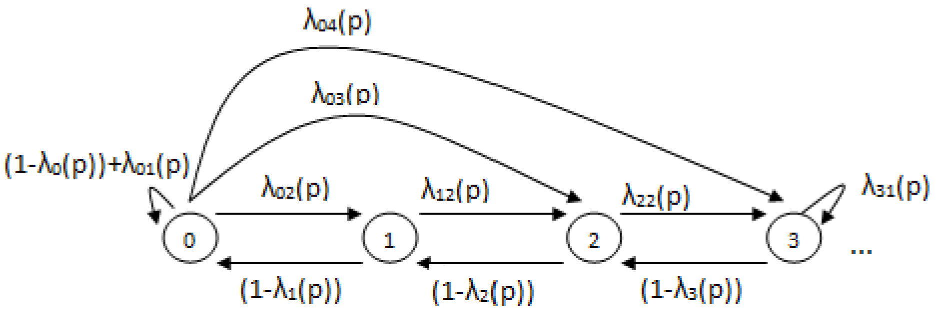


Figure 9: Transition diagram for SPFI scenario.

The steady state values may be directly found from flow balance equations. Let π_{wo} denote the steady-state probability that workload is *wo*. The steady-state profit per period is expressed as follows:

$$\Pi_{SPFI} = max_p \{\sum_{wo} \pi_{wo}(SPFI, p) (\sum_j \lambda_{wo,j}(p)(p - h(wo + j - 1))\}$$

Example 3.

Suppose for a system the customers' arrival probability is; λ=0.1, for each arriving customer the reservation price; r=10, the probability of service completion; μ=0.75, Θ_1~U[1,2], Θ_2=1 and there is no waiting cost. For a customer who brings 2 units of workload and already finds 2 units of system workload, we calculate the probability of entrance of this customer to the system; $\lambda_{22}(p)$ with respect to quoted prices. Notice that, since the workload of the arriving customer is 2 and finds 2 units of workload to be completed, the customer will be announced 4 units of waiting time. Figure 10 shows $\lambda_{22}(p)$ values with respect to p value. As the quoted prices increase, we observe that the probability of customer to purchase the service decreases. Although we illustrate the example for $\lambda_{22}(p)$ the relationship is valid for all $\lambda_{ij}(p)$ values.

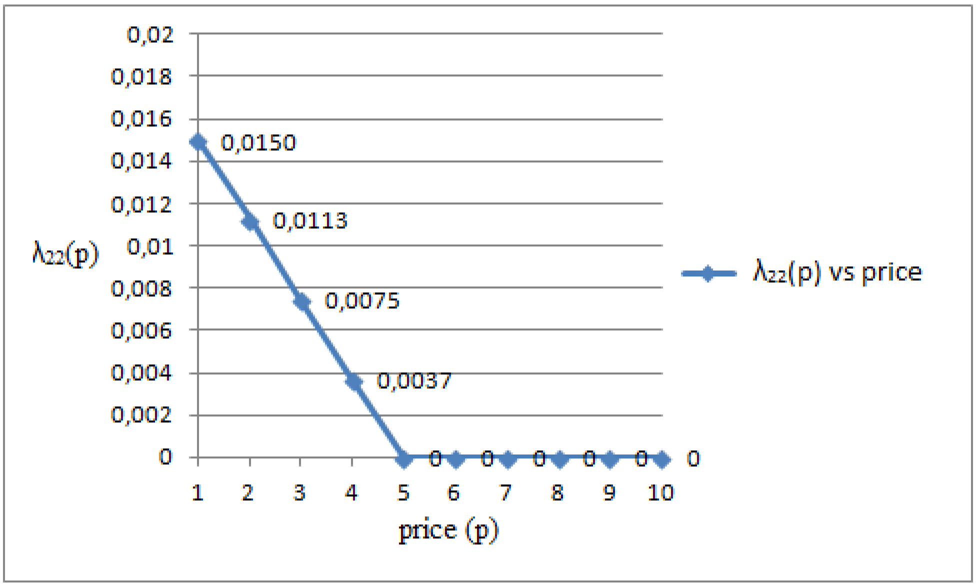


Figure 10: $\lambda_{22}(p)$ values in response to p

Again, as for the other SPs scenarios, we observe that up to a price threshold, the server is able to increase his profits. There is a peak price point where the service provider maximizes his profits, and prices higher than the threshold just decreases the profit that the server can obtain.

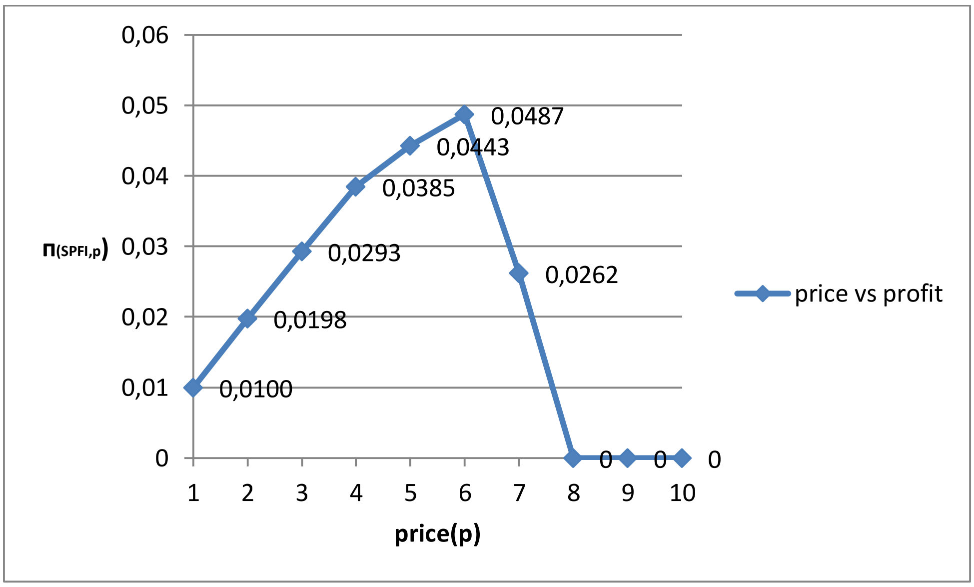


Figure 11: Profit with respect to price

4.4 Dynamic pricing under no information (DPNIs)

The scenario corresponds to the case where each arriving customer is announced the same w_1 and the same w_2. Since the service provider has the same amount of knowledge with the customers and each arriving customer is announced the same lead-time information, the prices that should be quoted to each customer should be the same. This implies that DPNIs scenario is equivalent to SPNIs scenario.

4.5 Dynamic pricing under partial information (DPPIs)

Under this scenario, the service provider can only have information regarding the number of customers in the system and makes delay announcement that fully reveals this information. Service provider may quote different prices to customers depending on the system state. Therefore, the probability of a customer to purchase the service may not vary solely due to the announced delay and its precision anymore, but also due to the quoted price. As in the other scenarios, the aim of the service provider is to maximize his profit per period. We model service provider's problem through stochastic dynamic programming. We consider the quotation of prices as the possible actions.

Defining $v(i)$ as the bias function under optimal policy and g as the expected gain per period under the expected average reward criteria the optimality equation is:

$v(i)+g=$

$$
\begin{cases}
\max\limits_{p\in P}\ \{\lambda_e^i(\mathrm{p})p - ih + \lambda_e^i(p)\mu\,\mathrm{v}(i) + \lambda_e^i(p)(1-\mu)\mathrm{v}\,(i+1) + \\
\left(1-\lambda_e^i(p)\right)\mu\mathrm{v}\,(i-1) \;\; + \left(1-\lambda_e^i(p)\right)(1-\mu)\mathrm{v}(i)\}\,, i \neq 0 \\
\max\limits_{p\in P}\{\lambda_e^i(p)p \;-\; ih + \lambda_e^i\;(p)\mu\,\mathrm{v}(i) + \lambda_e^i(p)\left(1-\mu\right)\mathrm{v}\;\;(i+1) + \left(1-\lambda_e^i(p)\right)\mathrm{v}(i)\Big\}, i = 0
\end{cases}
$$

The transition probabilities under price p are as follows:

$$
\mathrm{P}(\mathrm{k}|\mathrm{i},p)= \begin{cases}
\lambda_e^i(p)\mu \;+\left(1-\lambda_e^i(p)\right) \;\text{ for i} = \mathrm{k} = 0 \\
\lambda_e^i(p)\mu + \left(1-\lambda_e^i(p)\right)(1-\mu) \;\text{ for i } \neq 0 \text{ and k} = \mathrm{i} \\
\lambda_e^i(p)(1-\mu) \text{ for i} \geq 0 \text{ and k} = \mathrm{i}+1 \\
\left(1-\lambda_e^i(p)\right)\mu \;\text{ for i } \neq 0 \text{ and k} = \mathrm{i}-1 \\
0 \text{ otherwise}
\end{cases}
$$

The single stage expected reward function:

$\mathrm{r(i,p)} = \lambda_e^i(p)p - ih$, i≥0

Notice that our model satisfies the conditions for the application of MDP. Although system state goes to infinity, we observe that at some system state customers start to not to accept to join the system and therefore there exists a state that the system state will not go after. For all the scenarios, this observation is valid which implies that for all DP models, conditions for the use of MDP are satisfied.

Example 4.

Consider a system where the customers' arrival probability is λ=0.1, the service completion probability is μ=0.75. Assuming there is no waiting cost, all customers are partially informed about the delay, we observe the following pricing quotations with respect to the number of customers in the system:

Table 3: The quoted prices with respect to the number of customers in the system (excluding the arriving customer)

Number of customers in the system	0	1	2	3	4
Quoted prices to the customers	5	3	2	1	10

As observed from Table 3, as the number of customers in the system increases the service provider chooses to quote lower prices to the arriving customers. This is because; as the number of customers in the system increases the service provider announces higher expected waiting times. If he, at the same time, quotes higher prices, the probability of customers to join the system becomes too low and thus the provider's expected profit significantly decreases. After some point, however, the provider chooses to quote the higher prices and reject the customers. By rejecting an arriving customer the provider tries to keep the system state at some level and thus be always able to announce short waiting time.

4.6 Dynamic pricing under full information (DPFIs)

In this scenario, the service provider has perfect information regarding the workload in the system and reveals this information fully to the customers. Workload corresponds to the service time. Arriving customers are quoted a price and because the exact workload is known by the service provider, customers are announced the exact waiting time. As in the SPFI scenario, we consider each arriving customer as a workload and the workload brought by each customer is geometrically distributed with parameter μ. Defining (WO_n, Y_n) as the total workload of the system in period n and the workload brought by the arriving customer, we again resort to MDP to solve service provider's problem and introduce the recursive function.

$v(w_0,j)+g=$

$$\begin{cases} max_{p\varepsilon P}\{\mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0)(\mathrm{p}-\mathrm{h}(\mathrm{wo}+\mathrm{j}-1))+ \\ \mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0)\sum_{\mathrm{k}=1}^{\infty}\lambda \mathrm{b}(\mathrm{k})\mathrm{v}\,(\mathrm{wo}+\mathrm{j}-1, \mathrm{k})+ \\ \mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0)(1-\lambda)\mathrm{v}\,(\mathrm{wo}+\mathrm{j}-1, 0)+ \\ \big(1-\mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0)\big)\sum_{\mathrm{k}=1}^{\infty}\lambda \mathrm{b}(\mathrm{k})\mathrm{v}\,((\mathrm{wo}-1)^{+}, \mathrm{k}) \\ +(1-\mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0))(1-\lambda)\mathrm{v}\,((\mathrm{wo}-1)^{+}, 0)\} \quad \text{if wo} \geq 0,\ \mathrm{j} \neq 0 \\ max_{p\varepsilon P}\{\mathrm{H}(\mathrm{p}, \mathrm{wo}+\mathrm{j}, 0)(\mathrm{p}-\mathrm{h}(\mathrm{wo}+\mathrm{j}-1))+ \\ \sum_{\mathrm{k}=1}^{\infty}\lambda \mathrm{b}(\mathrm{k})\mathrm{v}\,((\mathrm{wo}-1)^{+}, \mathrm{k})+(1-\lambda)\mathrm{v}\,((\mathrm{wo}-1)^{+}, 0)\} \quad \text{if wo} \geq 0,\ \mathrm{j}=0 \end{cases} \tag{4}$$

where g and v(*wo*,*j*) denote the optimal gain and bias functions, respectively.

The expected reward is:

$$\mathrm{r}(\mathrm{wo}, \mathrm{j}, \mathrm{p}) = \begin{cases} \mathrm{H}(\mathrm{p}, wo+j, 0)(\mathrm{p}-h(wo+j-1)) & \text{if } j>0 \\ 0 & \text{if } j=0 \end{cases}$$

Based on Equation (4) we have the following transition probabilities:

$P(W_{n+1}=wo_2,\ Y_{n+1}=j_2 \mid W_n=wo_1,\ Y_n=j_1,\ p)$

$$= \begin{cases} \big(1-\mathrm{H}(\mathrm{p}, wo_1+j_1, 0)\big)\lambda \mathrm{b}(j_2) & \text{for } j_1 \neq 0, wo_2=(wo_1-1)^{+}, j_2 \neq 0 \\ (1-\mathrm{H}(\mathrm{p}, wo_1+j_1, 0)(1-\lambda) & \text{for } j_1 \neq 0, wo_2=(wo_1-1)^{+}, j_2=0 \\ \lambda \mathrm{b}(j_2) & \text{for } j_1=0, wo_2=(wo_1-1)^{+}, j_2 \neq 0 \\ (1-\lambda) & \text{for } j_1=0, wo_2=(wo_1-1)^{+}, j_2=0 \\ \mathrm{H}(\mathrm{p}, wo_1+j_1, 0)\lambda \mathrm{b}(j_2) & \text{for } j_1 \neq 0, wo_2=wo_1+j_1-1, j_2 \neq 0 \\ \mathrm{H}(\mathrm{p}, wo_1+j_1, 0)(1-\lambda) & \text{for } j_1 \neq 0, wo_2=wo_1+j_1-1, j_2=0 \end{cases}$$

Example 5.

Suppose we have a system such that the customers' arrival probability is λ=0.1, the probability of service completion is μ=0.75 and the reservation price for all the arriving customers is 10. Assuming there is no waiting cost, and customers are announced the exact waiting time, the quoted prices with respect to system state is provided in Table 4:

Table 4: The quoted prices with respect to arriving workload and system workload.

Arriving workload (Y_n) / System Workload (WO_n)	1	2	3	4	5	6	7	8	9	10
0	8	6	4	4	3	3	8	4	4	6
1	6	4	4	3	3	7	9	5	8	9
3	3	3	3	4	4	4	10	10	10	6
5	2	2	10	10	10	10	10	10	10	10

From Table 4 we observe that the service provider chooses to quote lower prices as the arriving workload amount increases. However, this decrease continues up to some level. After this level, making comment on the policy of price quotation becomes difficult. We are not able to observe a monotonic behavior for the price quotation. However, we can conclude that for the two customers that bring the same amount of workload, the provider often quotes lower price to the customer that arrives to a more crowded system and higher price to the other customer that arrives to a relatively empty system. If both the system workload and the arriving workload are high, on the other hand, the provider chooses to quote a higher price and rejects the arriving customer.

CHAPTER 5

ASYMMETRIC INFORMATION

In this setting, the service provider has perfect information and chooses the amount of information to reveal to the customers. The information is revealed in the form of delay announcements (w_1) and the precision of the delay (w_2).

5.1 Static pricing no information is revealed (SPNIa)

Under this scenario, we assume the service provider has perfect information regarding the system status but all arriving customers are announced the same delay and precision of delay. We present the following property.

Conjecture 1. The scenarios SPNIa and SPNIs are equivalent. In other words, if customers are not revealed delay information, then the service provider is indifferent between having no and perfect information on system status. Additional information does not any benefit to the provider.

Given that the prices under both systems are identical, in steady-state an arriving customer to either system will face a waiting time which has the same distribution. This implies, if all customers are to be quoted the same expected waiting time and the precision, those announcements must be the same under SPNIa and SPNIs.

□

Note that the results can be extended to the case where service provider has partial information regarding system status but reveals no information to the customers. Again, in that case whether the service provider has partial information on system status or no information would not make any difference in terms of customer response. Since in steady-state SPNIa and SPNIs are equivalent, we adopt the model of SPNIs for this scenario.

5.2 Static pricing partial information (SPPIa)

The service provider has perfect information and decides to reveal partial information to the customers. A customer is announced the delay and the precision of the delay based on the number of customers in the system. Similar to the result in Conjecture 1, because static pricing policy is adopted, extra information known by the service provider makes no difference on the profit. All customers are quoted the same price and customers choose to accept to purchase the service based on the announced delay (expected waiting time) and the precision of delay (the variance of the waiting time). Therefore, the model is equivalent to the SPPIs.

5.3 Static pricing full information (SPFIa)

The service provider has perfect information and all the customers are announced the exact waiting time. Therefore, by definition SPFIa and SPFIs are equivalent models.

5.4 Dynamic pricing under no information (DPNIa)

This scenario corresponds to the case which the service provider has perfect information about the total system workload, but all customers are announced the same delay (expected waiting time) and the precision of the delay (variance of delay). According to the system state price quotations to each arriving customer might differ. Defining the system state as (Wo_n, Y_n) (the system workload, the workload of the arriving customer), the following are introduced:

w_1: announced delay (steady-state expected system workload) to the arriving customer

w_2: announced precision of delay (steady-state variance of system workload announced to the arriving customer)

$$w_1=\sum_{\mathrm{wo}=0}\sum_{j=0}(wo+j)\pi(wo,j)_{(\mathrm{a})} \tag{5}$$

$$w_2=\sum_{wo=0}\sum_{j=0}(wo+j)^2\pi(wo,j)_{(\mathrm{a})} - w_1^2 \tag{6}$$

From (5) and (6), it may be observed that the steady-state delay information depend on the quoted prices. As the service provider quotes a different price to each state we solve the service provider's problem by MDP and introduce the following optimality equation.

$$\begin{cases} max_{p\varepsilon P}\{\mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2)(\mathrm{p}-\mathrm{h}(\mathrm{wo}+\mathrm{j}-1)) + \\ \mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2)\sum_{\mathrm{k}=1}^{\infty}\lambda\mathrm{b}(\mathrm{k})\mathrm{v}\,(\mathrm{wo}+\mathrm{j}-1,\mathrm{k}) + \\ \mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2)(1-\lambda)\mathrm{v}\,(\mathrm{wo}+\mathrm{j}-1,0) + \\ \left(1-\mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2)\right)\sum_{\mathrm{k}=1}^{\infty}\lambda\mathrm{b}(\mathrm{k})\mathrm{v}\,((\mathrm{wo}-1)^+,\mathrm{k}) \\ +(1-\mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2))(1-\lambda)\mathrm{v}\,((\mathrm{wo}-1)^+,0)\} \quad \text{if wo} \geq 0,\ \mathrm{j}\neq 0 \\ max_{p\varepsilon P}\{\mathrm{H}(\mathrm{p},\mathrm{w}_1,\mathrm{w}_2)(\mathrm{p}-\mathrm{h}(\mathrm{wo}+\mathrm{j}-1)) + \\ \sum_{\mathrm{k}=1}^{\infty}\lambda\mathrm{b}(\mathrm{k})\mathrm{v}\,((\mathrm{wo}-1)^+,\mathrm{k}) + (1-\lambda)\mathrm{v}\,((\mathrm{wo}-1)^+,0)\} \quad \text{if wo} \geq 0,\ \mathrm{j}=0 \end{cases} \tag{7}$$

subject to Equation (5) and Equation (6).

As prices are quoted dynamically, it is challenging to calculate the steady-state delay information and thus solve the above optimality equation. In order to find the steady-state expected waiting time and the steady-state variance of the waiting time in the system and thus be able to solve MDP defined in (7) we propose the following algorithm.

The algorithm steps are:

- ***Step 1.*** Let iter=0. Assign values to $w_1(iter=0)$, and $w_2(iter=0)$.
- ***Step 2.*** Given $H(p,w_1(iter),w_2(iter))$ solve the MDP (in Equation 4). Find the quoted price to each state under the corresponding optimal pricing policy. Let the vector of prices under the corresponding policy be denoted with $\bar{\boldsymbol{p}}$(iter). If $\bar{\boldsymbol{p}}$(iter)= $\bar{\boldsymbol{p}}$(iter-1), go to Step 3. Else go to Step 4.
- ***Step 3.*** Return the optimal expected profit.
- ***Step 4.*** Let iter=iter+1. Using $H(p,w_1(iter-1),w_2(iter-1))$ determine steady-state probabilities of states and through Equations (5) and (6) determine $w_1(iter)$, $w_2(iter)$. Go to Step 2.

□

Note that the algorithm does not guarantee to find an optimal solution. Also, there might be some cases that no convergence is observed.

Example 6.

Suppose we have a system such that the customers' arrival probability is λ=0.1, the probability of service completion is μ=0.75 and the reservation price for all the arriving customers is 10. The service provider has perfect information but choose to announce the steady state expected waiting time and the precision of the waiting time and can quote any price from the set p={1,2,..10}. The quoted prices with respect to system state when waiting cost is 0.5 is provided in Table 5.

Table 5: The quoted prices to customers with respect to arriving workload and system workload when h=0.5

Arriving workload (Y_n) / System Workload (X_n)	1	2	3	4	5	6	7	8	9	10
0	5	5	5	5	5	5	6	6	6	7
1	5	5	5	5	5	6	6	6	7	10
3	5	5	5	5	6	6	7	9	9	8
5	5	5	6	6	7	8	10	10	8	8
7	5	6	6	7	7	9	9	8	-	-
10	6	-	-	-	-	-	-	-	-	-

From Table 5, we observe that if there is waiting cost, for a constant arriving workload value, as the system workload increases the provider chooses to increase prices and after some point always chooses to reject customers. For a constant system workload, however, we cannot observe a monotonic price quotation behavior. We are not able to make a conclusion about the relationship between Y_n and price quotations.

□

5.5 Dynamic pricing under partial information (DPPIa)

Under this scenario, we assume the service provider has perfect knowledge about the system, he knows both the number of customers in the system, and the workload of each arriving customer. However, customers are only announced the expected waiting time and the variance of the expected waiting time in terms of the number of customers in the system. For the model we introduce the following:

wo_i^n: workload of the i^{th} customer currently in the system, $wo_i^n > 0$. Note wo_i^n denotes the remaining workload of the customer currently in the service at the beginning of period n.

$\overline{WO}_n$: a vector denoting workloads of customers in the system in period n (not including the arriving customer). $\overline{WO}_n = (wo_1^n, wo_2^n, \ldots wo_{X_n}^n)$ We assume $\overline{WO}_n$ is a vector with a dynamic dimension. When there is at least one customer, all elements of the vector are strictly positive. The dimension of the vector is determined by the number of customers in the system. If there are no customers in the system then $||\overline{WO}_n||=0$ otherwise, $||\overline{WO}_n||=X_n$

Note X_n is a function of workload vector $\overline{WO}_n$, and can be directly derived from $\overline{WO}_n$. We explicitly include X_n in the state definition for clarity. Note that X_{n+1} is either X_n-1 or X_n, or X_n+1.

$v(\overline{WO}_n, j, i.)$: the bias function under optimal policy given that the state is $(\overline{WOn}, j, i)$, Then:

$v(\overline{WO_n}, j, i.) = v((wo_1^n, wo_2^n, \ldots wo_i^n), j, i)$

The related function can be found on pages 59 and 60 .

5.6 Dynamic pricing under full information (DPFIa)

Since all arriving customers are perfectly informed, the model is the equivalent of DPFI under the case; the service provider has the same amount of information with the customers. For the details you may refer to DPFIs scenario.

v((wo$_1$,wo$_2$,…wo$_i$),j,i)+g=

$$
\begin{cases}
\max_{p\varepsilon P}\Big\{ H(p,w_1^i,w_2^i)p - ih + H\big(p,w_1^i,w_2^i\big)\sum_{j_2=1}^{\infty}\Big(\lambda b(j_2)v\big((wo_2\,,wo_3\,,..wo_L\,,j),j_2,i\,)\big)\Big) + \\
\big(1 - H(p,w_1^i,w_2^i)\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,((wo_2\,,wo_3\,,..wo_L\,),j_2,i-1\,)) + \\
\big(1 - H(p,w_1^i,w_2^i)\big)(1-\lambda)v\,\big((wo_2\,,wo_3\,,..wo_L\,),0,i-1\,)\Big\} \text{ if } wo_1\; = 1 \text{ and } j > 0 \\
\max_{p\varepsilon P}\Big\{ H(p,w_1^i,w_2^i)p - ih + \sum_{j2=1}^{\infty}\Big(\lambda b(j_2)v\big((wo_2\,,wo_3\,,..wo_L\,),j_2,i-1\,)\big)\Big) + \\
(1-\lambda)v\big((wo_2\,,wo_3\,,...wo_L\,),0,i-1\big)\Big\} \quad \text{if } wo_1\; = 1 \text{ and } j = 0 \\
\max_{p\varepsilon P}\Big\{ H(p,w_1^i,w_2^i)p - ih + H\big(p,w_1^i,w_2^i\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,((wo_1-1,wo_2\,,..wo_L\,,j\,),j_2,i+1\,)) + \\
H\big(p,w_1^i,w_2^i\big)(1-\lambda)v\,((wo_1-1,wo_2\,,..wo_L\,,j\,),0,i+1\,) + \\
\big(1 - H(p,w_1^i,w_2^i)\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)\,v\,((wo_1-1,wo_2\,,..,.wo_L\,),j_2,i\,) + \\
\big(1 - H(p,w_1^i,w_2^i)\big)(1-\lambda)v\,\big((wo_1-1,wo_2\,,..,wo_L\,),0,i\,)\Big\} \text{ if } \quad wo_1\; \geq 2 \text{ and } j > 0 \\
\max_{p\varepsilon P}\Big\{ H(p,w_1^i,w_2^i)p - ih + \sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,((wo_1-1,wo_2\,,..,.wo_L\,),j_2,i\,) + \\
(1-\lambda)v\,\big((wo_1-1,wo_2\,,..,.wo_L\,),0,i\,)\Big\} \text{ if } wo_1\; \geq 2 \text{ and } j = 0
\end{cases}
\tag{8}
$$

Equation 8 continues:

$$\begin{cases} \max_{p\varepsilon P}\{\, H(p,w_1^i,w_2^i)p - ih + H\big(p,w_1^i,w_2^i\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,((0),j_2,0\,) + H\big(p,w_1^i,w_2^i\big)(1-\lambda)v\,((0)\,,0,0\,) + \\ \big(1 - H(p,w_1^i,w_2^i)\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)\,v\,((0),j_2,0) + \big(1 - H(p,w_1^i,w_2^i)\big)(1-\lambda)v\,\big((0),0,0\big)\big\} \quad \text{if} \quad i = 0 \text{ and } j = 1 \\ \max_{p\varepsilon P}\{\, H(p,w_1^i,w_2^i)p - ih + H\big(p,w_1^i,w_2^i\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,(j-1,j_2,1\,) + H\big(p,w_1^i,w_2^i\big)(1-\lambda)v\,(\,j-1,0,1) + \\ \big(1 - H(p,w_1^i,w_2^i)\big)\sum_{j_2=1}^{\infty}(\lambda b(j_2)\,v\,((0),j_2,0) + \big(1 - H(p,w_1^i,w_2^i)\big)(1-\lambda)v\,\big((0),0,0\big)\big\} \quad \text{if} \quad i = 0 \text{ and } j > 1 \\ \max_{p\varepsilon P}\{\, H(p,w_1^i,w_2^i)p - ih + \sum_{j_2=1}^{\infty}(\lambda b(j_2)v\,((0),j_2,0\,) + (1-\lambda)v((0),0,0)\} \quad \text{if } i = 0 \text{ and } j = 0 \end{cases}$$

□

CHAPTER 6

COMPUTATIONAL ANALYSIS

This section presents the results of computational study and analyzes under which conditions which schemes are better for the service provider. Through this analysis, we aim to analyze whether adopting a more or less informative scheme is better for the service provider. Also, we would like to quantify the benefit of dynamic pricing beyond static pricing. Obviously, dynamic pricing is more profitable than static pricing. However, the delay information scheme and information asymmetry affect this benefit. By analyzing symmetric and asymmetric schemes, we analyze whether extra information known by the service provider yields a benefit for him or not. In §6.1 we first introduce our experimental setting. Then in §6.2.1 general observations and in §6.2.2 further observations related with each scenarios are provided.

6.1 Experimental Setting

We assume λ={0.1,0.2,0.3,0.4,0.5,0.6,0.7}. For all models μ is assumed to be 0.75. For the unit waiting cost three cases are analyzed, h={0,0.5,1}. Whenever a customer arrives to the system s/he is announced a waiting time w_1, variance of the waiting time, w_2, and quoted a price, p. We assume the service provider has various levels of price flexibility. Under low price flexibility, p={1,10}; under moderate flexibility, p={1,3,5,7,10}; and under high price flexibility, p={1,2,3,4,5,6,7,8,9,10}. In the sequel, price flexibility is abbreviated with pf, with levels 1,2 and 3, where pf=1 denotes the lowest flexibility level. The utility function of a customer was defined in Equation (1) as follows:

$$U(p,,w_1,w_2) := r - \Theta_1 c_1(w_1) - \Theta_2 c_2(w_2) - p \qquad (1)$$

In the equation r is assumed to be the same for all customers and set equal to 10. Θ_1 is assumed to be a random variable which is uniformly distributed between 1 and 2. We assume customers are differently sensitive to the precision of delay. While some might give less importance to the quoted variance of the waiting time, for some, the variance of the waiting time might have more importance. We assume customers have three levels of sensitivity to the precision of the waiting time. In order to provide these levels we assume Θ_2={0.5, 1, 2} corresponding to robust, moderately sensitive and highly sensitive customers. In the equation, $c_1(w_1)=1+w_1$, and $c_2(w_2)=w_2$. According to the defined parameters, the probability of customers to join the system is obtained as:

$$H(p,w_1,w_2)=\max(0;\min(1;(\frac{r-p-\theta_2 w_2}{1+w_1}-1)))$$

The customer sensitivity is abbreviated with cs, with levels 1,2 and 3, where cs=1 denotes the lowest sensitivity level.

In total, there are 7 distinct scenarios to analyze; SPNIs, SPPIs, SPFIs, DPPIs, DPFIs, DPNIa, DPPIa, (see sections 4 and 5) each requiring 126 instances. Due to computational difficulty we were not able to obtain the results for DPPIa, which leaves us with 6 distinct scenarios. For the asymmetric setting we base our conclusions on comparison of NIa and FIa. For obtaining the optimal profit per unit time and the optimal policy, under static pricing an exhaustive search is made. Under dynamic pricing, the corresponding stochastic dynamic program is modeled as a linear program and GAMSv24.2.2 modeling language with CPLEX solver is used. The solution times were negligible.

Notice that all the observations and results obtained in this chapter are sensitive to the experimental setting. The generality of the results might be investigated through different experimental settings.

6.2 Observations and Results

We present the results in two parts. First, general observations are made and no discussion is provided since the intuition behind is relatively straightforward. Then specific observations along with discussions are presented.

6.2.1 General Observations

- Under all scenarios, as the price flexibility increases, the profit either increases or remains the same.
- By Proposition 1 we know that SPNIs and SPNIa models are equivalent and by definition SPFIs and SPFIa models are equivalent. For this reason, under static pricing scheme, all observations that are made for symmetric scheme (in§6.2.2) are also valid for the asymmetric setting.
- For any delay information sharing scheme, profit under DP is at least as high as the profit under SP. Under both symmetric and asymmetric settings, the gap between static and dynamic pricing increases as price flexibility, arrival rate and the amount of information increases.

6.2.2 Further Observations

Observation 1. *Under static pricing and symmetric scheme when there is no waiting cost, for no information (NI) and partial information (PI) schemes, as customer sensitivity increases the profit decreases under all pricing flexibilities.*

Discussion. Under no information, the service provider announces each arriving customer the same steady-state expected waiting time and the variance of the waiting time. While, less sensitive customers give less importance to the announced variance of waiting time and more likely to join the system, more sensitive customers give more importance to the variance of the waiting time and are more likely to interpret the announced variance as high. Therefore, the probability of these customers to join the system is lower than less sensitive customers. That is why, as customer sensitivity increases, the probability of customers to join the system decreases and thus the profits decreases. Under partial information the provider announces the variance in terms of the number of customers in the system. In a similar fashion, if we compare robust and highly sensitive customers arriving to the same system state, more sensitive customers are less likely to enter the system and therefore the provider's profit decreases as the customer sensitivity increases. Under full information, the service provider announces the exact waiting time with perfect precision and therefore customer sensitivity loses its effect on the profits.

Observation 2. *Under static pricing and symmetric scheme, when there is no waiting cost:*

- ***a.*** *Under low price flexibility, and moderate λ, and under all customer sensitivities, NI always yields the highest profit.*
- ***b.*** *Under moderate price flexibility, and high λ and customer sensitivity, FI outperforms the other schemes. Under moderate λ and low and moderate customer sensitivity, PI may outperform other schemes.*
- ***c.*** *Under high price flexibility, under all λ's and customer sensitivities, FI outperforms the other schemes (except when λ=0.1, cs=1)*

Discussion. Under low price flexibility the service provider is not able to control demand effectively and no information could be the preferred scheme. For low λ, there is no significant difference among the schemes. Since the arrival probability is low, the system is often empty and the provider is able to announce short waiting time under all schemes. Under moderate λ and low price flexibility the provider might choose to use precision of information as a tool to control the demand and might prefer to lower the information accuracy.

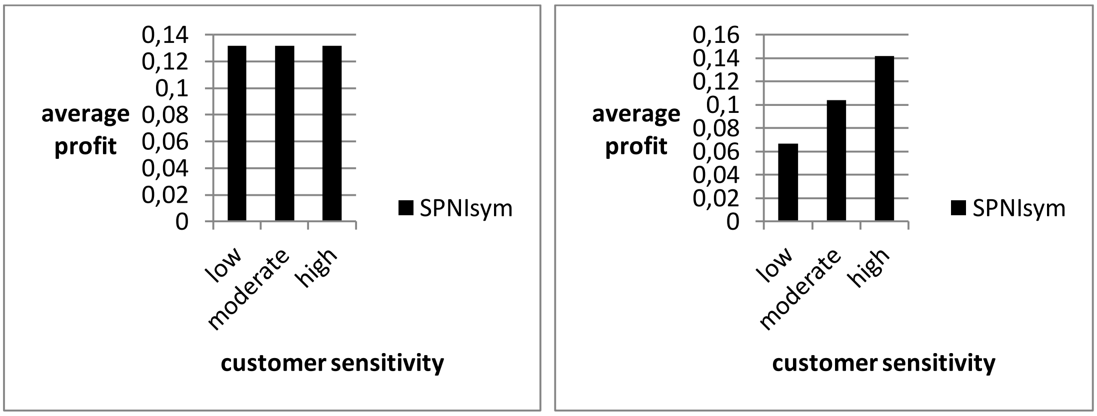
From Observation 1, we know that as the customer sensitivity increases the profit under SPPIs decreases and profit under SPFIs does not change. Therefore, in the existence of low or moderately sensitive customers PIs might perform better than FIs. We know that as λ, price flexibility and customer sensitivity increases FI outperforms other schemes. Therefore, under moderate price flexibility, the increase in λ results in the favor of FI. Under the combination of moderate λ and low and moderate customer sensitivities, however, FI may perform worse than PI. Under these settings, the provider uses delay accuracy as a tool to control the demand. Under NI, accuracy is very low this inaccuracy decreases the probability of customers to join the system, under FI because the provider gives perfect information, announcing high waiting times decrease the joining probability. PI provides equilibrium in terms of accuracy and expected waiting time and therefore we observe in some cases that PI outperforms FI

Observation 3. *Under static pricing and symmetric scheme when waiting cost is not low, under low price flexibility:*

- ***a.*** *For low or moderate λ, profits under SPNI are not affected by customer sensitivity.*
- ***b.*** *For high λ, under both NI and PI the profit increases with customer sensitivity.*

Discussion a. Figure 12a shows the average profit under SPNIs with respect to customer sensitivity levels under low price flexibility, under λ=0.2 and averaged over h=0.5 and h=1. As shown in the figure, profit is not affected by customer sensitivity. Under NI**,** when price flexibility is so low accept all and reject all are the only possible policies and when λ is relatively low, the provider is able to announce short steady-state waiting time and low variance of the waiting time. The announced waiting time and the variance of the waiting time are interpreted as relatively low even by the most sensitive customers and customers are likely to join the system. Effective arrival rate is not affected significantly. Therefore, the sensitivity of customers becomes unimportant in terms of the profit.

Discussion b. Figure 12b shows the average profit with respect to customer sensitivities under SPNIs when λ=0.6, price flexibility is low and averaged over h=0.5 and h=1. As shown from the figure, as customer sensitivity increases to the precision of the waiting time, profits under Nis increases. For high λ, when price flexibility is low; the service provider has not flexibility to trim the arrival rate but just has chance to accept or reject the customers. Although a decrease in customer sensitivity increases the probability of joining the system, this may not be something desirable as there is waiting cost. Under high waiting cost, uncontrolled arrival rate may result in negative profit. As a result an increase in customer sensitivity may increase the profit. When waiting cost is low, or when price flexibility is high, the provider has more flexibility to control the joining rates to the system and less sensitive customers are more desirable for the provider.



(a) λ=0.2 (b) λ=0.6

Figure 12: Average profit with respect to customer sensitivities under low price flexibility and averaged over h=0.5 and h=1

Low price flexibility together with unresponsiveness to customer sensitivity makes FI vulnerable to increase in λ. Thus in the presence of waiting cost, under high customer sensitivity and high λ, PI outperforms FI and NI. Figure 13 shows average profit under SPs schemes with respect to λ's under low price flexibility and high customer sensitivity, averaged over h=0.5 and h=1. As shown form the figure, as λ increases profits under SPFIs scheme decreases and SPPIs scheme outperforms SPFIs.

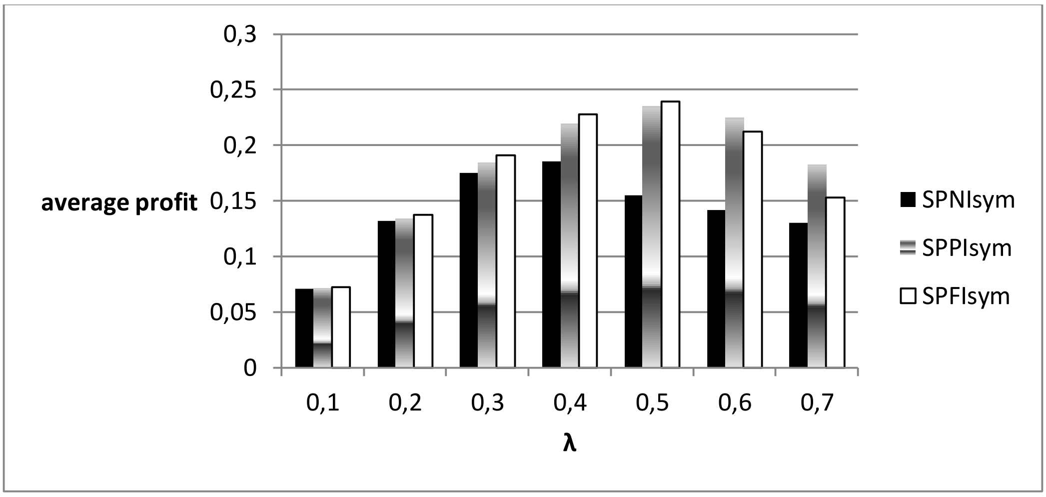


Figure 13: Average profit of SPsym schemes with respect to λ's under pf=1, cs=3 and averaged over h=0.5 and h=1

Observation 4. *Under static pricing and symmetric scheme, when waiting cost is not low, FI always outperforms less informative schemes under high price flexibility.*

Discussion. We observe that service provider may not benefit price flexibility under no or partial information, whereas it benefits the flexibility under full information. The reason is that, the impact of a change in price is much higher under less informative schemes, thus even if pricing flexibility has increased less informative schemes are more conservative in making a change in price to increase the profit.

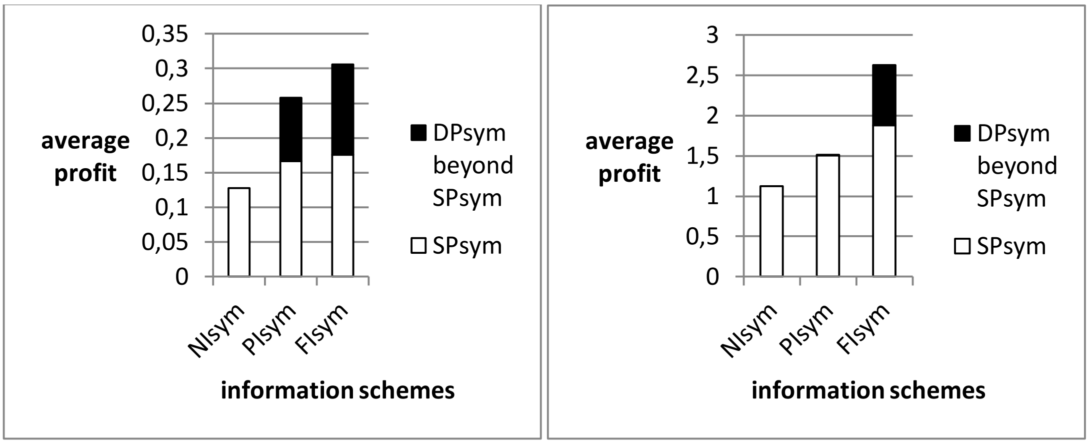
Under FI, the effective arrival rate can be more flexibly determined under a flexible price scheme. As the price flexibility increases, the service provider has increased control over the customer arrival. By quoting a high price s/he may make the arriving customers balk the system and may keep the system relatively "empty" and may always announce arriving customers short waiting times under the more informative scheme.

We observe that the gap between FI and other schemes increases with λ and customer sensitivity and price flexibility, in the favor of FI.

Observation 5. *Under symmetric scheme, as information sharing increases, the benefit of dynamic pricing beyond static pricing also increases.*

Discussion. As the amount of information revealed to the customers increases, the benefit of adopting dynamic pricing is more pronounced. From definition, we know that SPNIs and DPNIs models are equivalent and therefore as expected there is no profit difference between DPNIs and SPNIs. The striking result is, as shown in Figure 14, as price flexibility increases under partial information availability, adopting dynamic policy rather than static pricing does not make much difference in terms of profit. Figure 14a shows the average profits for SPs schemes and also average profit differences between DPs and SPs schemes under low price flexibility. As shown, under low price flexibility for both PIs and FIs, dynamic pricing yields significantly higher profits compared to static pricing. Figure 14b, shows the average profits under static pricing and the difference between dynamic pricing and static pricing under high price flexibility. As shown, for PIs there is no significant difference between dynamic and static pricing.

We conclude that even if service provider has moderate information availability on system status, quoting different prices to arriving customers does not bring any benefit. In other words, unless the system status is perfectly observable, quoting the same menu price to all customers is preferable.



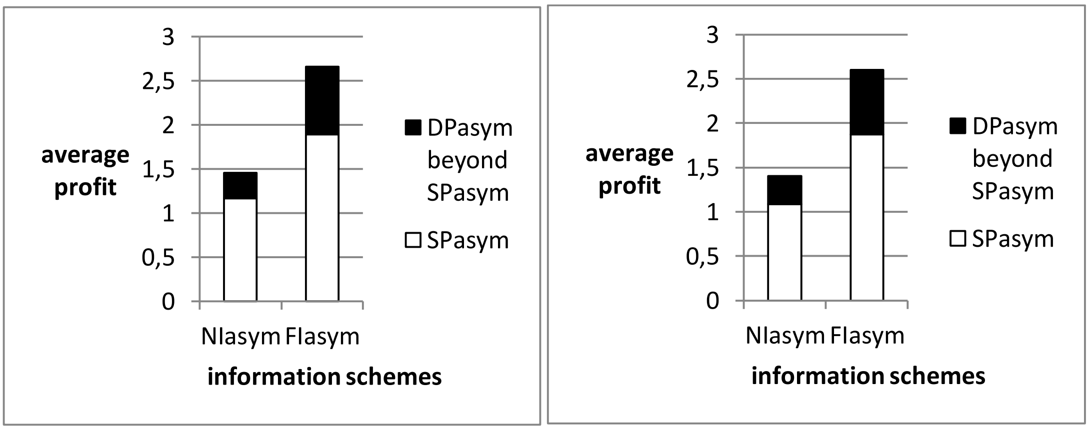
(a) pf=1 (b) pf=3

Figure 14: Average profit under DPsym and SPsym under, averaged over all cs, λ’s and h=0.5 and h=1.

Observation 6. *Under asymmetric scheme, the benefit of dynamic pricing beyond static pricing increases with:*

(i) the amount of information revealed to the customers,

(ii) the waiting cost for NI.



(a) h=0.5 (b) h=1

Figure 15: Average profit under DPasym and SPasym underpf=3, averaged over all cs, and λ’s.

Discussion. From Figure 15 (a) and (b) profit difference between dynamic pricing and static pricing might be observed for NIa and FIa schemes, under h=0.5 and h=1, respectively. As for symmetric scheme, also for asymmetric scheme the benefit of dynamic pricing is mostly observed by full information. Independently from the waiting cost, if the provider chooses to reveal all information he has, it is better for him to adopt dynamic pricing. However, we observe that under FI, as holding increases, the benefit of dynamic pricing decreases when price flexibility is low. The reason is when price flexibility is low increase in the waiting cost affects the profits significantly, however, under high price flexibility profits are nor significantly affected. For NIa also choosing dynamic pricing is better but if waiting cost is low, the benefit of dynamic pricing is not significant. As waiting cost increases, the benefit of dynamic pricing increases.

CHAPTER 7

CONCLUSION

In this thesis, we study three different schemes of delay information which we call no information, partial information and full information. For the pricing problem, we study two schemes; static pricing and dynamic pricing. For the service provider we assume either he has the same level of information with the customers and has either no information, or partial information or full information (symmetric information) or has full information and chooses the amount of information to be revealed with the customers (asymmetric information). We model the system as a stochastic discrete-time Markovian queue, and model pricing problems through Markov decision process. We compare several delay information and pricing schemes and through analytical results and numerical study identify under which conditions which schemes are more preferable. We consider the service provider has varying levels of price flexibility. Furthermore, we assume customers are either robust, or moderately or highly sensitive to the precision of the delay. We observe that customer sensitivities might significantly affect the provider's profit and the provider should consider customer sensitivity levels while deciding delay information and pricing schemes.

For the symmetric scheme our findings are:

- If the provider adopts static pricing, and there is no waiting cost, under low price flexibility and moderate λ, NI always yields the highest profit independently from customer sensitivity. Under moderate flexibility and high λ and high customer sensitivity PI may outperform the other schemes and under high price flexibility for all λ's and customer sensitivities, FI outperforms the other schemes. For very low λ, while the most informative scheme (FI) yields the highest profit, the least informative scheme (NI) yields a higher profit than partial information scheme (PI).

- If the service provider adopts static pricing and if there is no waiting cost, under no information (NI) and partial information (PI) except for very low λ (λ=0.1), profits decreases as customer sensitivity increases. For very low λ, customer sensitivity does not significantly affect the profits. For all sensitivity levels the same profits are obtained.

- If the service provider adopts static pricing and if the waiting cost is not low, under low price flexibility and under low and moderate λ, profits are not affected by the customer sensitivity under NI. Under high λ, under both NI and PI the profit increases with customer sensitivity.

- Dynamic pricing always yields a higher or the same profit compared to static pricing. The benefit of dynamic pricing increases as the information sharing increases.

- Under DPPIs scheme, the provider quotes lower prices as the number of customers in the system increases. In order to keep the system relatively empty, after some state he chooses to quote the highest possible price and reject the arriving customers. By such a pricing quotation policy, he is always able to announce short waiting time and low variance of the waiting time. Under DPFIs schemes we are not able to observe a monotonic behavior for the price quotation.

For the asymmetric scheme our findings are:

- All observations done under static pricing and symmetric scheme are also valid for static pricing and asymmetric scheme. Under no and partial information schemes, additional information to the provider does not have extra benefit. Under full information, by definition, symmetric and asymmetric schemes become equivalent.

- Under asymmetric scheme, the benefit of dynamic pricing beyond static pricing increases with the amount of information revealed to the customers and with the waiting cost under NI.

- Under DPNIa, as the arriving workload increases, the provider initially chooses to quote higher prices but then lowers. If the amount of arriving workload is too high, however, the provider chooses to reject customers.

Our observations show that the preferences for delay information scheme and pricing schemes might be different depending on the market conditions (arrival probabilities), sensitivity level of the customers, and conditions of the service provider; the waiting cost and price flexibility of the provider.

Our study can be extended in several directions. The observations done in this thesis are sensitive to modeling assumptions. Changing the parameter settings might result in different observations. For all the models, we assume that the sensitivity indicating parameters; Θ_1 is uniformly distributed between 1 and 2 and Θ_2 is a constant value. Furthermore, we set $c_1(w_1)$ and $c_2(w_2)$ functions as linear functions of w_1 and w_2, respectively. We might make our observations for different distributions of Θ_1 and Θ_2 and for nonlinear $c_1(w_1)$ and $c_2(w_2)$ functions which implies different structure for utility function.

For all the models we assume that the capacity is infinite and there is a single server. We may model the scenarios with a certain capacity and/or increase the number of servers. We assume customers just decide whether to enter the system or not. We might model the systems in the existence of strategic customers who also aim maximize their own utility and act in this manner. For all the models, we only consider the provider's profit. Like in the paper of Hassin (2007) we might also take the social welfare into account.

Finally, for all the models we assume the information provided by the provider is considered to be truthful by the customers. We might consider the case, which the provider does not always provide truthful information.

REFERENCES

Afeche, P., & Pavlin, M. (2016). Optimal Price-Lead Time Menus for Queues with Customer Choice: Priorities, Pooling & Strategic Delay. *SSRN Electronic Journal,* 1-25.

Allon, G., Bassamboo, A., & Gurvich, I. (2011a). “We Will Be Right with You”: Managing Customer Expectations with Vague Promises and Cheap Talk. *Operations Research*, 59(6), 1382-1394.

Allon, G., & Bassamboo, A. (2011b). The Impact of Delaying the Delay Announcements. *Operations Research*, 59(5), 1198-1210.

Ata B. & Olsen, T.L. (2009). Near-Optimal Lead-Time Quotation and Scheduling Under Convex-Concave Customer Delay Costs. *Operations Research*, 57(3), 753-768.

Bendre, A. B., & Nielsen, L. R. (2013). Inventory Control in a Lost-Sales Setting with Information about Supply Lead Times. *International Journal of Production Economic*s, 142(2), 324-331.

Boyaci, T., & Ray, S. (2003). Product Differentiation and Capacity Cost Interaction in Time and Price Sensitive Markets. SSRN *Electronic Journal*, 5, 18-36.

Cachon, G. P., & Feldman, P. (2010). Dynamic versus Static Pricing in the Presence of Strategic Consumers. Typescript.

Cachon, G. P., & Fisher, M. (2000). Supply Chain Inventory Management and the Value of Shared Information. *Management Science*, 46(8), 1032-1048.

Chen, F., & Yu, B. (2005). Quantifying the Value of Lead Time Information in a Single-Location Inventory System. *Manufacturing & Service Operations Management*, 7(2), 144-151.

Choi, H.P., Blocher, J.D., & Gavirneni, S. (2008). The Value of Sharing Production Yield Information in a Serial Supply Chain. *Production and Operations Management* 7, 614-635.

Çelik, S., Maglaras, C. (2008). Dynamic Pricing and Lead-Time Quotation for a Multiclass Make-to-Order Queue. *Management Science*, 54(6), 1132-1146.

Ding, D., Ou, J., & Ang, J. (2014). Analysis of Ticket Queues with Reneging Customers. *Journal of the Operational Research Society*, 66(2), 231-246.

Dobson, G., & Pinker, E. J. (2006). The Value of Sharing Lead Time Information. *IIE Transactions*, 38(3), 171-183.

Feng, Q. (2010). Integrating Dynamic Pricing and Replenishment Decisions under Supply Capacity Uncertainty. *Management Science*, 56(12), 2154-2172.

Gavirneni, S., Kapuscinski, R., & Tayur, S. (1999). Value of Information in Capacitated Supply Chains. *Management Science,* 45(1), 16-24.

Gayon, J., Talay-Değirmenci, I., Karaesmen, F., & Örmeci, E. L. (2009). Optimal Pricing and Production Policies of a Make-To-Stock System With Fluctuating Demand. *Probability in the Engineering and Informational Sciences,* 23(02), 205.

Guo, P., & Zipkin, P. (2007). Analysis and Comparison of Queues with Different Levels of Delay Information. *Management Science,* 53(6), 962-970.

Guo, P., & Zipkin, P. (2009). The Effects of the Availability of Waiting-Time Information on a Balking Queue. *European Journal of Operational Research*, 198(1), 199-209.

Güllü, R. (1997). A Two-Echelon Allocation Model and the Value of Information under Correlated Forecasts and Demands. *European Journal of Operational Research*, 99(2), 386-400.

Hafizoğlu, A. B., Gel, E. S., & Keskinocak, P. (2016). Price and Lead Time Quotation for Contract and Spot Customers. *Operations Research*, *64*(2), 406-415.

Hassin, R. (1986). Consumer Information in Markets with Random Product Quality: The Case of Queues and Balking. *Econometrica*, 54(5), 1185.

Hassin, R. (2007). Information and Uncertainty in a Queuing System. *Probability in the Engineering and Informational Sciences*, 21(03), 361.

Jouini, O., Akşin, Z., & Dallery, Y. (2011). Call Centers with Delay Information: Models and Insights. *Manufacturing & Service Operations Management,* 13(4), 534-548.

Koenig, M., & Meissner, J. (2010). List Pricing versus Dynamic Pricing: Impact on the Revenue Risk. *European Journal of Operational Research*,204(3), 505-512.

Larsen, C. (1998). Investigating Sensitivity and the Impact of Information on Pricing Decisions in an M/M/1/ ∞ queueing model. *International Journal of Production Economics,* 56-57, 365-377.

Palaka, K., Erlebacher, S., & Kropp, D. H. (1998). Lead-Time Setting, Capacity Utilization, and Pricing Decisions under Lead-Time Dependent Demand. *IIE Transactions*, 30(2), 151-163.

Ray, S., & Jewkes, E. (2004). Customer Lead Time Management When Both Demand and Price are Lead Time Sensitive. *European Journal of Operational Research,* 153(3),

Shone, R., Knight, V. A., & Williams, J. E. (2013). Comparisons between Observable and Unobservable M/M/1 Queues with Respect to Optimal Customer Behavior. *European Journal of Operational Research,* 227(1), 133-141.

Simchi-Levi, D., & Zhao, Y. (2003). The Value of Information Sharing in a Two-Stage Supply Chain with Production Capacity Constraints. *Naval Research Logistics,* 50(8),

So, K. C., & Song, J. (1998). Price, Delivery Time Guarantees and Capacity Selection. *European Journal of Operational Research,* 111(1), 28-49.

So, K. C. (2000). Price and Time Competition for Service Delivery. *Manufacturing & Service Operations Management,* 2(4), 392-409.

Spenner, P., & Freeman, K. (2012, May). To Keep Your Customers, Keep It Simple. Harvard Business Review.

Swamiathan, J. M., Sadeh, N. M., & Smith, S. F. (1995). The Robotics Institute Cariiegie Mellon University Pittsburgh PA. Typescript,

Tian, Y., & Pan, X. (2008). The Value of Downstream Information Sharing on Upstream Supply Chain. *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing.*

Uzun, H.,& Poturak, M. (2014). Factors Affecting Online Shopping Behavior of Consumers. *European Journal of Social and Human Sciences,* 3, 163-170. Retrieved March 20, 2017.

Webster, S. (2002). Dynamic Pricing and Lead-Time Policies for Make-to-Order Systems. *Decision Sciences*, 33(4), 579-600.

Whitt, W. (1999). Improving Service by Informing Customers about Anticipated Delays. *Management Science,* 45(2),

Zhang, G. (2015). Joint Lead Time and Price Quotation : Dynamic or Static? (Master's thesis, Lingnan University) (pp. 1-69). Hong Kong.

Zhao, X., Stecke, K. E., & Prasad, A. (2011). Lead Time and Price Quotation Mode Selection: Uniform or Differentiated? *Production and Operations Management*, 21(1), 177-193.