Studies in the Algorithmic Pricing of Information Goods and Services

Meenal Chhabra

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science & Application

Sanmay Das, Chair
Anil K.S. Vullikanti, Co-Chair
Naren Ramakrishnan
Ilya O. Ryzhov
David Sarne

February 7, 2014
Blacksburg, Virginia

Studies in the Algorithmic Pricing of Information Goods and Services

Meenal Chhabra

(ABSTRACT)

This thesis makes a contribution to the algorithmic pricing literature by proposing and analyzing techniques for automatically pricing digital and information goods in order to maximize profit in different settings. We also consider the effect on social welfare when agents use these pricing algorithms. The digital goods considered in this thesis are electronic commodities that have zero marginal cost and unlimited supply e.g., iTunes apps. On the other hand, an information good is an entity that bridges the knowledge gap about a product between the consumer and the seller when the consumer cannot assess the utility of owning that product accurately e.g., Carfax provides vehicle history and can be used by a potential buyer of a vehicle to get information about the vehicle.

With the emergence of e-commerce, the customers are increasingly price sensitive and search for the best opportunies anywhere. It is almost impossible to manually adjust the prices with rapidly changing demand and competition. Moreover, online shopping platforms also enable sellers to change prices easily and quickly as opposed to updating price labels in brick and mortar stores so they can also experiment with different prices to maximize their revenue. Therefore, e-marketplaces have created a need for designing sophisticated practical algorithms for pricing. This need has evoked interest in algorithmic pricing in the computer science, economics, and operations research communities.

In this thesis, we seek solutions to the following two algorithmic pricing problems:

- In the first problem, a seller launches a new digital good (this good has unlimited supply and zero marginal cost) but is unaware of its demand in a posted-price setting (i.e., the seller quotes a price to a buyer, and the buyer makes a decision depending on her willingness to pay); we look at the question — how should the seller set the prices in order to maximize her infinite horizon discounted revenue? This is a classic problem of learning while earning. We propose a few algorithms for this problem and demonstrate their effectiveness using rigorous empirical tests on both synthetic datasets and real-world datasets from auctions at eBay and Yahoo!, and ratings on jokes from Jester, an online joke recommender system. We also show that under certain conditions the myopic Bayesian strategy is also Bayes-optimal. Moreover, this strategy has finite regret (independent of time) which means that it also learns very fast.

- The second problem is based on search markets: a consumer is searching for a product sequentially (i.e., she examines possible options one by one and on observing them decides whether to buy or not). However, merely observing a good, although partially informative, does not typically provide the potential purchaser with the complete information set necessary to execute her buying decision. This lack of perfect information about the good creates a market for intermediaries (we refer to them as experts) who can conduct research on behalf of the buyer and sell her this information about the good. The consumer can pay these intermediaries to learn more about the good which can help her in making a better decision about whether to buy the good or not. In this case, we study various pricing schemes for these information intermediaries in a search-based environment (e.g., selling a package of k reports instead of selling a single report or offering a subscription based service). We show how subsidies can be an effective tool for a market designer to increase the social welfare. We also model quality level in the experts and study competition dynamics by computing equilibrium strategies for the searcher and two experts with different qualities. Surprisingly, sometimes an improvement in the quality of the higher-quality expert (holding everything constant) can be pareto-improving: not only that expert's profit increase, so does the other expert's profit and the searcher's utility.

Acknowledgments

My efforts towards my Ph.D. would not have been fruitful without the contributions of a few people. Here, I would like to thank and acknowledge them to the best of my abilities.

First and foremost, I would like to express my sincere gratitude towards my advisor Sanmay Das for mentoring and funding me. He introduced me to my research topic at the beginning of my Ph.D. and mentored me very patiently throughout the process. He has been extremely helpful and encouraging. His positive attitude about the results whether positive or negative has been very motivating to me. I cannot imagine a more supportive advisor. I am very thankful to him for guidance.

I am very grateful to all my Ph.D. committee members. I would like to express my sincere gratitude towards Dr. Naren Ramakrishnan and Dr. Anil Vullikanti for guiding and providing helpful suggestions throughout the process. I am very thankful to Dr. David Sarne for giving me the opportunity to work with him. He has guided me throughout and is like a second advisor to me; half of my thesis is work done in collaboration with him. He has been extremely helpful, and I have learnt a lot from him. I am very grateful to Dr. Ilya Ryzhov, from the University of Maryland for his help, ideas, and insights in the dynamic pricing project. I learned a lot about operations research literature from him. I would like to thank all of them for being really prompt in replying to queries and generously providing me with their valuable time and help.

Special thanks to Professor Boleslaw Szymanski and the entire SCNARC team for providing me the opportunity to work with them. I am also thankful to Professor Elliot Anshelevich for teaching various interesting courses and his help and guidance in a research project. I am very grateful to the faculty at RPI and Virginia Tech for teaching various exciting courses.

My lab mates Allen, Mithun and Zhuoshu have been great friends to me and always gave me an ear for all sorts of questions. They have always been very helpful in providing suggestions, discussing ideas, homeworks etc., and also making our lab a lively place. I am extremely thankful to all of them for being supportive. I would like to specially thank Mithun for his efforts in proofreading my preliminary examination report and this thesis. I am also grateful to my floormates at Lally and KW II for all the technical and the non-technical discussions.

Many thanks to the writing center at RPI and VT for helping me with my writing. I would

also like to thank the staff at the CS department at RPI, VT and WUSTL: Chris, Terry, Sharon, Carol, Ginger, Lori, and Lauren for taking caring of all the administrative work.

On a personal note, I am very thankful to all my family, friends and relatives. I am very grateful to my parents, husband, and sisters for their unconditional love, support and strength. I am indebted to my uncle Rajeev for mentoring me throughout my life and always encouraging me to follow my dreams. I am specially thankful to him for making me take my first step in computer science by helping me choose it as my major in undergrad, his long discussion sessions, and also for saving the newspaper strips for interesting articles. I am also very thankful to all my friends in the US and back home for always being there for me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This thesis focuses on algorithmic pricing — the method of automatically setting or updating the price of a commodity depending on factors like demand and competition to achieve a desired goal like maximizing profits, capturing market share. Airline pricing is a notable example of successful algorithmic pricing. With the recent proliferation of online marketplaces and improvements in computing power, the use of algorithmic price-setting agents is on the rise. Many online marketplaces like Amazon.com and Buy.com use intelligent agents to dynamically adjust their prices in response to current demand and competitors' prices. In the past, Buy.com employed software agents to scrape competitors' websites to learn their commodity prices. While this simple algorithm enabled Buy.com to undercut its competitors and attract more price-sensitive buyers, it also resulted in a decrease in its profit. This illustrates that a naive approach to algorithmic pricing may not yield desired results [85].

Algorithmic pricing problems have traditionally been studied in economics or operations research. However, a common practice in these fields is to simplify a problem until analytical results are obtained. This approach does help in building intuition about the problem but the inherent simplifications sometimes render the solutions unusable in the real world. For example, in order to make the problem of maximizing revenue when the demand is unknown, more tractable, Harrison et al. [50] assume that there are only two possible demand functions, and hence reduce the problem to one of identifying which of these two functions captures the true demand.

The need for designing efficient algorithms in order to increase revenue has recently fostered interest in algorithmic pricing in the computer science research community. In fact, the Association of Computing Machinery (ACM) organizes an annual conference named "Economics and Computation" (that also lists "Revenue optimization, pricing, and payments" as one of the topics of interest in its call for papers) and publishes a journal "ACM Trans-

actions on Economics and Computation", both of which are focused on problems at the intersection of computer science and economics. Top conferences in Artificial Intelligence like the international conference on Autonomous agents and Multiagent Systems (AAMAS), the Association for the Advancement of Artificial Intelligence conference on AI (AAAI) etc., are also showing an increasing interest in problems related to revenue maximization.

From a computer science perspective, problems of maximizing revenue in various scenarios fit naturally within a reinforcement learning framework and thus can be solved empirically using dynamic programming techniques. Consider, for example, the revenue maximization problem for pay-per-click ad auctions like Google Adwords. In these auctions the advertisers bid for their ad to get displayed. The winner's ad is shown but she only pays if the ad gets clicked. The auctioneer wants to choose the ad which maximizes her revenue. This revenue depends not only on the advertiser's bid but also on the propensity of the ad to be clicked because the auctioneer does not get paid until the ad is clicked. The auctioneer can learn about the performance of the ad from its history and thus make better decisions while choosing a winner. Another related example is a design of a recommendation engine. Shopping websites like Amazon wants to provide useful recommendations to potential buyers in order to make more sales. Yet another example, which we study in detail in this thesis, is that of maximizing revenue when the underlying demand is unknown. Each of these examples can be modeled as a reinforcement learning problem because of the inherent exploration-exploitation dilemma: these agents have to learn from their interactions with the environment but at the same time have to make decisions to optimize their goal.

This thesis concentrates on the algorithmic pricing problem in specific sequential decision making environments where the utility of an agent making a decision depends on a sequence of actions and not just the current action. There is an interdependence among actions taken at different steps (i.e., the current action affects the subsequent actions). Consider a sequential auction setting, where two complementary goods are to be auctioned one by one. In a truthful seting, the winner of the first round of the auction is more likely to bid higher in the second round than she would have bid had she lost in the first round. In general, the utility of a bidder cannot be calculated by treating the two auctions as separate one-shot decision making problems (episodic) because the first auction determines how much the bidder values the second item [46].

In sequential-decision making, some research efforts focus on designing generic algorithms while others try to provide specific results based on the particular information structure of the problem under consideration. This thesis belongs to the latter category: we study two specific pricing problems which are described below.

1.2   Pricing Digital Goods

The first problem that this thesis deals with is that of pricing a digital good in a posted-price auction setting. A digital good is a good that is sold in a digital format and has zero marginal cost of production, i.e., there is no cost to produce an additional copy of the good. Therefore, such goods can be reproduced in infinite quantities. Some examples of digital goods are iTunes apps/songs/movies and Kindle e-books. In a posted-price auction, the seller quotes the selling price for the item, and on observing it the buyer decides whether to buy or not.

In our model, time proceeds in rounds, and in each round a single buyer arrives with a unit demand. The assumption of having a unit demand is commonly used in the literature owing to the non-perishable nature of digital goods [58]. Each buyer is characterized by a privately known valuation of the good which determines the maximum amount they are willing to pay for this good. The buyers are myopic in nature in the sense that they buy the item if the posted price is less than their valuation in the current round without considering price fluctuations in future rounds. The seller can update her price at the beginning of each round. The goal of the seller is to maximize her infinite-horizon discounted revenue for a newly launched digital good whose demand is unknown.

The problem of maximizing revenue in the infinite supply case is very different from that of finite supply auctions for the following reasons. Firstly, finite supply goods often have associated marginal and inventory costs. Secondly, in the unlimited supply case, the problem depends only on the information state of the seller and is stationary with respect to time whereas in the limited supply case, time becomes another dimension in determining the state of the seller. Therefore, the solution proposed for one is not directly reducible to that for the other.

If the seller is aware of how much each buyer is willing to pay for the good then the profit maximizing approach is to charge exactly (or slightly below) what each buyer is ready to pay. However, it is hard to estimate the valuation for each buyer. Perhaps more importantly, this strategy of charging different buyers differently, called price discrimination, often causes anger and loss of trust among the consumers and introduces complications as the consumers try to find ways to manipulate the seller. So it is seldom used in practice, e.g., when Amazon experimented with this approach for selling DVDs in 2000, it faced criticism from users [1]. For this reason, price discrimination strategy is not considered further.

In this thesis, instead of learning every individual's valuation, the seller assumes that all valuations are independent and identically distributed (i.i.d.) draws from some distribution and hence tries to set a single profit maximizing price. Of course, the problem is trivial if the seller knows the distribution function completely; however, we assume that the valuation distribution has a known form but unknown parameters. The seller tries to learn these

---

[1]see `http://news.cnet.com/2100-1017-245326.html`, last accessed on January, 2014

parameters from buyers' decisions while maximizing the profit at the same time.

This problem of simultaneous learning and optimizing exhibits a classic exploration-exploitation dilemma: the seller faces the challenge of striking a balance between acquiring more informa-tion to make better decisions in later rounds and exploiting currently available information in order to maximize immediate returns. For example, in this case of pricing a digital good, the seller may exploit the available information and keep charging a price that has yielded the highest revenue of all the explored prices so far but is in fact not optimal. However, on exploring other prices, she will get more information and may find a truly optimal price.

Problems exhibiting an exploration-exploitation dilemma are often modeled in the form of multi-armed bandits. A multi-armed bandit is a generalization of a slot machine commonly seen in a casino. In a slot machine, the player pulls the arm and gets a reward associated with the arm; similarly, in a multi-armed bandit, the player has the option of choosing an arm (option) from a set of multiple arms, and the reward is given depending on the selected arm. Assuming an agent is rational and wants to maximize her reward, she will choose an option that gives her the best utility/reward. However, if she is unaware of these rewards but wants to maximize her utility over a series of trials rather than just one trial then she can learn the rewards distribution while trying to optimize her utility.

Consider the digital good pricing problem: if the seller has a set of prices to choose from then these prices can be thought of as arms of a multi-armed bandit, the seller wants to find a price which will maximize her revenue over a series of trials which is analogous to selecting a reward maximizing arm. It is often assumed that all the prices are independent of each other when reducing the pricing problem to the multi-armed bandit [58]. This is not a suitable assumption because if a rational utility-maximizing buyer decided to buy at some price q, then she would also buy at any price less than q. Despite that, several of these schemes have been shown to possess good properties in terms of asymptotic regret for the seller's revenue maximization problem in the unlimited supply setting. Blum et al. [15] discuss the application of Auer et al.'s [2] EXP3 algorithm for the adversarial multi-armed bandit problem to posted price mechanisms, showing a worst-case adversarial bound. Kleinberg and Leighton [58] show that the regret bounds for Auer et al.'s [3] UCB1 bandit algorithm for independent and identically distributed (i.i.d.) settings in the posted price context is logarithmic in time. UCB1 is intended to minimize regret even in finite-horizon contexts, so we would expect it to perform relatively well. However, these algorithms rarely perform very well in terms of utility received in even simulated posted price auction settings – for example, in Conitzer and Garera's comparison of EXP3 with gradient ascent and Bayesian methods [27], or even in different applications, as found by Vermorel and Mohri on an artificially generated dataset and a networking dataset [89]. There are also some technical limitations on using multi-armed bandits for the pricing problem — for example, bandit algorithms have traditionally been studied in a framework where the arms are statistically independent of each other. However, possible prices of a commodity are interdependent, and utilizing this information can lead to potentially higher profits. Another shortcoming of the bandit approach is that the seller is restricted to a discrete set of prices from which it must pick

the optimal price. One needs information about the discrete set of allowable prices which may not be available. If one were to know the bounds of the region in which the best price lies then she could discretize the region. The problem with manually discretizing is that if the discretization is done finely then there will be a larger number of arms which will slow down the learning. On the other hand, if it is coarse then due to loss in precision one may not have the best price selected in the set. In this thesis, we aim to design algorithms which perform well in practice for digital goods pricing by overcoming these limitations.

Another challenge the seller faces in the pricing problem is that each time a buyer makes a decision, she only informs the seller whether she decided to buy or not; she does not reveal her true valuation. If the buyer had disclosed her true valuation, then the seller could have learned the true distribution easily. However, in this case the seller has to learn the valuation from censored information: when the buyer decides to buy (resp. not buy), the seller only knows that the price she set was less (resp. more) than the buyer's valuation.

We approach the problem of digital goods pricing in steps. In Chapter 3, we propose a stylized model of the posted-price auction setting. We partially relax the problem of learning from completely censored information and assume that the buyer discloses her true valuation in case of a purchase but not when she decides not to buy the good. This is not a very strict assumption for the sale of digital goods because they are non-perishable, therefore a buyer who has already purchased the good is not likely to come back to buy the same item again and is not affected by future price fluctuations. The seller has to learn from partially censored information in this case.

We suggest a learning-based approach where the seller assumes that the true valuations are i.i.d. draws from a uniform distribution with an unknown parameter. The mild relaxation of disclosing the true valuation in case of a buy allows us to prove a surprising result that the greedy (myopic) strategy is Bayes optimal under certain conditions. Further, we can show a finite time-independent regret bound (i.e., asymptotic growth rates on the opportunity cost incurred by using a suboptimal strategy) for the partially censored information case while the best regret bounds are at least logarithmic in time for the completely censored case.

To check the robustness of our proposed strategy, we compare its performance with other bandit-based schemes on three real-world datasets (containing bidding data from eBay and Yahoo! auctions and ratings for jokes from Jester, an online joke recommender system[2]) where our modeling assumptions are violated [55, 45, 1]. Surprisingly, we find that our approach outperforms not only the other partial information-based pricing scheme but also a Gittins index-based pricing technique in the complete information case (i.e., when the seller is aware of the true valuation both in case of a buy and no buy).

In Chapter 4, we take a step closer to reality, and assume that buyers do not disclose their valuations in either case. This implies that the seller gets completely censored signals, i.e.,

---

[2]Jester can be found at `http://eigentaste.berkeley.edu/user/index.php`. The data can be downloaded from `http://eigentaste.berkeley.edu/dataset/jester_dataset_1_1.zip`

after interacting with a buyer, she just knows that the price she quoted is either less (in case of a buy) or more (in case of no buy) than the buyer's valuation of the good. The seller still assumes that the valuations are drawn from a uniform distribution. We extend our partial-information pricing set-up using the knowledge gradient strategy (a Bayesian one-step-look-ahead strategy[83]) to this completely censored case. We also suggest another approximately optimal and scalable approach to pricing a good, and evaluate the performance of various strategies empirically. Our proposed algorithms are very robust and outperform state-of-the-art pricing strategies not only in the environment where our assumptions hold but also in many cases where they are violated (like real-world datasets).

1.3   Pricing Information Services in Search Markets

The latter part of the thesis focuses on pricing information goods/services in sequential search environments. A buyer does not always have access to the same information as the seller about the commodity she wishes to purchase. An information good or service provider sells this missing information which may bridge the information gap between the buyer and the seller. Some examples of information goods are reviews provided by sites like Yelp and Angie's list about commercial businesses, and vehicle history reports by agencies like Carfax and Autocheck.

What is search? In the Artificial Intelligence (AI) community, the term "search" is used almost exclusively in the context of intelligently searching for the optimal solution to a problem. However, in this thesis, we use the term "search" in a sense that is more commonly encountered in the economics literature. This is best explained with the help of the following examples: a consumer searching for a product, a worker looking for a job, a manager looking to hire an assistant, optimal stopping problems etc.

Consider an example of optimal stopping where a manager is looking to hire an assistant (commonly known as the secretary problem). There are n possible candidates who can be ranked according to the order of the manager's preference. The possible candidates are interviewed in a random order one by one, and the manager has to decide immediately after interviewing a candidate whether she wants to hire the interviewee or not. In other words, the problem is to decide when the manager should stop searching. The manager gets a positive utility only if she hires the best candidate, otherwise she gets zero utility. The optimal strategy in the face of uncertainty is to maximize the chance of hiring the best secretary. Surprisingly, this has a very non-intuitive optimal solution: the manager should reject the first $\frac{n}{e}$ candidates, then choose the first candidate who is better than all the candidates interviewed so far [43].

Search problems are related to the secretary problem described above. Consider a slight variation in the above problem. The manager first screens the résumés of interested candi-

dates, then invites selected candidates for an interview one at a time, and reimburses their travel expenses. Each candidate provides a certain utility to the manager on being hired, and the goal of the manager is to maximize her overall utility. Since there is a cost involved in interviewing each candidate (reimbursing travel costs, time consumed in waiting for the next candidate etc.), the manager may not want to interview all the candidates. Therefore, choosing the best candidate may not always yield maximum overall utility owing to the search cost accumulated on the way. When the distribution of the candidates' utility values is known, the optimal strategy is a reservation value-based strategy, i.e., the manager should hire the first candidate with utility greater than a fixed threshold (the reservation value). This threshold is equal to the expected utility the manager can get by rejecting the current candidate and resuming the search.

Consider another example: a consumer is looking to buy a used car. She typically examines cars one at a time. On inspecting a car, she forms some idea about its quality and thus the value of buying that car. However, she may not observe the true quality of the car under investigation; the car may look great but may not have good maintenance history, or may have an accident or salvage title in the past which reduces its value. Thus, what she observes is a noisy signal about the true value of the opportunity. After investigation, she has to decide immediately without waiting for future options to materialize whether to buy the car or continue searching for a better deal. The optimal strategy in this case is also a reservation value-based strategy under certain conditions which will be discussed later in the thesis.

Lack of perfect information creates a market for intermediaries who can disambiguate these observed noisy signals for a payment. Due to the presence of these information-intermediaries (henceforth called "experts"), a consumer has an additional choice, apart from accepting or rejecting the opportunity, she can pay a fee and query these information-intermediaries to know the true value of the opportunity before making a decision. Carfax and Autocheck are two examples of information-intermediaries for the problem of buying a used car. Angie's list is another example of such information provider for local businesses. An interesting question that comes to mind is how these information-intermediaries should price their services when the consumers are strategic and optimize their utility. This is a game-theoretic set up where both the expert and the searcher are strategic agents, and the utility of either depends not only on their decision but also on the decision of the other agent. The searcher's strategy affects how many times the expert will be queried, and the expert's price is deducted from the utility of the searcher everytime she queries the expert.

Chapter 5 concentrates on how an expert should set the price per query optimally in a strategic environment. The expert sets a price per query, and the searcher responds by finding an optimal search strategy for the given price of the query, thus impacting the revenue of the expert. This problem can be formulated as a Stackelberg leader-follower game where the expert is a leader and makes a first move by setting the price, and the searcher is a follower who takes an action that optimizes her utility given the price of the query [40].

Services like Carfax not only sell one query at a time, but also offer a set of k queries together for a price which is less the total price of k separate queries. Particularly, for Carfax, one query costs \$39.99 but a package of 5 queries costs only \$44.99. Such non-linear pricing is also commonly observed in the sale of non-information goods (e.g., in supermarkets). A special case of this non-linear pricing is subscription-based pricing where the service provider allows unlimited usage ($k \to \infty$) for a fixed fee. For example, Carfax offers an unlimited monthly query package for \$54.99. This subscription-based pricing is usually offered for digital goods (e.g., Netflix, Rdio, etc.). Many questions arise in this case such as: how do these service providers decide how much to charge and what package size to offer? Why did Carfax choose only 3 options - a la carte, a package of 5 queries, and an unlimited subscription? In Chapter 6, we seek answers to some of these questions. We find the optimal package size and price combination when the expert has the option to choose between the following different pricing schemes apart from a la carte to maximize her expected profit:

1. Subscription pricing: The expert charges a fixed price for an unlimited queries. The fixed price is chosen optimally.

2. Package pricing: In this case, the expert can choose to sell a package of multiple queries with the size and the price of the package being chosen optimally.

We explore the conditions under which each of these forms of pricing performs best (a la carte, subscription-based pricing, package pricing).

To derive the optimal strategies for the expert and the searcher, we assume that the opportunities observed by the searcher appear exogenously. Now, we would like to study a platform which connects the searchers with these opportunities and thus with the expert. For the used car case, `autotrader.com` is an example of such a platform which connects the searchers looking to buy used cars to the car sellers. In order to attract more users to its website, `autotrader.com` would like to improve the user experience by improving their utility. It can do so by providing cheaper access to the information from an expert like Carfax or Autocheck. One of the ways of doing so is to subsidize the expert fee by making payments to the expert on behalf of its users. In this manner, the expert's profit is not affected but the consumer's utility will increase owing to cheaper access to information. However, subsidies may lead to less overall social welfare because they result in the overconsumption of resources by consumers due to cheaper cost of the resource. Surprisingly, we find that the subsidies for the expert services in this search environment can in fact increase the overall social welfare. The social welfare is maximized when the expert charges the user a price equal to the marginal cost of the query, i.e., the platform fully internalizes the cost to the expert of producing the extra report. Therefore, for services with no marginal cost, it makes sense for the platform to provide them for free to the consumers. The details will be discussed in Chapter 7.

In Chapters 5, 6 and 7 we assume that experts provide perfect information but this may not be true in reality. For example, not all mechanics may report to Carfax, therefore, the

vehicle history reports by Carfax might be missing some entries. We relax this assumption in Chapter 8 and model experts who are themselves noisy and hence may not accurately disambiguate the searcher's signal. The level of noise in the expert's signal is a direct measure of its quality — less the noise, better the quality and vice versa.

We often see that there are multiple firms providing a similar product/service. For example, both Carfax and Autocheck provide vehicle history reports for used cars. What happens when there are multiple experts in the market, and the searcher can choose the expert that provides her better utility? When the experts have identical quality: (1) if they also have the same marginal costs, then in equilibrium they will all quote prices equal to their marginal cost and make zero profits; (2) if they have different marginal costs, then in equilibrium all experts but the one with the lowest marginal cost will price their services at their respective marginal costs. The expert with the lowest marginal cost will undercut the expert with the second lowest marginal cost and will capture the entire market. However, when the experts are differentiated by quality then the market dynamics in equilibrium are non-trivial to derive. One interesting case is when the two experts have different qualities and the marginal cost of production is proportional to their quality (i.e., the expert with higher quality has to pay higher costs for providing the service). Many questions arise in this case. First, what is the optimal strategy for the searcher? Second, when should the searcher prefer one expert over the other? Third, how should the experts price their services in equilibrium when the quality is set exogenously? Fourth, what are the equilibrium properties when the experts can control not only their price but also their quality? We discuss these issues in detail in Chapter 8.

1.4   Chapter Description

1. Chapter 3: Digital Goods Pricing with Partially Censored Information. This is joint work with Dr. Sanmay Das and Dr. Ilya Ryzhov. Some parts of this chapter are also based on work published in AAMAS 2011 [23].

2. Chapter 4: Digital Goods Pricing with Completely Censored Information. This is joint work with Dr. Sanmay Das and is based on work published in AAMAS 2011 [23]. Some parts of this chapter are also joint work with Dr. Ilya Ryzhov.

3. Chapter 5: Pricing Information Services in One-sided Search. This chapter is based on the work published in AAMAS 2011 [24] and European Journal of Operational Research (EJOR) [25]. This is joint work with Dr. Sanmay Das and Dr. David Sarne.

4. Chapter 6: Non-linear Pricing. This chapter is based on the work presented in AMEC 2011 (non-archival workshop) and a journal paper published in EJOR [25]. This is joint work with Dr. Sanmay Das and Dr. David Sarne.

5. Chapter 7: Market Design - Can Subsidies Increase Social Welfare? This chapter is based on the work published in AAMAS 2011 [24] and EJOR [25]. This is joint work with Dr. Sanmay Das and Dr. David Sarne.

6. Chapter 8: Noisy and Competitive Experts. This chapter is based on work accepted for publication at AAMAS 2014 [26]. This is joint work with Dr. Sanmay Das and Dr. David Sarne.

Chapter 2

Literature Review

With the increasing popularity of online shopping, automated dynamic pricing mechanisms are gaining importance. More and more marketplaces are experimenting with price optimization these days in order to maximize their profit. For example, if a user checks the price for an item at Amazon but decides to come back later to buy, the price is not guaranteed to remain same. With more computing power available it is becoming easier to analyze an item's purchase history and estimate the demand, and thus price the item optimally. The first half of this thesis addresses the problem of learning the underlying demand for digital goods from its purchase history and setting prices in order to maximize revenue.

Another dimension along which pricing decisions are optimized is to choose to sell a package of k items instead of a single item. In most cases, the price for the package is different than the total price for buying k items separately; this is called non-linear pricing. For example, one bottle of mineral water costs \$1 but a crate of 6 bottles costs \$4, and not \$6 as one would expect. An interesting special case of non-linear pricing is subscription-based pricing, which is often used for information services. In subscription pricing, there is a fixed fee to access the information service an unlimited number of times (i.e., $k = \infty$). For example, one can stream unlimited video from Netflix for \$7.99 a month (as of January 2014). If the seller is aware of the demand, then she can experiment with different non-linear pricing strategies, and find the package size (k) which yields the maximum revenue. The second half of the thesis explores these non-linear pricing mechanisms in search markets for information goods. In search markets, a consumer observes opportunities one at a time and after examining each opportunity, she has to decide whether to accept it or not without knowing the future options. On examining the opportunity, she does not observe the true value of the opportunity, instead a noisy signal correlated with the true value. She can get correct information about the opportunity by paying a fee to an information-intermediary/expert. In this setting, the expert (or the seller of an information good) has the freedom to set not only the price but also the number of queries to be packaged together in order to maximize her revenue.

Below, I discuss related work for both of these broad sets of problems.

2.1 Dynamic Pricing of Digital Goods

Consider a take-it-or-leave-it auction setting, where a seller quotes a price for a good to be sold to the buyer, and depending on her valuation, the buyer either decides to buy the good or leave. This is also known as a posted-price auction and is how a variety of items in the real world are sold. In this thesis, we study revenue maximizing pricing strategies for selling a digital good in a posted-price setting when the demand is unknown. Time proceeds in rounds and a single buyer with unit demand arrives in each round. The seller can update the asking price at the beginning of each round. As these digital goods (e.g., iTunes apps) have zero marginal cost and unlimited supply, the seller wants to set a price in each round such that it maximizes her expected infinite-horizon discounted profit. At any time, the price offered by the seller affects not only her current state revenue but also the information she gathers about the true demand which in turns influences her future decisions. Therefore, from the seller's perspective, the most important question is how she should quote the price in order to maximize her long term profit every time a buyer arrives.

There has been a lot of work on multiple variants of dynamic pricing/revenue management in the computer science, economics and operations research literatures. Examples include both finite-horizon and infinite-horizon models, seasonal goods, storage goods, strategic and non-strategic buyers etc. I will now turn to discussing some of the important work in this area.

Our setting conforms to a reinforcement learning scenario, and we encounter an exploration-exploitation dilemma because the response of an individual customer is uncertain, and the seller wants to maximize her profit. Some researchers propose reducing the problem of optimal pricing to that of choosing a best arm in a multi-armed bandit, a problem extensively studied in the reinforcement learning literature [81, 58, 15, 61]. The seller selects a discrete set of possible prices where each price corresponds to an arm of a multi-armed bandit. The probability of success of that arm is equal to the probability of the buyer buying at the corresponding price. The seller can now use any bandit algorithm like UCB1 [58], EXP3 [15], or a Gittins index-based policy [61] to set the price in each round. Some of these algorithms have known regret bounds when applied to thepricing problem. For example, Kleinberg and Leighton [58] derive logarithmic regret bounds when the valuation for each buyer is an independent and identically distributed (i.i.d.) draw from a fixed time-invariant unknown distribution and the bandit algorithm used is UCB1. However, in practice UCB1 does not perform very well [23, 27, 89]. Blum et al. [15] prove regret bounds for the EXP3 algorithm given by Auer et al. [2] which is designed for adversarial situations rather than for the i.i.d. case. As expected, experiments show that EXP3 performs worse than UCB1 in an i.i.d. situation [89, 27].

However, there are a few pitfalls of viewing the pricing problem as a multi-armed bandit problem. First, in the classical framework, the arms of the bandits are assumed to be independent which is not true for the pricing problem. If the buyer decides to buy at any

price q, then she will buy at any price less than q also. Second, the requirement of knowing a discrete possible set of prices is also strict.

Recently, there has been some work on correlated bandit models [71, 83, 84]; however, these models assume a specific dependence structure which cannot fit as is to our problem. For example, Mersereau et al. [71] show that complete learning can be achieved by the myopic policy if the correlations are highly structured, even leading to finite cumulative regret for undiscounted arms.

Dimicco et al. [35] propose two simulation-based approaches to dynamic pricing over a finite time horizon. The first strategy is to adjust prices with the goal to sell all the goods by the last day of the fixed horizon and not before. The second is to update the price by reviewing the sales of the day before. If change made on the previous day increased the revenue, then the price is changed in the same direction, if it produced less revenue then the price is changed in the opposite direction. The first strategy outperforms the second in most cases except in the case of valuation decreasing with time as the first strategy holds the item until the end to sell. Babaioff et al. [4] also study a limited-supply posted-price auction in finite time horizon. He formulates this problem as a multi-armed bandit, which is non-intuitive because most of the optimization algorithms in the multi-armed bandit literature assume discounted infinite horizon rewards. They propose a prior-free mechanism with regret $O((k \log n)^{\frac{2}{3}})$ where k is the number of arms in a bandit and n is the time.

Gallien [41] study the optimal pricing problem when the buyers are also strategic. He proposed a dynamic programming-based mechanism for a seller to maximize the expected discounted revenue in a finite horizon setting, when both types of participants, the seller and the buyers, are risk-neutral. The buyers arrive sequentially with unit demand and their arrival epochs follow a known renewable process. The buyers' valuations are i.i.d. draws from a known distribution. The buyers are strategic in the sense that they can misrepresent their valuation and their arrival time. The proposed mechanism achieves the highest expected discounted revenue while satisfying the incentive compatibility constraint. However, the truthfulness constraint is only on the valuations: i.e., a dominant strategy for the bidder is to disclose her true valuation. The solution does not guarantee that the bidders cannot misrepresent their arrival time, and thus, the bidders in the context of their paper are time-strategic. The author compares the performance of the dynamic programming-based optimal mechanism with fixed posted price and online auction empirically for specific settings and finds that the dynamic programming-based mechanism performs the best. The improvement of dynamic programming-based strategy over fixed pricing is very small, whereas the improvement is significant compared to online auctions. However, the proposed dynamic programming-based mechanism and optimal fixed-price auction are not as robust as online auction — if the parameters (prices) are not chosen optimally then the penalty for using dynamic programming-based method or optimal fixed pricing is much higher than that for the online auction method.

Balcan et al. [7] also focus on a generalized pricing problem in a strategic environment. The

authors reduce the problem of designing incentive compatible mechanisms for revenue maximizing pricing to standard algorithmic problems (non-incentive-compatible) using sample-complexity techniques from machine learning. Some examples of revenue maximizing pricing problems are: auction of digital goods to indistinguishable bidders, attribute auctions, item pricing in combinatorial auctions etc. They show that if there exists an optimal (or β-approximate) algorithm for finding the price in order to maximize revenue then that algorithm can be used to give a $(1 + \epsilon)$ (or $\beta(1 + \epsilon)$) approximation for incentive compatible mechanism design under certain assumptions. However, the downside of this approach is that it cannot be used in an online setting.

Besbes and Zeevi [13] formulate a non-stationary demand model for a problem where the underlying valuation distribution changes at some point; the seller knows a priori both the valuation distributions but not the point where the distribution changes. The authors derive the bounds for the worst-case regret and propose strategies that achieve the lower bound performance.

Bertsimas and Perakis [12] study the problem of learning demand and setting optimal prices for a good when a buyer can purchase multiple units of that good. In this setting, they assume a linear demand model (with additive random noise) with two unknown parameters (the slope and the intercept) which are estimated using the past demand observation through least squares. With that estimation, they employ dynamic programming to calculate the optimal price both in a monopoly and in an oligopoly. Least squares based estimation is one of the most widely used methods to estimate unknown parameters of a model with multi-unit demand; Besbes and Zeevi [14] and Harrison et al. [49] also use a similar model and a least squares estimation technique to show theoretical properties of semi-myopic pricing policies. On the other hand, Lobo and Boyd [64] use a Bayesian approach by setting Gaussian priors on the unknowns to propose an approximate method based on convex optimization to solve the dynamic program. Similarly, Balvers and Cosimano [8] also use a Bayesian approach to estimate the parameters. We, on the other hand, assume that each buyer demands only one unit of a good like much of the work in digital goods pricing.

Harrison et al. [50] consider a version of the revenue maximization problem with only two possible candidate distributions for product demand; over time, the decision-maker accumulates evidence in favor of one hypothesis or the other. The authors show that even in this simple model, the myopic Bayesian policy can lead to incomplete learning, i.e., the policy may converge to a sub-optimal price. They also suggest a modification of the myopic Bayesian strategy such that it has finite regret and performs well in simulations. On the other hand, Farias and Van Roy [39] consider a more general case where the demand is a function of an unknown scalar parameter, which can potentially be equal to any positive value. A Bayesian conjugate prior is then used to model the decision-maker's beliefs about the value of this parameter and the margin for error inherent in the beliefs. Cope [28] formulates the problem assuming a fixed number of allowed prices with a Dirichlet prior on the joint distribution of purchase probabilities in order to take the interdependence among prices into account.

Chapman et al. [22] also consider a take-it-or-leave-it pricing setting; however, they want to sell exactly one good in each auction with finite number of bidders. The auctioneer quotes a price to each bidder one by one. If a bidder's valuation is greater than the price asked then she buys the item and the auction finishes, otherwise the seller moves to the next bidder. If no bidder buys the item, the auctioneer receives zero utility. The authors focus on the incomplete information case where the bidder's valuations are i.i.d. draws from a known distribution with unknown parameters. In this case, the auctioneer maintains some prior belief over the unknown parameters and updates her belief after each auction even though she can use the information from each of the bidders individually to refine the priors. The authors compare the offers calculated using two strategies over two auctions, the first of which is to maximize the expected revenue according to the auctioneer's belief, and the second is to choose a price that maximizes the information gain. For the latter strategy, one might expect that the revenue lost in order to gain more information about the valuation distribution earlier in the process may be compensated for by revenue earned in the later stages due to improved information. Surprisingly, it turns out that this is not true.

Although our model is very different from that of Chapman et al. [22], we also assume that the functional form of the valuation is known to the seller but the parameters are not. The seller maintains a belief distribution over the unknown parameters. Under this assumption, the pricing problem can be easily formulated as a dynamic program, and hence the optimal pricing policy can be characterized using a Bellman equation. The exact solutions can be intractable but practical strategies can be developed either by solving a simplified version of the problem to optimality, or by considering heuristics that can be shown to have desirable regret bounds analytically or competitive performance empirically. A simple example is the myopic strategy which has been shown to perform well in certain settings [50]. Another example of such a heuristic which has shown promising results is the use of the knowledge gradient technique which is a Bayesian one-step-look-ahead strategy [83]. Carvalho and Puterman [21] also show that a one-step look-ahead strategy achieves significantly better finite-time performance than a myopic strategy in an empirical study. We extend this knowledge gradient strategy to our pricing problem. We also propose another scalable approach towards solving the Bellman equation to get close to optimal results and demonstrate its robustness by experimentation. We also consider a simplification of the pricing problem where the buyer provides extra information by disclosing her valuation if she decides to buy an item. In this case, we show not only the existence of a finite regret bound (independent of time) but also the Bayes optimality of the greedy strategy under certain conditions.

2.2   Expert-mediated Sequential Search

In Chapters 5, 6, 7 and 8 of this thesis we study the pricing problem for an information-intermediary in a one-sided sequential search environment. Models of sequential search are applicable to many problems like job, mate, house, or employee search. For example, consider

a consumer looking to buy a used car. She investigates cars one at a time, and makes one of two decisions: she either buys the car under consideration and terminates the search, or keeps looking for cars, till she finds one. We assume that these opportunities appear exogenously. If an opportunity is rejected, then the consumer has to pay a search cost to access another opportunity. Therefore, it may not be optimal for the searcher to query forever to get the best deal as this cost may accumulate and outweigh the advantage of the best deal. Hence, one has to settle for a trade-off between the search cost and the value of the opportunity accepted. However, with the rise in e-marketplaces, search costs have diminished drastically because large amounts of information can now be accessed and sorted quickly. Earlier, a consumer interested in buying a car would have to physically visit the car dealers to learn about the available options, but nowadays, one can easily procure this information online.

Another factor to consider is that often on observing the opportunity, the true value may not be realized. Instead, the consumer observes a noisy signal which is correlated with the opportunity. In this used car example, on observing a car, a consumer may not get information about the drive train, repair and maintainence profile, accident history, salvage title in the past etc., which factor into the true value of the car. So she may want to consult an information-intermediary who can do research on behalf of her and provide better information about the opportunity for a fee. This creates a market for such information-intermediaries who aid the searcher by disambiguating the noisy signal. Agencies like Carfax and Autocheck are real world examples of such experts for the used car case. A part of this thesis is focused on the pricing schemes of these experts.

Now, I will discuss the related work for the problem described above. The relevant literature can be divided into following topics:

1. Sequential search

2. Quality certification and competition dynamics

3. Non-linear pricing

2.2.1 Sequential Search

The searcher's problem is to find an optimal decision-making rule, which fits perfectly within a sequential search framework [67, 63]. Sequential search problems have been studied extensively in the literature. In many cases, the solution to such a problem turns out to be a *reservation value* rule [90, 70]. For example, if the searcher realizes the true value of the opportunity on observing it and is aware of the underlying true value distribution, then the optimal strategy is given by the reservation rule where the reservation value is equal to the expected utility the searcher receives by rejecting the current opportunity.

Some examples of sequential search include job, employee, mate, and property search, and optimal stopping problems. Janetos [54] analyzes and compares different methods of mate

selection where a female is looking to find a mate from a pool of n males. Each male is characterized by a fitness value which determines how preferable he is to other males. A female (searcher agent) may be limited by various factors like time, mobility to find mates or memory to remember them. The author analyzes the following matching mechanisms: random matching, fixed-threshold strategy, best-of-n-males, and a different threshold at each stage. As n (number of males) increases, the strategy with a different threshold at each stage significantly outperforms the fixed-threshold strategy. Of course, the algorithm that achieves the best expected fitness is best-of-n-males. However, there is a considerable cost associated with the use of this algorithm, e.g., memorizing the information and mobility to reach out to the males met previously, which can reduce the overall utility. Also, the marginal increase in the expected fitness of the selected male decreases with increase in n (diminishing returns).

Though most of the work assumes that one observes the true value of the opportunity directly, there has been some work on the noisy information case too. Hey and McKenna [53] model noisy search for goods characterized by two attributes - price and quality. Each good has an n-period lifetime. There is a search cost c in observing a new item; however, after paying the search cost only the price is observed and not the quality. The quality can be observed only after purchasing the item. After observing the quality, the searcher decides whether she wants to keep the item for its lifetime or dispose of it some time in between and start searching again. The utility of the searcher from using the item is a linear combination of the price paid and the quality of the item. The authors derive and analyze the optimal strategy, and find that it is not always reservation-based. Lim et al. [62] consider a sequential search model where the opportunity is characterized by multiple attributes, and the utility from selecting a given opportunity is an additive function of its (initially unknown) attributes, which can be revealed one by one by paying a fee; they show that the optimal strategy in this model is double-threshold based. However, their additive utility structure makes the problem very different from ours.

Wiegmann et al. [91] investigate a multi-attribute mate search problem where the males arrive stochastically and the females have to choose a mate. The quality of the mate is represented by some attributes which are not known to the female. She can either reject the potential mate or pay to know the hidden attributes one by one till all the attributes are visible. After all the attributes are revealed, the female can either accept, or reject the mate and resume the search.

McCardle [68] derives a two-thresholds-based strategy for an organization which is considering adopting a new technology, but is uncertain about how profitable it might be. The organization can acquire some information about the technology through a costly process and update its beliefs about the profitability. The authors show that at each time step it is optimal for the organization to (1) reject the technology if the estimate is less than the lower threshold, (2) accept it if the estimate is greater than the upper threshold, and (3) keep investing to get more information, otherwise.

Monahan [73, 74] studies an optimal stopping problem in the case of a two-state Markov

chain where one is a good state and other is a bad state. The decision maker is not aware of her current state and the choice in front of the decision maker is to accept the current state and get the associated reward, reject the current state, or pay a fee to learn more about the current state. Monahan [74] shows that the optimal strategy is a two-thresholds-based strategy in case of perfect information, where the thresholds are on the probability of being in the good state.

The searcher's problem with noisy signals in the presence of an entity that can provide a more accurate estimate of the opportunity for some fee was dealt with by MacQueen [66] (similar to our a la carte case). He shows that the optimal strategy in this case under certain mild assumptions is a double-reservation-value strategy such that the searcher rejects all the opportunities with signals below the lower threshold, accepts the first signal which is above the upper threshold, and consults the expert by paying some fees if the signal lies in between. On consulting the expert, the true value is revealed and the problem is simplified. We extend the work of MacQueen [66] to the task of calculating the optimal strategy for the searcher under various pricing schemes. His focus is on finding a strategy when the cost of information is fixed, and the problem he solves is more searcher-centric. However, we consider a strategic scenario where we allow the expert to decide not only on the price of the information but also on the pricing scheme: a la carte, package pricing, or subscription-based.

Moscarini and Smith [76] also analyze a two-state world from a slightly different perspective. They consider a situation where a decision-maker has to pick one of two actions, each of which is preferred in a different state. They consider the question of how much costly information it makes sense to acquire before making a decision.

2.2.2   Quality Certification and Competition Dynamics

The expert can be thought of as a third party certifier providing quality ratings for a product. Dranove and Jin [36] summarize the literature on third party certifiers. Most prior work either focuses on the seller's incentives to either disclose complete information or hide her quality, or on the strategic behavior of certifiers [9, 10]. In our model, the incentives for experts to acquire more information are dependent on the costs of such information acquisition, but there is no value to hiding information once acquired. We also study the strategic issues arising from the competition dynamics of experts with different levels of information. The major difference between a search setting such as ours and the literature on ratings (for example, credit ratings) is as follows: a central theme of the ratings world is that organizations (or sellers) pay certifiers in order to get rated (the issuer pays model). The business model is different for our expert, since she provides services to buyers directly. These certifiers are struggling to stay in the market so they may have incentives to provide generous ratings [38]. On the contrary, the expert may have to make a decision on the quality of information to acquire in order to compete, but there are no incentives for misrepresentation. Thus the resulting competition among certifiers creates completely different dynamics than in a search

setting.

Another line of related research is on equilibrium analysis in duopoly when firms produce similar goods with different qualities. The experts, in our setting, compete on price and quality, so we focus on literature geared towards finding Bertrand solutions. Motta studies a two stage game in which the two participating firms first choose their quality and then select price or quantity when the marginal cost varies with quality in order to compare Cournot and Bertrand equilibrium outputs [77]. Crampes and Hollander study how a duopoly equilibrium can be perturbed if there is a minimum quality requirement [30]. Economides studies oligopolistic competition with an infinite number of firms who can select quality from a given range, and he studies equilibrium characteristics with both fixed and variable costs of production [37]. However, in all of these papers, the model used are simple enough to be solved analytically. As far as we are aware of, we are the first to study competition in information provision in search markets.

2.2.3 Non-linear Pricing

Another major related strand of literature is on non-linear pricing. Sundararajan [86] analyzes optimal pricing for information goods (those with low or zero marginal costs of production) with a particular focus on explaining when and why subscription models (fixed-fee pricing) dominate usage-based or a la carte pricing from the seller's perspective. Starting from the empirical observation that unlimited subscription models are usually available in online services, where vanishing marginal costs and low search costs dominate, Sundararajan explains why this would occur, and analyzes the optimal combination of fixed-fee and non-linear usage based pricing as a function of the characteristics of the market. His work places special emphasis on the transaction costs of administering usage as a significant factor when marginal cost of producing an extra unit is low or zero. Our work focuses on sequential search with informational experts, and as a specific case we analyze environments with informational services with low marginal costs (for example, producing an additional report on a car's history). An interesting difference arises when we look at the impact of search costs on the optimality of fixed-fee pricing, because search costs are an intrinsic feature of the environment and directly impact the consumer rather than the expert. Our work is also somewhat related to recent literature on the bundling of information goods [6, 42]. Much of this literature focuses on deriving general conditions under which bundling of complementary or substitute goods makes sense. While packages of expert services are, of course, a different case, we are able to extend the intuitions from the theory of product bundling of [42] to the search domain. For example, in product bundling of information goods, if consumers' values for future goods do not decrease too quickly, bundling can be optimal for the seller. Similarly, in search, the marginal utility of an additional query to the searcher turns out to be an important factor in the expert's decision as to whether to offer fixed-fee or usage-based pricing, and whether to package multiple queries into indivisible groups that must be purchased together (such schemes are common: for example, as of June 2013, Autocheck offers

a single car history report for \$29.99, and an unlimited package for one month for \$44.99).

2.2.4 Contributions

In this thesis, we compute the Stackelberg equilibrium for the searcher and the monopolist expert in a noisy sequential search environment by characterizing the optimal search strategy for the searcher and the profit maximizing strategy for the expert when she can offer to sell her queries in a package of size k (non-linear pricing). Although, there have been investigations in the area of optimal search strategy for the searcher in the presence of information brokers, the expert has been considered exogenous. To the best of our knowledge no one has considered the information-brokers to be endogenous to the sequential-search environment and modeled them as strategic agents. We also analyze how a platform, which connects the searcher and the expert, can improve the additive social welfare by subsidizing the expert and derive the optimal subsidy to be given to the expert in order to maximize the social welfare.

We also study the implications for this search process when the expert does not provide perfect information. In the real world, an expert may not have access to perfect information, or the marginal cost of providing information may be dependent on the expert's quality and the expert may choose to offer noisy information in order to maximize her profit. This can lead to interesting competition dynamics among experts. We derive the optimal search strategy in a duopolistic setting when the two experts provide noisy information. We analyze the equilibrium outcomes not only when an expert can set the price but also when she can choose the quality to offer.

Though competition dynamics is a well studied topic, we did not find relevant literature geared towards sequential search setting. In some cases we were able to draw intuition about analyzing equilibrium characteristics from the literature but most of the research focused on selecting the price and quality is based on simple models that can be solved analytically.

Chapter 3

Digital Goods Pricing Using Partially Censored Information

3.1 Introduction

We consider a sequential pricing problem faced by a seller with an unlimited supply of identical goods, who interacts with a stream of buyers with i.i.d. valuations. This is a classic problem sometimes referred to as "learning and earning". The unlimited supply world is applicable to digital goods, which have zero or very low marginal costs of production / storage e.g., iTunes apps. Consider a seller of such a product who is interested in maximizing her infinite horizon discounted profit. There is an underlying valuation distribution on the potential population of buyers, reflecting how much each potential buyer values that product. The seller is not aware of the true valuation of her product but can learn it over time through interaction with various buyers. Lack of information about the valuation introduces an exploration vs. exploitation tradeoff as on the one hand the seller wants to exploit the price which has been yielding the highest revenue so far, and, on the other hand, in order to learn the demand curve the seller may have to compromise on some profit in order to explore with other prices.

Dynamic pricing is a fundamental problem class within revenue management [87], with numerous variations covering perishable and non-perishable goods, homogeneous and in-homogeneous customer populations, finite and infinite inventories and time horizons, and known and unknown demand distributions. The exploration/exploitation trade-off arises when the response of an individual customer is uncertain, and the underlying distribution of the uncertainty is itself unknown. The optimal policy in these problems can be characterized using Bellman's equation, but exact solutions are typically intractable. Practical strategies can be developed either by solving a simplified version of the problem to optimality, or by considering heuristics that can be shown to have desirable regret bounds, i.e., asymptotic

growth rates on the opportunity cost incurred by using a suboptimal strategy. The latter approach has enjoyed special popularity in dynamic pricing problems: see e.g., [15, 17, 49]. In many settings, a simple myopic (greedy) pricing strategy is sufficient to achieve sublinear growth in the opportunity cost over time, possibly requiring a minor modification in order to avoid "confounding prices" that provide no new information [50]. However, such results are dependent on the modeling assumptions made in defining the problem;

We consider a model in which buyers arrive independently in a sequential manner with unit demand; the seller quotes a price and the buyer gives her decision about buying or not buying the item. The seller learns the demand curve over time and adjusts the price accordingly. This setting is also adopted by Blum et al. [15] and Kleinberg and Leighton [58], and in numerous other studies. Variations on this model, retaining the core assumption of independent sequential customer arrivals, have also been studied. Besbes and Zeevi [13] formulated a non-stationary demand model for a problem where the underlying valuation distribution changes at some point; the seller knows *a priori* both valuation distributions, but not the exact point in time when the change occurs from one distribution to the other. den Boer and Zwart [34] investigate another dimension of the problem, where the seller is assumed to have finite inventory; in this case, a maximum likelihood-based myopic pricing strategy is again shown to possess desirable regret bounds.

Some proposed solutions reduce the pricing problem to an instance of the well-known multi-armed bandit problem [80, 61, 58]. The set of allowable prices is discretized, with each price representing an arm of the bandit. The probability that the arm will yield a reward is equal to the probability that the buyer will buy the product at that price. Numerous algorithms with desirable regret bounds are available for multi-armed bandits [58, 57, 3, 44], though these techniques do not always work well in practice [23, 27, 89] (see Chapter 4 for detailed discussion on bandit algorithms). Another class of methods for this problem is the class of knowledge gradient policies [83, 78], which are simple to implement and often produce competitive empirical performance. In the context of pricing, however, the multi-armed bandit model is not ideal. The classical bandit framework assumes the arms of the bandit to be independent, which seems unrealistic when two arms are taken to represent similar prices for the same good. Furthermore, if the arms are independent, even an optimal strategy has a strictly positive probability of converging to a suboptimal price, known as incomplete learning [81, 69, 16]. More recent work allows for correlations between arms. Ryzhov et al. [84], and Mersereau et al. [71] show that complete learning can be achieved by a myopic strategy if the correlations are highly structured, even leading to finite (bounded) expected cumulative regret with undiscounted rewards. Nonetheless, existing models for correlated arms do not allow for the monotonic structure of a demand function; see Chhabra and Das [23] for more discussion of this issue.

We consider a different issue in dynamic pricing, namely the problem of *censored information*. Like many other studies, we assume that each customer has a certain valuation of the product, a concrete economic quantity that the customer compares against the quoted price in order to decide whether to buy the item. However, in practice, the decision-maker typically

does not observe this quantity, only whether it was greater than the quoted price (if the purchase occurred) or smaller (if not). Censored information makes it difficult to apply Bayesian inference on the valuation distribution. Harrison et al. [50] circumvent this issue by using a binary prior; however, if the set of possible valuation distributions is continuous (e.g., normal or uniform with unknown parameters), a censored observation will only lead to a conjugate belief model in certain special cases. For instance, Lariviere and Porteus [60] examines such a model with an exponential demand distribution, a gamma prior on the rate parameter, and censored information when demand exceeds available inventory.

In this chapter, we consider a different model where censored information admits a conjugate Bayesian belief. We assume that customer valuations are i.i.d. uniform on the interval $[0, Z]$, where Z is unknown. The uniform valuation distribution is equivalent to a linear demand assumption, in the sense that the probability of a purchase is a linear function of the price. Linear demand of this kind is one of the standard models considered in the literature [69, 50]. In this setting, the Pareto distribution is a natural choice for representing our prior beliefs about Z [33]. Our analysis shows that the posterior distribution remains Pareto as long as customers reveal their true valuations after buying the product; we do not, however, require them to do so if they do not buy. Intuitively, once the customer has purchased a digital good, she will presumably never need to purchase it again, and thus would be willing to reveal her true valuation to the seller. On the other hand, if no purchase occurs, the customer may wish to return to the seller later, and thus has an incentive not to reveal her true valuation. The main result here is that a myopic pricing strategy is optimal for the uniform-Pareto learning model with censored information. This result dovetails with earlier work on myopic strategies for structured learning problems, including, but not limited to, the dynamic pricing context. We motivate this result by first deriving a knowledge gradient (KG) strategy, a Bayesian one-step look-ahead method previously applied in ranking and selection [48] and multi-armed bandit problems [84]. Even with censored information (buyer reveals only when she buys but not otherwise), the price chosen by KG turns out to be exactly identical to the price chosen by the myopic strategy. We then show that the myopic strategy is optimal among all strategies that price below the largest valuation observed thus far, and that the cumulative regret incurred by this strategy is finite, that is, bounded by a constant that does not depend on the time index, even over an infinite time horizon. Our analysis shows that myopic pricing continues to retain its desirable theoretical properties in a setting with censored information. This is one of the first results showing Bayes optimality of myopic pricing in a dynamic pricing setting, rather than deriving asymptotically optimal regret bounds. We also present empirical evidence, using three real-world datasets, that the uniform-Pareto model offers a convenient way to model beliefs while promoting good performance.

3.2 The Model

Consider a seller of digital goods who interacts with a stream of non-adversarial buyers, each with unit demand. The seller wishes to set prices sequentially so as to maximize infinite horizon discounted revenue $\pi = \sum_{t=0} \delta^t \pi_t$, where δ is a discount factor and π_t is the profit from the transaction that occurs at time t.

The buyers' valuations are i.i.d draws from a known distribution function $f_v(x)$ with unknown parameters. At each time t, the seller quotes a price q_t. An arriving buyer with valuation v_t sees the quoted price, and chooses not to buy if $v_t < q_t$ and chooses to buy otherwise.

We assume that if the buyer chooses to buy, she also reveals her valuation v_t to the seller. This is not a very strict assumption as the buyer's demand has been fulfilled and it does not hurt the buyer to disclose her valuation. It is unlikely that the buyer will want to buy the same good again as digital goods are not perishable. If the buyer decides not to buy then she does not disclose her valuation. This makes sense because she may have repeat interactions with the seller and would prefer the seller not to know her true valuation. Therefore, in this case, the seller receives left-censored values: she knows that the true valuation of the buyer, in the case that she does not buy, is less than the quoted price, but does not know by how much.

To price the item optimally the seller has to learn the buyers' valuations over time. We study optimal pricing when the underlying valuation distribution is known to the seller to be uniform in $[0, Z]$ where Z is finite but she is unaware of the value of Z. The seller faces an exploration-exploitation dilemma due to the uncertainty in the value of Z. The assumption on our valuation distribution is equivalent to the assumption of having a linear demand model which is one of the most widely used models in the pricing literature [69, 71, 23].

There exists a single optimal price if the valuation distribution $f_v(x)$ is known to the seller. In our model, this is equivalent to knowing the value of Z *a priori*. Then the single optimal price can be calculated as follows:

$$\Pr(\text{Buy}|q, Z) = \begin{cases} 1 - \frac{q}{Z} & q < Z \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_{\text{opt}} = \max_q (q \Pr(\text{Buy}|q, Z))$$

$$q_{\text{opt}} = \frac{Z}{2} \Rightarrow \pi_{\text{opt}} = \frac{Z}{4}$$

Since the underlying valuation distribution is known to be uniform, we model the seller as maintaining a Pareto prior on Z, the unknown parameter of the uniform distribution. The Pareto is the conjugate prior for the uniform in this case. The seller's beliefs are then fully represented by the two parameters (a and b) of the Pareto distribution:

$$f_Z(x; a, b) = \frac{ab^a}{x^{a+1}} \quad x > b$$

Note, that the expected value of Z is not finite if $a < 1$. In order for the seller to always have finite mean belief distribution we assume that $a > 1$ at all times.

3.3 The Seller's Bayesian Updates

We start by simply showing the form of the updates for a seller given any quoted price q. In the next section, we discuss strategies for actually setting q. The seller quotes a price q to an arriving buyer, and the buyer either decides to buy and disclose her valuation to the seller, or leave. Depending on the information, the seller gets from the buyer's decision, she updates her belief on Z as follows for the two cases:

The buyer buys: When the buyer chooses to buy, she also discloses her valuation, which comes from a uniform distribution. Since we have a Pareto prior on the value of Z which is conjugate to the uniform distribution, our posterior also remains Pareto:

$$\Pr(\text{Buy}|q, Z = x) = \begin{cases} 1 - \frac{q}{x} & q < x \\ 0 & \text{otherwise} \end{cases}$$

$$\Pr(\text{Buy}|q) = \int_0^\infty \Pr(\text{Buy}|q, Z = x) f_Z(x) \, dx$$

$$= 1 - F_Z(M; a, b) - \frac{qa}{b(a+1)} (1 - F_Z(M; a+1, b)) \quad M = \max(q, b)$$

$$f_Z(x|\text{Buy}) = \frac{f_Z(x) f_v(v|x)}{\int_0^\infty f_Z(x) f_v(v|x) \, dy}$$

$$= \begin{cases} f_Z(x; a+1, \max(v, b)) & x > \max(v, b) \\ 0 & \text{otherwise} \end{cases}$$

The buyer does not buy: If the buyer does not reveal her valuation, the observation that the seller receives is no longer a sample from a uniform distribution, so the seller's posterior may not be Pareto. The seller receives a signal that the price is higher than the buyer's valuation. However, interestingly, it turns out that as long as the current price is less than the parameter b of the Pareto distribution, the posterior is, in fact, Pareto. We later show that it is suboptimal for the seller to price the item above b, so her optimal selection of price

always leads to a Pareto posterior.

$$\Pr(\neg\text{Buy}|q, Z = x) = \begin{cases} \frac{q}{x} & q < x \\ 1 & \text{otherwise} \end{cases}$$

$$\Pr(\neg\text{Buy}|q) = \int_0^\infty \Pr(\neg\text{Buy}|q, Z = x) f_Z(x)\, dx$$

$$= F_Z(M; a, b) + \frac{qa}{b(a+1)}(1 - F_Z(M; a+1, b)) \quad M = \max(q, b)$$

$$f_Z(x|\neg\text{Buy}) = \frac{f_Z(x)\Pr(\neg\text{Buy}|x)}{\Pr(\neg\text{Buy})}$$

$$\begin{cases} \frac{q f_Z(x)}{x \Pr(\neg\text{Buy}|q)} & x \geq q \\ \frac{f_Z(x)}{\Pr(\neg\text{Buy}|q)} & x < q \end{cases}$$

Therefore, we see that if the parameter $b > q$ (the price) then the posterior is $f_Z(x; a+1, b)$ (i.e., Pareto with parameters $a+1$ and b).

3.4 Strategies

The problem in front of the seller is how to set the price q_t at any point in time t. The seller's goal is to maximize her infinite horizon discounted profit. As the seller is not aware of the underlying demand, it is hard to find the optimal price analytically. Therefore, we study the following two strategies for the seller: (1) the Bayesian myopic pricing strategy and; (2) the knowledge-gradient (i.e., one-period Bayesian look-ahead) strategy.

3.4.1 Bayesian Myopic Pricing Strategy

The simplest strategy is to price the item greedily, in order to maximize the expected profit from the next interaction with a buyer. At time t:

$$\mathbb{E}(\pi_{t_{\text{myopic}}}) = \max_{q_t}(q_t \Pr(\text{Buy}|q_t))$$

$$= \max_{q_t}\left(q_t(1 - F_Z(M; a, b)) - \frac{q_t^2 a(1 - F_Z(M; a+1, b))}{b(a+1)}\right) \quad M = \max(b, q_t)$$

The myopic profit is maximized when $\max(b, q_t) = b$ and $q_t = \frac{b(a+1)}{2a}$ and it is:

$$\mathbb{E}(\pi_{t_{\text{myopic}}}) = \frac{b(a+1)}{4a} \tag{3.1}$$

For details, see Appendix A.1.1.

Setting initial priors   Let a_0 and b_0 represent the initial value of the parameters (at time $t = 0$) for the Pareto distribution. From the updates, we know that the b parameter never decreases from its initial value. We also know that the myopic price is $\frac{b(a+1)}{2a} \geq \frac{b}{2}$ when the seller's state parameters are a and b. Thus, the minimum price that can be offered using the myopic strategy is $\frac{b_0}{2}$. However, if the optimal price is less than $\frac{b_0}{2}$ then the myopic strategy will never converge to the correct price, and if this minimum price $\left(\frac{b_0}{2}\right)$ is greater than the maximum possible valuation then the myopic profit will be zero.

We want to find how close to the optimal can the myopic strategy perform in such a setting where the priors are misspecified. If the valuations are drawn from a uniform distribution $[0, Z]$ then the optimal price to charge is $Z/2$ which results in an expected profit of $Z/4$. Consider a setting where $\frac{Z}{2} < \frac{b_0}{2} \Rightarrow Z < b_0$. At time $T = t$, according to the Bayesian updates the a parameter is incremented by one in each round therefore $a_t = a_0 + t$; the value of the b parameter increases only if the observed valuation in case of a buy is greater than the b parameter, however the valuation will always be less than $Z < b_0$ so $b_t = b_0$. Note, that if $\frac{b_0}{2} > Z$, the expected profit using the myopic strategy will be zero. When $\frac{b_0}{2} < Z$, the ratio of the profit earned using the myopic strategy and the optimal fixed price strategy when the true distribution is known at time t is:

$$\frac{\pi_{t_{\text{myopic}}}}{\pi_{t_{\text{opt}}}} = \frac{\frac{b_t(a_t+1)}{2a_t}\left(1 - \frac{b_t(a_t+1)}{2a_t Z}\right)}{Z/4} < \frac{2b_0}{Z}\left(1 - \frac{b_0}{2Z}\right) \tag{3.2}$$

For $b_0 > Z$, Equation 3.2 decreases as b_0 increases and eventually becomes zero when $b_0 = 2Z$. Therefore, if the seller has no information about the minimum value of the true Z then she should conservatively set the value of b_0 very low in order to avoid being in a situation where she can never learn the true price.

3.4.2   Knowledge-Gradient

The knowledge-gradient (KG) strategy is a one-period Bayesian look-ahead strategy. The seller assumes that, starting from the next round, her beliefs will be fixed forever, and she will choose the myopic price every time after that. Mathematically:

$$\pi_{KG} = \max_{q_t}\left(\pi_t + \frac{\delta}{1-\delta}\left(\pi_{t+1_{\text{Myopic}}}\right)\right)$$

Then:

$$\mathbb{E}(\pi_{KG}) = \max_{q_t}\left(\mathbb{E}(\pi_t) + \frac{\delta}{1-\delta}(\Pr(\neg\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\right.$$
$$\left. + \Pr(\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1|\text{Buy}})))\right) \tag{3.3}$$

Surprisingly, it turns out that the KG optimal price (from Equation 3.3) is equal to the myopic price.

Theorem 1 *The KG optimal price for the uniform $[0, Z]$ valuation distribution with left-censored observations is equal to the myopic price when the seller maintains a Pareto prior on Z.*

Proof: To calculate the KG optimal price, we first calculate the expected myopic profit for the $(t+1)^{\text{st}}$ round. As we saw in Section 3.3, the posterior distribution when the buyer does not buy in our setting depends on the value of $\max(b, q_t)$; we evaluate the two cases: $b \geq q_t$ and $b < q_t$ separately:

Case 1: $\max(b, q_t) = b$

As discussed in Section 3.3, when the buyer decides not to buy the good (i.e., $v < q_t$), the posterior is also Pareto $f_Z(x; b, a+1)$. Therefore, the myopic profit (as given by Equation 3.1) for this case is $\frac{b(a+2)}{4(a+1)}$. The expected profit when the buyer does not buy is given by:

$$\max_{q_{t+1}}(\mathbb{E}(\pi_{t+1}|\neg\text{Buy})) = \frac{1}{\Pr(\neg\text{Buy}|q_t)} \left(\int_b^\infty dx f_Z(x) \int_0^{q_t} dv \frac{b(a+2)}{4(a+1)} f_v(v|x) \right) \quad (3.4)$$

If the buyer decides to buy (i.e., $v > q_t$), she also discloses her valuation; the posterior is Pareto and is $f_Z(x; \max(b, v), a+1)$ where v is the buyer's valuation. The myopic profit (using Equation 3.1) in case of a buy is given by $\frac{\max(b,v)(a+2)}{a+1}$. The problem here is that the profit depends on the valuation of the buyer which will be known to the seller only after the current round is over. The seller only knows that if the buyer decides to buy then her valuation is greater than the price. The expected profit at time $t+1$ when the buyer buys is given by:

$$\max(\mathbb{E}(\pi_{t+1}|\text{Buy})) = \frac{1}{\Pr(\text{Buy}|q_t)} \left(\int_b^\infty dx f_Z(x) \int_{q_t}^x dv \frac{\max(b,v)(a+2)}{4(a+1)} f_v(v|x) \right)$$

$$= \frac{1}{\Pr(\text{Buy}|q_t)} \left(\int_b^\infty dx f_Z(x) \left(\int_{q_t}^b dv \frac{b(a+2)}{4(a+1)} f_v(v|x) + \int_b^x dv \frac{v(a+2)}{4(a+1)} f_v(v|x) \right) \right)$$

$$(3.5)$$

The total expected profit at time $(t+1)$ is given by the probability-weighted sum of Equations

3.4 and 3.5:

$$\mathbb{E}(\pi_{t+1_{\text{Myopic}}}) = \Pr(\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1}|\text{Buy})) + \Pr(\neg\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1}|\neg\text{Buy}))$$

$$= \int_b^\infty dx f_Z(x) \left(\int_0^{q_t} dv \frac{b(a+2)}{4(a+1)} f_v(v|x) + \int_{q_t}^b dv \frac{b(a+2)}{4(a+1)} f_v(v|x) + \int_b^x dv \frac{v(a+2)}{4(a+1)} f_v(v|x) \right)$$

$$= \int_b^\infty dx f_Z(x) \left(\int_0^b dv \frac{b(a+2)}{4(a+1)} f_v(v|x) + \int_b^x dv \frac{v(a+2)}{4(a+1)} f_v(v|x) \right)$$

$$= \frac{a^2 b(a+2)}{4(a-1)(a+1)^2}$$

The expected KG profit using Equation 3.3 is given by:

$$\mathbb{E}(\pi_{KG}) = \max_{q_t} \left(\mathbb{E}(\pi_t) + \frac{\delta}{1-\delta}\left(\mathbb{E}(\pi_{t+1})_{\text{Myopic}}\right) \right)$$

$$= \max_{q_t}\left(\mathbb{E}(\pi_t)\right) + \frac{\delta}{1-\delta}\mathbb{E}\left(\pi_{t+1}\right)_{\text{Myopic}}$$

$$= \frac{b(a+1)}{4a} + \frac{\delta}{1-\delta}\left(\frac{a^2 b(a+2)}{4(a-1)(a+1)^2}\right) \tag{3.6}$$

Surprisingly, the expected profit at time $(t+1)$, given by Equation 3.6, does not depend on q_t. Therefore, the KG optimal strategy when $b > q_t$ is the same as the myopic strategy.

Case 2: $\max(b, q_t) = q_t$ The probability of not buying a good as described in Section 3.3 is given by:

$$\Pr(\neg\text{Buy}|q_t) = F_Z(q_t) + \frac{q_t a}{b(a+1)}(1 - F_Z(q_t; a+1, b))$$

$$1 - \frac{1}{a+1}\left(\frac{b}{q_t}\right)^a$$

We can show that the maximum profit possible when pricing the good such that $q_t > b$ is always less than that when $b > q_t$ (see Appendix A.1.2), so the seller will never choose a price greater than b.[1] Therefore, the optimal one-step look-ahead pricing strategy is identical to the myopic strategy. This result suggests that the myopic strategy is in fact optimal, even when information is censored. We will now examine this issue. □

3.5   Optimality of the Myopic Strategy

To argue the optimality of the Bayesian myopic strategy, we first restrict the set of allowable strategies in the following manner. We consider only those strategies that, for any t, choose

---

[1]Intuitively, it does not make sense for the seller to price the item higher than the myopic price, because the information content in case of a no buy is much less than the information content in case of a buy.

prices q_t below the current value of the Pareto parameter b. From Section 3.4, we know that the myopic strategy satisfies this condition.

$$q_t = \frac{b(a+1)}{2a} < b \quad \text{for } a > 1$$

Now consider a model where the seller gets perfect information after each interaction with a buyer. That is, the seller observes the exact valuation of the buyer regardless of whether the buyer purchases the product. The updates in this complete-information setting for both buying and not buying are Pareto:

$$f_Z(x; a, b|v) = f_Z(x; a+1, \max(b, v))$$

After k buyers have visited, let $m(k)$ represent the maximum of the valuations of these k buyers. If the seller starts with initial parameters $(a = a_0, b = b_0)$ for a Pareto prior on the value of Z ($v \sim U[0, Z]$) at time $t = 0$, the posterior distribution after k buyers have visited is given by:

$$f_Z(x; a_k, b_k|k \text{ buyers have visited}) = f_Z(x; a_0 + k, \max(b_0, m(k)))$$

Note that the parameter b_0 also provides the information that the seller initially assumes $Z > b_0$ because $f_Z(x; a_0, b_0) = 0$ for $x < b_0$ for the Pareto distribution, so it is advisable to choose a small value for b_0.

Now, return to our setting, where the buyer discloses her valuation if she chooses to buy, but not otherwise. Since we are only considering strategies that price below the current value of the Pareto parameter b, the conjugacy of our belief model is preserved even with censored information. Let $m_b(k)$ represent the maximum of all the valuations observed among those who chose to buy after a sequence of buyers has arrived. After k buyers have visited the seller, the posterior distribution is given by:

$$f_Z(x|k \text{ buyers have visited}) = f_Z(x; a_0 + k, \max(b_0, m_b(k)))$$

Observe that the a parameter of the Pareto distribution is always equal (incremented by 1) in both the complete-information and partial-information models. Let $b_c(k)$ and $b_p(k)$ represent the b parameter for the Pareto distribution after k buyers have arrived for complete and partial information respectively. We now show that $b_p(k)$ is actually equal to $b_c(k)$ for any allowable strategy, as long as the two sellers start with the same prior.

Theorem 2 *At any time $k > 0$, the parameters of the posterior distribution in our settings (partially-censored information about buyers' valuation) for a Pareto prior are equal to the parameters of the posterior distribution with complete information using any pricing strategy that prices the item less than the b parameter of the Pareto distribution about the buyers' valuation if the initial value of the parameters is chosen to be equal (i.e., $b_p(k) = b_c(k)$ if the same b_0 is chosen for both complete information and partial information model).*

Proof: At time $k = 0$ we assume that the Pareto prior in both the complete and partial information cases is the same (i.e, $b_p(0) = b_c(0) = b_0$ and $a_c(0) = a_p(0) = a_0$.

The a parameter is incremented by one every time a buyer arrives under either setting, so it is equal for both complete and partial information at any time k (i.e, $a_p(k) = a_c(k) = a_0 + k$). We will now show the equality of the b parameter using induction:

Let's assume that at any time k, $b_p(k) = b_c(k)$; using this assumption we will show that $b_p(k + 1) = b_c(k + 1)$ no matter what the buyer decides.

Let v_k represent the valuation of the kth buyer.

Case 1: The buyer buys When the buyer decides to buy, she discloses her valuation, v_k, both for the partial and complete information setting. Therefore, $b_p(k + 1) = \max(v_k, b_p(k))$ and $b_c(k + 1) = \max(v_k, b_c(k))$. Since we assume that $b_p(k) = b_c(k)$, therefore, in case of a buy $b_p(k + 1) = b_c(k + 1)$.

Case 2: The buyer does not buy If the buyer decides not to buy, she does not disclose her valuation in partial information setting, therefore, the b parameter remains same: $b_p(k + 1) = b_p(k)$.

The buyer discloses her valuation in complete-information setting. We assume that the price, q_{k_c}, is less than $b_c(k)$, and at no buy — $v_k < q_{k_c}$, therefore, $b_c(k + 1) = \max(b_c(k), v_k) = b_c(k)$. Thus, in case of a no buy also $b_p(k + 1) = b_c(k + 1)$ because $b_p(k) = b_c(k)$ by assumption.

For both cases above, we show that $b_c(k + 1) = b_p(k + 1)$ given $b_c(k) = b_p(k)$ and it is already assumed that $b_p(0) = b_c(0)$. Therefore, by the induction hypothesis, $\forall \, k \quad b_p(k) = b_c(k)$. $\quad\square$

Corollary 1 *Among all strategies in the allowable set, the myopic pricing strategy is optimal.*

Since every strategy in the allowable set produces the same sequence of posterior distributions in either the complete-information setting or the censored-information setting, it follows that the same cumulative revenue will be generated in both cases. However, the myopic strategy is optimal in the case of complete information. If we observe the buyer's exact valuation regardless of the buyer's decision, it follows that the sequence of posterior distributions is independent of the price. Thus, any pricing decision made at time t will have no effect on future revenues after time t, and it is optimal to price myopically. Finally, since any allowable strategy will generate the same revenue under partial information, the myopic strategy will continue to outperform any other strategy.

It remains to argue that the myopic strategy continues to be optimal when we expand the set of allowable strategies to allow arbitrary pricing decisions (include those above the Pareto parameter b). We offer the following intuition in support of this claim:

1. Non-myopic decisions can be optimal if they allow us to collect new information that compensates for lost revenue over time. However, in our setting, higher prices increase the likelihood of a lost sale, and thus the risk that we will receive less information rather than more.

2. Reducing our ability to collect information should also reduce the revenue we can generate. Thus, the maximum achievable revenue under partial information should not be greater than the maximum revenue under complete information. However, the revenues generated by the myopic strategy are the same in both settings.

We conclude that the myopic strategy is optimal for the uniform-Pareto pricing problem with censored information. This result relates to earlier work by Harrison et al. [50] and Mersereau et al. [71] on greedy and semi-greedy policies in Bayesian dynamic pricing. We provide further evidence in support of the myopic policy in the form of a finite regret bound.

3.6   Regret Bounds

For a multi-armed bandit with independent arms the regret grows as $\Omega(\log T)$ [59]. Mersereau et al. [71] demonstrate the existence of finite regret in a special case with correlated arms where the expected reward of each arm is a linear function of an unknown scalar with known prior. The existence of finite regret in general means that the algorithm learns very fast. We show that, the myopic Bayesian algorithm in our model also has finite regret, implying that it learns Z quickly.

Theorem 3 *The regret for the myopic Bayesian policy for the partially-censored information setting (i.e, the buyers disclose their valuation in case of a buy) when the true valuations are drawn from the uniform distribution in $[0, Z]$, where Z is unknown and the seller maintains a Pareto prior, is finite.*

Proof: We begin by considering the complete information setting, where the true customer valuations are observed regardless of the customer response. Note that, in this setting, the Pareto parameters a_t and b_t do not depend on the pricing strategy.

First, note that the time t profit is given by $\pi_q^t = q_t \left(1 - \frac{q_t}{Z}\right)$. Now, the regret is defined in the case where Z is known.

$$\text{Regret}(Z, T, \text{Myopic}) = \sum_{t=0}^{T} \delta^t \mathbb{E}\left(\left(\max_{q_t}(\pi_q^t) - \max_{q_t}(\mathbb{E}_Z(\pi_q^t)))\big| Z\right)\right) \tag{3.7}$$

Since Z is itself a random variable, we take an expectation over the prior distribution:

$$\text{Risk}(T, \text{Myopic}) = \mathbb{E}_Z(\text{Regret}(Z, T, \text{Myopic})$$

$$= \mathbb{E}_Z\left(\sum_{t=0}^{T} \delta^t \left(\frac{Z}{4} - \frac{b_t(a_t + 1)}{4a_t}\right)\right)$$

$$= \sum_{t=0}^{T} \delta^t \left(\frac{a_t b_t}{4(a_t - 1)} - \frac{b_t(a_t + 1)}{4a_t}\right)$$

$$= \sum_{t=0}^{T} \delta^t \left(\frac{b_t}{4a_t(a_t - 1)}\right)$$

$$< \frac{E(Z)}{4} \sum_{t=0}^{T} \frac{\delta^t}{a_t(a_t - 1)} \quad (b_t \text{ is bounded by the value of true Z })$$

$$= \frac{E(Z)}{4} \sum_{t=0}^{T} \delta^t \left(\frac{1}{a_t - 1} - \frac{1}{a_t}\right)$$

$$< \frac{E(Z)}{4} \left(\frac{1}{a_0 - 1} + (\delta - 1) \sum_{t} \frac{\delta^t}{a_t}\right)$$

$$< \frac{E(Z)}{4(a_0 - 1)}. \tag{3.8}$$

Now, recall that the myopic strategy produces the same sequence of posterior parameters in the censored information setting. It follows that the expected revenue of this strategy remains unchanged, so the bound continues to hold. $\qquad \square$

3.7   Empirical Results

Many algorithms do not perform well on real datasets even though they have good convergence bounds. This is perhaps due to the stringent theoretical assumptions needed to get good bounds. To check the robustness of our model, we conduct an empirical test on a valuation distribution derived from real data. We use three different datasets described as follows:

1. **eBay auction dataset** We use the eBay auction dataset[2] to synthesize a valuation distribution [55]. This dataset contains bidding information from eBay for several auctions of Palm Pilot M515 PDAs. While this is not a digital good, it is one of the few datasets available that would allow us to report a valuation distribution. In

---

[2] downloaded from `http://www.modelingonlineauctions.com/datasets`; last accessed January, 2014

(a) eBay auction dataset   (b) User ratings on a joke from the Jester dataset   (c) Bids on a keyword from Yahoo! sponsored search auction
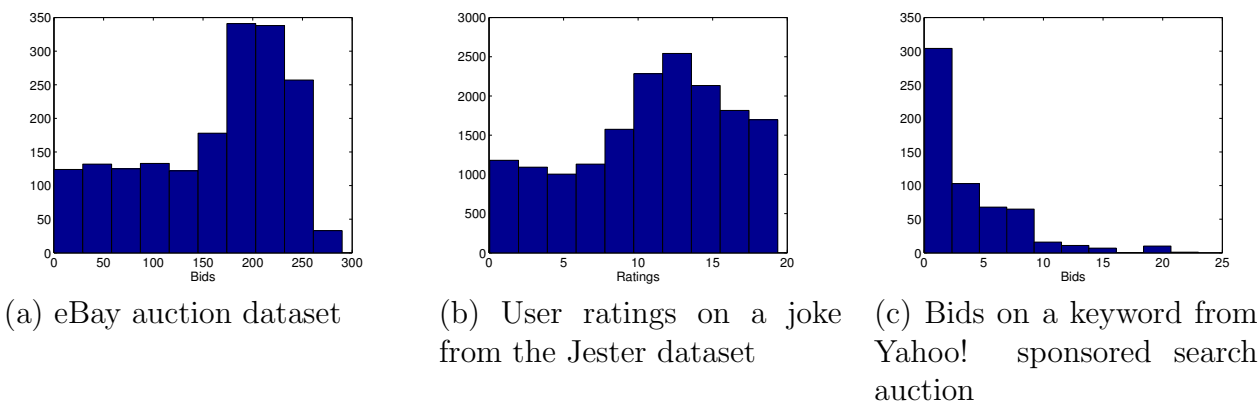
Figure 3.1: (a) Histogram of the highest bid of each bidder from the eBay auction data. A large number of bids lie in between 175 and 250; perhaps the market value of the item is in this range. (b) Ratings of a joke from 16452 users. Originally a rating was a number between -10 to 10. However, to avoid negative numbers we add 10 to all the ratings. (c) Histogram of the highest bid of each bidder for one of the keywords from the Yahoo! sponsored search keyword auction data. Clearly, the valuations are not uniformly distributed for any of these datasets.

order to come up with realistic valuations based on this dataset, we first combined bidding data from all 350 auctions. We then found the single highest bid put in by each bidder, which is a proxy for how much she values the item (since that is the maximum amount she is willing to pay, unless she is the winner, in which case her valuation may be somewhat higher). We found 1783 unique bidders, and 350 Palm Pilots were sold. Therefore, it is possible that some of the valuations at the higher end of the distribution could actually be even higher, because a winning bidder has no need to reveal her true valuation with an even higher bid. Figure 3.1 (a) shows the valuation distribution extracted from this data.

2. **Jester dataset** This dataset contains ratings for a set of 100 jokes by 29,483 users on a scale from -10.00 to 10.00 (continuous rating) from Jester[3], an online joke recommender system [45]. Not all users have rated all the jokes, so the number of ratings for each joke varies. We pick one joke arbitrarily and its associated ratings (by 16,452 users). Here, we use the rating for the joke as a proxy to its valuation. Although, it is not a valuation dataset, we assume that a higher rating implies a higher value and thus can be used as a reliable proxy. Since the ratings can be negative, while using them as valuations we add 10 to each of the ratings to move it to a scale of 0 to 20 so that all our valuations are non-negative without any loss of information. Figure 3.1 (b) shows

---

[3]Jester can be found at `http://eigentaste.berkeley.edu/user/index.php`. The data can be downloaded from `http://eigentaste.berkeley.edu/dataset/jester_dataset_1_1.zip`; last accessed January 2014.

the valuation distribution extracted from the ratings of the selected joke.

3. **Yahoo! advertiser bidding data** Yahoo! webscope has released data for advertisers' bids on the top 1000 keywords for sponsored search auction from June 15, 2002 to June 14, 2003 [1][4]. We choose a keyword which had the highest number of unique bidders (588 unique bidders) and use the bidders' highest bid as a proxy to their valuation similar to what we did for the eBay auctions dataset. We removed the three largest bids (values: 400, 50, 50) from the dataset as outliers because the remaining 585 bids were all less than 23. Figure 3.1 shows that the bids form a heavy-tailed distribution.

Clearly, neither of these underlying valuation distribution is uniform, so this is a realistic test for our pricing algorithm. We compare the Bayesian myopic pricing strategy to the two modifications based on Gittins index based strategy discussed below:

Gittins Index based strategy: This is a multi-armed bandit algorithm. We first discretize the prices — each price is equivalent to an arm of the multi-armed bandit. We have Bernoulli arms (the success probablity of each arm is equal to the probability that the buyer decides to buy at that price) and the seller maintains a Beta prior $B(\alpha, \beta)$ on the probability of success of each arm. Using this, the Gittins indices (or the dynamic allocation indices) are calculated for all the arms and the arm with the highest index is selected [61, 23]. On playing the arm, the parameter α is incremented if the buyer decides to buy and β is incremented if the buyer decides not to buy.

This Gittins strategy is optimal for a multi-armed bandit when all the arms are independent. While the arms for the pricing problem are not independent, this approach has shown good results [61, 23]. We experiment with some variations of the Gittins strategy to incorporate dependence.

In order to discretize the arms from the dataset of unique bidders, we find the minimum and the maximum bids and discretize the price uniformly in that range. The minimum bid is $0.01, and the maximum bid is $290. We fix the number of arms at 20. The two modifications of the Gittins index strategy that incorporate information on the price structure are as follows:

1. **Modified Gittins index strategy for our partial-information setting:** In our partial-information model, when the seller observes the true valuation of the good (in case of a buy), we update the beliefs on all the arms based on the true valuation seen. In case of a no buy, the seller does not get information about the true valuation but the seller knows that if a buyer decided not to buy at any price q, then she would decide the same for any price greater than q. Therefore, the seller increments the β parameter of all the arms with price greater than or equal to the current arm.

---

[4]downloaded from `http://webscope.sandbox.yahoo.com/catalog.php?datatype=a`; last accessed in January 2014

(a) eBay auction data   (b) Joke rating dataset   (c) Yahoo! sponsored search
                                                                       dataset
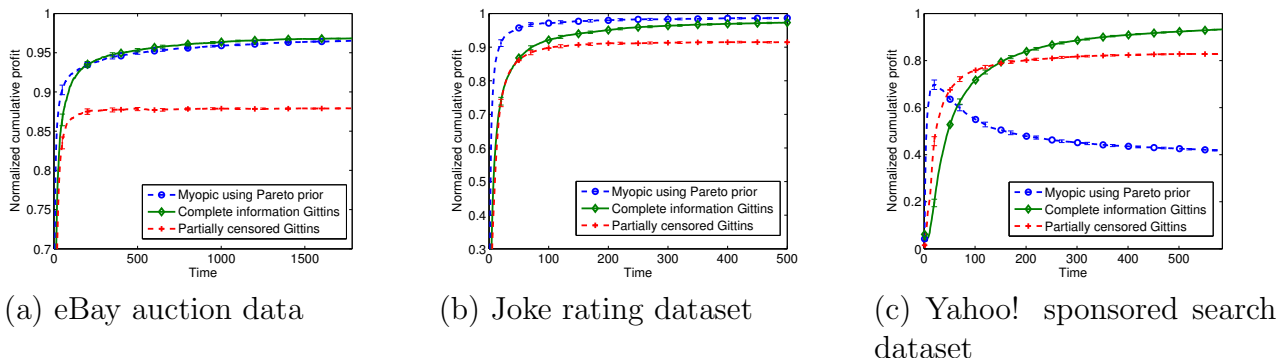
Figure 3.2: Comparison of the Pareto-prior based myopic technique with two proposed Gittins index based strategies on three dataset shown in Figure 3.1. Each graph shows the normalized cumulative profit averaged over 1000 iterations and 95% confidence intervals. Surprisingly, in (a) and (b) the Pareto-prior based myopic strategy for the partially censored case (the buyer disclosed her valuation in case of a buy) either close to or outperforms completely-informed Gittins-index-based strategy that receives valuation information both in case of a buy and no buy. However, in case of (c), the Pareto-prior based myopic strategy does not perform well. Perhaps, the heavy-tailed distribution for (c) is not suitable to be modeled as uniform.

2. **Modified Gittins index strategy for the complete-information setting:** In the complete information world, the seller observes the valuation both in case of a buy and a no buy, and thus updates all the arms every time a buyer visits. We use this as an upper bound on the performance of a Gittins-index style strategy, even though we do not have complete information.

For the eBay and the Yahoo! auction dataset, we average our results over multiple random permutations of the bids. However, the Jester dataset is much bigger (more than 16000 ratings for the selected joke) so we instead repeatedly sampled 500 unique buyers from the valuation data and allowed our seller strategies to interact with a stream of buyers with those valuations. We ran 1000 such experiments. The discount factor, δ, for these experiments is 0.95. Figures 3.2 (a), (b) & (c) show the performance of the three strategies discussed above on theeBay auction dataset, the Jester dataset and Yahoo! keyword auction data respectively. In Figure 3.2 (a) & (b) the Pareto prior based myopic strategy surprisingly performs better than not only the Partial information based Gittins strategy but also the Gittins-index strategy with complete information. However, in Figure 3.2 (c) the Pareto-prior based strategy does not perform well. This is perhaps because the fat-tailed keywords valuation distribution is not suitable for being modeled as a uniform distribution. The optimal price is $\sim$ \$5 but the seller sometimes observes a higher valuation during a buy which pushes the seller's belief to a price far higher than the optimal; the seller is unable to re-adjust, leading to a decrease in cumulative profit over time.

We also perform large-scale analysis on the Jester and the Yahoo! sponsored search auction datasets. For the Jester dataset, we use all the 100 jokes (equivalent of having 100 different goods being sold independently). For each joke, we first calculate the average discounted and undiscounted profit by averaging the respective profits over various iterations. The Jester dataset is a dense dataset and has a lot of ratings for each joke. In each iteration, a sample of 500 unique users are picked uniformly at random. We also compare the number of times each algorithm performs best for 100 jokes. For the Yahoo! dataset, we choose top 100 keywords which have the highest number of unique bidders (the number of bidders varied from 126 - 588 depending on the keyword). We find that bids on some keywords had outliers e.g., for one keyword with 167 bids all bids but one are less than 20, and one bid is at 1990. We remove such outliers with winsorization by censoring the data whose percentile lies outside of [1 99]. Since, the number of bidders is small, instead of sampling we generate random permutations of the bids for each keyword and average the profits over them. These profits are averaged over 1000 iterations.

We find that for all the 100 jokes and for 83 out of 100 keywords the Pareto-prior based myopic strategy performs the best in terms of average discounted profit. For the Yahoo! dataset, 8 times the complete-information based Gittins strategy and the remaining 9 times the partial-information based Gittins strategy are the winners.

The table below summarizes the average ratio of the total discounted profit earned using that policy and the total discounted profit achievable by a single fixed price strategy with prior knowledge of the true underlying valuation distribution for the two datasets:

Dataset	Modified Gittins with partial information	Modified Gittins with complete information	Pareto-prior based myopic strategy
Jester dataset	0.7577 ± 0.0020	0.7677 ± 0.0024	0.9072 ± 0.0024
Yahoo! dataset	0.5945 ± 0.0053	0.5437 ± 0.0079	0.6500 ± 0.0065

The Pareto-prior based myopic strategy outperforms the other two even in terms of the average discounted profits. Surprisingly, the modified Gittins index based strategy using partial information (which has access to the same amount of information as the Pareto-prior based strategy) yields higher discounted profits compared to the modified Gittins-index-based strategy using complete information in the Yahoo! dataset. The reason for this could be that initially the partial-information based Gittins policy performs better and due to discounting, the profits which have been earned early on are important. Even though the complete-information based Gittins strategy may have eventually outperformed the partial-information based Gittins, it would not have been captured in the discounted profits. We also compare the sum of undiscounted profits for the three strategies. The ratio of average profit earned and the profit achievable by a single fixed price strategy with prior knowledge of the true valuation distribution for different strategies are shown below:

Dataset	Modified Gittins with partial information	Modified Gittins with complete information	Pareto-prior based myopic strategy
Jester dataset	0.9162 ± 0.0014 (39)	0.9734 ± 0.0003 (0)	0.9735 ± 0.0009 (61)
Yahoo! dataset	0.7487 ± 0.0042 (22)	0.7963 ± 0.0064 (59)	0.5819 ± 0.0128 (19)

In the above table the number in the parenthesis correspond to the frequency of the corresponding strategy being the winner (e.g., on 61 out of 100 jokes the Pareto-prior based myopic strategy leads to highest undiscounted profits). The performance of the complete information based Gittins strategy is better than the partial information based Gittins strategy in terms of undiscounted profit for both the datasets indicating that this strategy eventually outperforms the other Gittins-index based policy. Surprisingly, the performance of the Pareto-prior based myopic strategy in terms of undiscounted rewards is worst for the Yahoo! dataset. Perhaps, the distribution of the bids for most of the keywords is fat-tailed like that of in Figure 3.1 (c), and is hard to be modeled by a uniform distribution which leads to suboptimal profits like in Figure 3.2 (c).

3.8 Discussion

There has been a resurgence of interest in the classic "learning while earning" problem of dynamic pricing in recent years, due in no small part to the increasing prevalence of electronic transactions. This also makes the problem more relevant for the infinite inventory case of digital goods. Much of the recent work in this area has found that myopic strategies are surprisingly effective, but has typically focused on proving regret bounds and discussing the asymptotic optimality of myopic policies and whether or not they will eventually learn the "right" price. We take a complementary approach, and show, for an important class of valuation distributions and the appropriate conjugate prior (the uniform distribution on $[0, Z]$ with a Pareto prior) when the myopic strategy is, in fact, Bayes optimal even when the seller receives partially censored information, and at the same time learns quickly. Our extensive analysis on the real-world datasets help in achieving better understanding of when linear-demand based myopic strategies discussed in this chapter can be optimal or close to it, and when they fail (e.g., heavy-tailed distribution).

An important future direction is to perform similar analyses with alternative valuation distributions and to understand which distributions are more suitable in real-world for such settings. Another direction is to understand optimality properties and regret bounds when the seller is faced with completely-censored information, instead of the partially-censored case we deal with in this chapter. In the latter case in particular, it may be difficult to find conjugate priors, so innovative algorithmic techniques may need to be used.

Chapter 4

Digital Goods Pricing Using Completely Censored Information

4.1 Introduction

In this chapter, we turn out attention to a modification of the digital goods pricing problem presented in Chapter 3. In that chapter, we had considered a special case where the buyer also discloses her valuation to the seller in case of a buy and proved the optimality of the Bayes-myopic strategy under certain assumptions. We also showed that the Bayes-myopic strategy has a finite expected regret (independent of time) which means that it also learns very fast. Here, we relax the constraint of the partial disclosing of buyers' valuations. The seller only receives a buyer's "binary" decision to buy or not. This increases the complexity of the problem since the seller now receives less information and has to learn from completely censored signals.

Contributions   In this chapter, we study the problem of revenue maximization in posted-price auctions of digital goods from the perspective of reinforcement learning and maximizing flow utility, rather than trying to achieve asymptotic regret bounds. We evaluate algorithms on simulated buying populations, with valuations distributed uniformly, exponentially, and log-normally as well as on the valuations synthesized using the real-world datasets from auctions of Palm PDA M515 at eBay [55], Yahoo! [1] keyword auctions and a jokes ratings dataset from Jester an online joke recommender system [45]. We find that regret-minimization algorithms from the multi-armed bandit literature are slow to learn in practice, and hence impractical, even for simple distributions of valuations in the buying population. We propose three alternatives: (1) a scheme based on Gittins indices that starts with different priors on the arms based on the knowledge that purchases at higher prices are less likely; (2) An extension of the knowledge gradient (i.e., Bayesian one-step look-ahead)

strategy for the completely censored case from the partial-information setting described in Chapter 3; (3) a new reinforcement learning algorithm for the problem, called LLVD, that is also based on a plausible linearity assumption on the structure of the demand curve. LLVD, however, maintains a Beta distribution as the seller's belief state, updating it using a moment-matching approximation. LLVD is (approximately) optimal when the linearity assumption holds, and empirically performs well not only for several families of valuation distributions but also on real-world datasets that violate the linearity assumption.

4.2   The Model

In our setting buyers arrive in a stream, each with an independent and identically distributed (i.i.d.) valuation v of the good from an unknown underlying distribution f_V. f_V can have support on $[0, \infty)$. At each instant in time, the seller quotes a price $q_t \in [0, \quad \infty)$, a potential buyer arrives with $v_t \sim f_V$, and chooses to buy if $v_t \geq q_t$ and not to buy otherwise. The seller has access to the history of her own pricing decisions, as well as the purchase decisions made by each arriving buyer. Her goal is to sequentially set q_t so as to maximize infinite horizon discounted profit. For any given distribution of buyer valuations f_V, under the assumption that buyer valuations are i.i.d. draws from f_V at each point in time, there is a single optimal price q_{OPT} that maximizes the seller's expected revenue.

4.3   Algorithms

Here we describe the four algorithms we compare for this problem: (1) a knowledge gradient based algorithm and a myopic algorithm; both of which based on the learning model from Chapter 3; (2) our new parametric algorithm, LLVD; (3) a Gittins-index based strategy with appropriately chosen priors; (4) UCB, a regret-minimizing algorithm from the multi-armed bandit literature.

4.3.1   Pareto-prior based Myopic and Knowledge-Gradient Strategies

Similar to Chapter 3, we assume that the underlying valuation distribution is known to the seller to be uniform on $[0, Z]$ where Z is finite but she is unaware of the value of Z. Since the underlying valuation distribution is known to be uniform, we model the seller as maintaining a Pareto prior on Z, the unknown parameter of the uniform distribution. The Pareto is the conjugate prior for the uniform in this case. The seller's beliefs are then fully represented

by the two parameters (a and b) of the Pareto distribution:

$$f_Z(x; a, b) = \frac{ab^a}{x^{a+1}} \quad x > b$$

The buyer buys In Chapter 3 we assumed that when the buyer buys, she also reveals her true valuation. Due to that assumption, the posterior turned out to be Pareto. However, in this case the posterior is no longer Pareto.

$$f_Z(x|\text{Buy}) = \begin{cases} \frac{(1-\frac{q}{x})f_Z(x)}{\Pr(\text{Buy}|q)} & x > \max(q, b) \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

The buyer does not buy The updates in this case are similar to the partially censored case discussed in Chapter 3. The posterior is Pareto ($f_Z(x; a+1, b)$) if $b > q$ otherwise not.

$$\begin{aligned} f_Z(x|\neg\text{Buy}) &= \frac{f_Z(x)\Pr(\neg\text{Buy}|x)}{\Pr(\neg\text{Buy})} \\ &= \begin{cases} \frac{qf_Z(x)}{x\Pr(\neg\text{Buy}|q)} & x \geq q \\ \frac{f_Z(x)}{\Pr(\neg\text{Buy}|q)} & x < q \end{cases} \end{aligned} \tag{4.2}$$

The problem with learning from censored information is that the posterior distribution does not have the same distribution type as the prior. One way of tackling this problem is to find a distribution which has the same form as prior and also approximates the true posterior distribution well [23, 79, 32]. We consider the following two approaches of finding approximate posterior distribution:

- **Moment matching based approximation:** In this case, the approximate posterior distribution is calculated such that the first two moments about mean of the approximate distribution and the true posterior distribution are equal [32, 23]. We use this approach in Chapter 4 for approximating a posterior distribution to a Beta distribution [23].

- **Density projection method:** We find an approximate distribution with the same form as the prior such that the KL divergence of the approximate distribution from the true distribution is minimized [79]. This technique is also often used for model evaluation [20].

Let $f_{\text{tp}}(x)$ represent the true posterior distribution (either $f_Z(x|\text{Buy})$ or $f_Z(x|\neg\text{Buy})$) and $f_p(x; a', b')$ represent a Pareto distribution. We want to find a' and b' such that $f_p(x; a', b')$ approximate the true posterior distribution $f_{\text{tp}}(x)$. We now turn to exploring these two techniques for our setting.

Moment matching based approximation

Let $\mathbb{E}_{\text{tp}}(Z)$ and $\mathbb{E}_{\text{tp}}(Z^2)$ represent the first two moments of the true posterior distribution. In this case we want to find (a', b') of the Pareto distribution $f_p(x; a', b')$ such that its first two moments (or, equivalently, the mean and the variance) are the same as that of the posterior distribution:

$$\frac{a'b'}{a'-1} = \mathbb{E}_{\text{tp}}(Z)$$

$$\frac{a'b'^2}{a'-2} = \mathbb{E}_{\text{tp}}(Z^2)$$

$$\Rightarrow a' = 1 + \sqrt{\frac{\mathbb{E}_{\text{tp}}(Z^2)}{\mathbb{E}_{\text{tp}}(Z^2) - (\mathbb{E}_{\text{tp}}(Z))^2}} \quad b' = \frac{(a'-1)\mathbb{E}_{\text{tp}}(Z)}{a'}$$

When the buyer decides to buy, $\mathbb{E}_{\text{tp}}(Z)$ and $\mathbb{E}_{\text{tp}}(Z^2)$ can be computed as follows:

$$\mathbb{E}_{\text{tp}}(Z) = \int_{\max(q,b)}^{\infty} x \frac{(1-\frac{q}{x})f(x)}{\Pr(\text{Buy}|q)} \, dx$$

$$= \frac{1}{\Pr(\text{Buy}|q)} \left(\left(1 - F_Z(\max(q,b); a-1, b)\right)\frac{ab}{a-1} - q\left(1 - F_Z(\max(q,b); a, b)\right) \right)$$

$$= \frac{1}{\Pr(\text{Buy}|q)} \left(\left(\frac{b}{\max(q,b)}\right)^{a-1}\frac{ab}{a-1} - q\left(\frac{b}{\max(q,b)}\right)^{a} \right) = \frac{\frac{a\max(q,b)}{a-1} - q}{1 - \frac{aq}{\max(q,b)(a+1)}}$$

$$\mathbb{E}_{\text{tp}}(Z^2) = \int_{\max(q,b)}^{\infty} x^2 \frac{(1-\frac{q}{x})f(x)}{\Pr(\text{Buy}|q)} \, dx$$

$$= \frac{1}{\Pr(\text{Buy}|q)} \left(\left(1 - F_Z(\max(q,b); a-2, b)\right)\frac{ab^2}{a-2} - \frac{abq}{a-1}\left(1 - F_Z(\max(q,b); a-1, b)\right) \right)$$

$$= \frac{1}{\Pr(\text{Buy}|q)} \left(\left(\frac{b}{\max(q,b)}\right)^{a-2}\frac{ab^2}{a-2} - \frac{aqb}{a-1}\left(\frac{b}{\max(q,b)}\right)^{a-1} \right)$$

$$= \frac{a\max(q,b)\left(\frac{\max(q,b)}{a-2} - \frac{q}{a-1}\right)}{1 - \frac{aq}{\max(q,b)(a+1)}}$$

The following expression show how $\mathbb{E}_{\text{tp}}(Z)$ and $\mathbb{E}_{\text{tp}}(Z^2)$ can be calculated when the buyer decides to not buy:

$$\mathbb{E}_{\text{tp}}(Z) = \int_b^{\max(q,b)} \frac{x f_Z(x)}{\Pr(\neg\text{Buy}|q)} \, dx + \int_{\max(q,b)}^{\infty} \frac{q x f_Z(x)}{x \Pr(\neg\text{Buy}|q)} \, dx$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q)} \left(\frac{ab}{a-1} F_Z(\max(q,b); a-1, b) + q\big(1 - F_Z(\max(q,b); a, b)\big) \right)$$

$$\mathbb{E}_{\text{tp}}(Z^2) = \int_b^{\max(q,b)} \frac{x^2 f_Z(x)}{\Pr(\neg\text{Buy}|q)} \, dx + \int_{\max(q,b)}^{\infty} \frac{q x^2 f_Z(x)}{x \Pr(\neg\text{Buy}|q)} \, dx$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q)} \left(\frac{ab^2}{a-2} F_Z(\max(q,b); a-2, b) + \frac{aqb}{a-1} \big(1 - F_Z(\max(q,b); a-1, b)\big) \right)$$

Posterior approximation using density projection

We find the approximate posterior distribution such that the KL divergence from true distribution is minimized where the relative entropy or KL divergence measures the discrepency between two distributions. According to Cover and Thomas [29]:

> The relative entropy $D(p||q)$ is the measure of inefficiency of assuming that the distribution is q when the true distribution is p.

Mathematically:

$$D(p||q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) \, dx$$

In our case, $p(x)$, the true distribution, is the exact posterior distribution $f_{\text{tp}}(x)$ (calculated using Bayes rule in Equations 4.1 and 4.2 above), and $q(x)$ is the Pareto distribution, $f_p(x; a', b')$, which we want to use as a proxy to the true distribution in order to characterize the belief state of the seller. In this technique, we want to find parameters (a', b') of the Pareto distribution such that the KL divergence $D(f_{\text{tp}}||f_p)$ is minimized, i.e., the approximation is closer to the true distribution. Here, minimizing KL divergence is equivalent to minimizing the cross-entropy because the entropy of the true posterior is constant:

$$D(f_{\text{tp}}||f_p) = \int_{-\infty}^{\infty} f_{\text{tp}}(x; a', b') \log(f_{\text{tp}}(x)) \, dx - \int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(f_p(x; a', b')) \, dx$$

$$= -H(f_{\text{tp}}) + H(f_{\text{tp}}, f_p)$$

In the above equation, $H(f_{\text{tp}})$ is the entropy of the true posterior distribution, which is a constant, and $H(f_{\text{tp}}, f_p)$ is the cross entropy. So we will use cross entropy instead to find the parameters of the approximate distribution.

When the buyer buys then for the cross-entropy to be finite $b' \leq \max(q, b)$; similarly, in case of a no buy $b' \leq b$.

$$
\begin{aligned}
H(f_{\text{tp}}, f_p) &= -\int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(f_p(x; a', b')) \, dx \\
&= -\int_{-\infty}^{\infty} f_{\text{tp}}(x) \big(\log(a') + a' \log(b') - (a' + 1) \log(x) \big) \, dx \\
&= -(\log(a') + a' \log(b')) + (a' + 1) \int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(x) \, dx
\end{aligned}
$$

Clearly $H(f_{\text{tp}}, f_p)$ is decreasing in b'. Thus b' is equal to $\max(q, b)$ when the buyer decides to buy and b otherwise. To find the optimal value of a', we calculate the first derivative and equate it to zero. We then compute the second derivative and verify that the condition for minimum is satisfied.

$$
\begin{aligned}
H(f_{\text{tp}}, f_p) &= -(\log(a') + a' \log(b')) + (a' + 1) \int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(x) \, dx \\
\frac{dH(f_{\text{tp}}, f_p)}{da'} &= -\frac{1}{a'} - \log(b') + \int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(x) \, dx \\
a' &= \frac{1}{\int_{-\infty}^{\infty} f_{\text{tp}}(x) \log(x) \, dx - \log(b')} = \frac{1}{\mathbb{E}_{\text{tp}}(\log(x)) - \log(b')} \\
\frac{d^2 H(f_{\text{tp}}, f_p)}{da'^2} &= \frac{1}{a'^2} > 0 \quad \text{condition for minimum is satisfied}
\end{aligned}
$$

Now we calculate the specific values of the a' parameter of the approximate Pareto distribution for both the buy and no buy cases:

The buyer buys

$$
\begin{aligned}
b' &= \max(q, b) \\
a' &= \frac{1}{\int_{\max(q,b)}^{\infty} f_Z(x|\text{Buy}) \log(x) \, dx - \log(\max(q, b))} \\
a' &= a \left(\frac{1 - \frac{aq}{\max(q,b)(a+1)}}{1 - \frac{qa^2}{\max(q,b)(a+1)^2}} \right)
\end{aligned}
$$

From the update equations, we can see that $a' < a$ when the buyer buys. This can also lead to $a' < 1$. However, we assume that the Z parameter has a finite mean but the mean of

the posterior distribution will not finite if $a' < 1$. To avoid this we artificially constrain the value of a' to be greater than 1 (i.e., whenever $a' < 1$ according to the update, we set it to a value greater than 1). While doing moment-matching, the a parameter of the posterior distribution is always greater than 2 (because the first two moments are finite only if $a > 2$) so we do not have to put such artificial constraints.

The buyer does not buy   In this case, the posterior is already Pareto when $b > q$ with parameters $a' = a + 1$ and $b' = b$, so we only consider the situation when $b < q$:

$$b' = b$$

$$a' = \frac{1}{\int_b^q \frac{\log(x) f_Z(x)}{\Pr(\neg Buy|q)} dx + \int_q^\infty \frac{q \log(x) f_Z(x)}{x \Pr(\neg Buy|q)} dx - \log(b)}$$

$$= a \left(\frac{1 - \frac{1}{a+1} \left(\frac{b}{q}\right)^a}{1 - \frac{2a+1}{(a+1)^2} \left(\frac{b}{q}\right)^a + \log(\frac{b}{q}) \left(\frac{b}{q}\right)^a \frac{1}{a+1}} \right)$$

Here, $a' > a$ so as long as $a > 1$, we do not have to put any artificial constraints.

Calculate of myopic prices in the completely censored case   The myopic price is only dependent on the current state parameters hence it is same for the completely and the partially censored case. From Equation 3.1, if the state parameters are a and b then the myopic price is $\frac{b(a+1)}{2a}$. Since, the posterior distribution is not a Pareto, on observing the buyer's decision, the true posterior is approximated to a Pareto using either of the two algorithms discussed for approximating the posterior.

Calculation of KG optimal prices in the completely censored case   In order to compute the KG optimal price at any time t, we first have to computer the myopic profit at time t + 1. For doing so, we first calculate the approximate posterior distribution using either of the two methods discussed above. Then the myopic profit can be calculated using Equation 3.1. The KG price can then be calculated by maximizing the expected one-step-look-ahead profit using Equation 3.3.

4.3.2   The LLVD Algorithm

This algorithm is also based on the assumption that the underlying valuation distribution is uniform. This implies that the probability of an arriving buyer choosing to go through with a purchase at quoted price q is a linear function of q, $\Pr(\text{Buy}|q, \gamma) = 1 - \gamma q$ (for $q < \frac{1}{\gamma}$).

This forms the basis for our learning algorithm, which we call "Linear Learning of Valuation Distributions" (LLVD).

Under the linearity assumption we want to maximize total expected (discounted) revenue. The seller's state space is now the space of distributions over γ. In order to make this a tractable state space to work with, we enforce that the seller always represents her beliefs as a Beta distribution ($\gamma \in [0, 1]$). The state space can then be parametrized by the two parameters of the Beta distribution. We need to derive the state space transition model and the reward model in order to solve for the seller's optimal policy. In the following, $f(\gamma; \alpha, \beta)$ represents the density function for the Beta distribution, $F(\gamma; \alpha, \beta)$ represents the cdf for the Beta distribution, and $F_k(\gamma)$ represents $F(\gamma; \alpha + k, \beta)$.

Transition Model An arriving buyer is quoted a price q and decides whether or not to buy at that price. She will buy if her valuation is less than or equal to the price quoted. The seller updates her own distribution over γ based on whether or not the arriving buyer bought the good. Consider the Bayesian updates in two cases:

1. Buyer does not buy:

$$P(\neg\text{Buy}|\gamma, q) = \begin{cases} \gamma q & \text{for } q < 1/\gamma \\ 1 & \text{otherwise} \end{cases}$$

$$f(\gamma|\neg\text{Buy}) = \frac{f(\gamma; \alpha, \beta)P(\neg\text{Buy}|q, \gamma)}{qF_1(1/q) + 1 - F_0(1/q)}$$

For $q < 1$, the normalizing constant is 1 and the true posterior is Beta. When $q > 1$ the posterior need not be Beta, so we approximate the posterior by a Beta distribution, $f_b(x; \alpha', \beta')$, using the moment-matching based approximation discussed in Section 4.3.1 (i.e., find α' and β' such that the first and second moment of the true posterior and the approximate Beta distribution are same). This yields a pair of simultaneous equations for α' and β':

$$\frac{\alpha'}{\alpha' + \beta'} = \frac{q\mathbb{E}(\gamma^2)F_2(1/q) + \mathbb{E}(\gamma)(1 - F_1(1/q))}{(q\mathbb{E}(\gamma)F_1(1/q) + 1 - F_0(1/q))}$$

$$\frac{\alpha'(\alpha' + 1)}{(\alpha' + \beta')(\alpha' + \beta' + 1)} = \frac{q\mathbb{E}(\gamma^3)F_3(1/q) + \mathbb{E}(\gamma^2)(1 - F_2(1/q))}{(q\mathbb{E}(\gamma)F_1(1/q) + 1 - F_0(1/q))}$$

2. Buyer buys:

$$f(\gamma|\text{Buy}) = \frac{f(\gamma; \alpha, \beta)(1 - \gamma q)}{\int_0^{1/q} f(\gamma; \alpha, \beta)(1 - \gamma q)d\gamma}$$

$$= \frac{f(\gamma; \alpha, \beta)(1 - \gamma q)}{(F_0(1/q) - q\mathbb{E}(\gamma)F_1(1/q))}$$

$$= \frac{f(\gamma; \alpha, \beta)(1 - \gamma q)}{(F_0(1/q) - q\mathbb{E}(\gamma)F_1(1/q))}$$

Again, we approximate the true posterior with a Beta distribution by matching the first and second moments.

$$\frac{\alpha'}{\alpha' + \beta'} = \frac{\mathbb{E}(\gamma)F_1(1/q) - qF_2(1/q)\mathbb{E}(\gamma^2)}{F_0(1/q) - q\mathbb{E}(\gamma)F_1(1/q)}$$

$$\frac{\alpha'(\alpha' + 1)}{(\alpha' + \beta')(\alpha' + \beta' + 1)} = \frac{\mathbb{E}(\gamma^2)F_2(1/q) - q\mathbb{E}(\gamma^3)F_3(1/q)}{F_0(1/q) - q\mathbb{E}(\gamma)F_1(1/q)}$$

Let M and S represent first and second order moments respectively. Solving these equations yields update rules $\alpha' = \frac{MS - M^2}{M^2 - S}$ and $\beta' = \frac{(1 - M)\alpha'}{M}$.

Reward Model Let π denote the discounted long-term revenue and δ the discount factor. Let $P(q) = \Pr(\text{Buy}|q)$. Then $\pi = q_0 P(q_0) + \sum_{t=1}^{\infty} \delta^t q_t P(q_t)$. The first term, $\pi_0 = q_0 P(q_0)$ is the expected reward at this particular instant, from the next action. We can compute the expected value of this term:

$$P(q) = \int_0^{1/q} (1 - \gamma q)f(\gamma; \alpha, \beta)d\gamma$$

$$= F(1/q; \alpha, \beta) - q\mathbb{E}(\gamma)F(1/q; \alpha + 1, \beta)$$

$$= F(1/q; \alpha, \beta) - q\mu F(1/q, \alpha + 1, \beta) \tag{4.3}$$

where $\mu = \alpha/(\alpha + \beta)$.

$$\pi_0 = q_0(F(1/q_0; \alpha, \beta) - q_0\mathbb{E}(\gamma)F(1/q_0; \alpha + 1, \beta))$$

$$= q_0(F(1/q_0; \alpha, \beta) - q_0\mu F(1/q_0, \alpha + 1, \beta)) \tag{4.4}$$

The Bellman Equation In a risk-neutral framework, we can similarly take expectations over γ and derive the appropriate Bellman equation:
$V(\alpha_t, \beta_t) = \max_q (qP(q) + \delta V')$, where α_t and β_t represent the seller's belief at time t and

$$V' = P(q)V(\alpha_{t+1}, \beta_{t+1}|\text{Buy}) + (1 - P(q))V(\alpha_{t+1}, \beta_{t+1}|\neg\text{Buy})$$

Obviously, if γ were known to the seller, the optimal action would be the optimal myopic action, and it would yield a discounted expected revenue of:

$$\pi = \max_q (q(1 - \gamma q) + \sum_{t=1}^{\infty} \delta^t q(1 - \gamma q))$$

$$= \max_q \frac{q(1 - \gamma q)}{1 - \delta} = \max_q \frac{q(1 - \gamma q)}{1 - \delta} \tag{4.5}$$

This equation is maximized at $q = \frac{1}{2\gamma}$, in our environment, yielding $V = \frac{1}{4\gamma(1-\delta)}$.

Solving for the optimal policy   Various issues arise in trying to solve such a system. A value-iteration type method would rely on a reasonable functional approximation of the value function in order to converge to a correct estimate. We use a different approach by first restricting the problem to a space where table-based value iteration can be applied, and then extrapolating to the complete space. We start by restricting to values of q between 0 and 1.

The $q < 1$ case: Equation 4.3 reduces to $P(q) = (1 - \mu q)$, therefore Equation 4.4 reduces to $\pi_0 = q_0(1 - \mu q_0)$ because $F(1/q) = 1$ for the Beta distribution as $q < 1$. Equation 4.5 is maximized at $q = \min(1, \frac{1}{2\mu})$, in our environment, yielding $V = \min(\frac{1}{4\mu(1-\delta)}, \frac{1-\mu}{1-\delta})$. Since the transition model is known (the fact that the true posterior is Beta when the buyer buys for $q < 1$ is helpful in efficient implementation), all that remains in order to discretize and apply value iteration is to specify some boundary conditions on the model. The boundary conditions correspond to having a high degree of certainty about the value of γ. We assume that when the variance of the Beta distribution becomes less than 0.001, γ can be assumed to be known to the seller, and it is then equal to μ. In order for this technique to be consistent, we need to show that once the variance is sufficiently low, it will not be the case that it again starts increasing. We can show that in expectation the variance decreases in every iteration for $q < 1$; the proof is omitted due to space considerations.

This yields the final algorithm: we use value iteration to solve for the value function on a grid for $\alpha, \beta \in [0.1, 200]$, but we pre-fill all spaces where α, β are such that the variance of the distribution is less than 0.001. Figure 4.2 (V1) shows the value function for $\delta = 0.95$, as a function of α and β.

Extending to $q > 1$:   We expect the value function computed using table-based value iteration to closely approximate the universally "correct" one for regions where the optimal value of q is less than 1. Therefore, we fit a regression line using values from the value function matrix where $\mu > 0.6$ (implying that the optimal q is probably lower than 0.85). Empirically, we find that the value function is close to linear in $\frac{1}{\mu}$ and $\frac{1}{\alpha+\beta}$ (see Figure 4.1). So we approximate the value function for the whole space as:
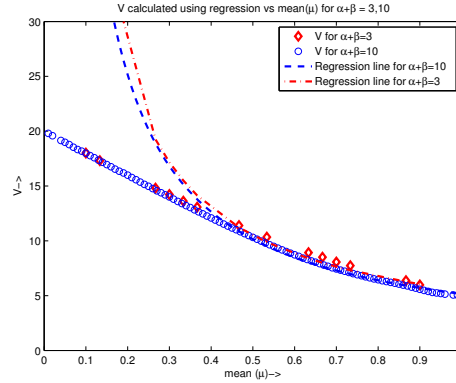
Figure 4.1: Comparison of the regression line with data from the value iteration table for different values of $\alpha + \beta$. Note the very tight match in the domain where the optimal q would be expected to be less than 1. The regression function allows LLVD to generalize this to the entire space (notice the difference between the line and the data points for lower values of μ, which correspond to higher optimal values of q).

$$V(\alpha, \beta) = a_1 \frac{\alpha + \beta}{\alpha} + a_2 \frac{1}{\alpha + \beta} \tag{4.6}$$

Figure 4.2 shows that this is a good approximation over the entire space. Now, at any time $T = t$, with the belief state (α, β) we can find the q which maximizes the given equation.

$$\pi = \max_{q_t} \left(V(\alpha, \beta) + \delta(\Pr(\text{Buy}|q_t)V(\alpha', \beta'|\text{Buy}) + (1 - \Pr(\text{Buy}|q_t))V(\alpha'', \beta''|\neg\text{Buy})) \right)$$

Here α', β', α'' and β'' are functions of q_t,α and β, price offered at time T=t. These values can be calculated as discussed above by comparing the first two moments.

Implementation notes: In our experiments, we compute the value function using $\delta = 0.95$. The best fit regression line is obtained for $a_1 = 4.99$ and $a_2 = 1.5147$; for convenience we use $a_1 = 5$ and $a_2 = 1.5$. The LLVD based seller then learns online, constantly updating her belief on γ, and choosing the price that maximizes the value function at any instant.

4.3.3 Bandit Schemes

Multi-armed bandit algorithms are often applied to Dynamic pricing [81]. The different pricing options are the arms of the bandit and the goal is to find the arm that maximizes infinite horizon discounted reward. The downside of such approaches is that one needs to have fixed arms, and there is no "information sharing" between arms. While reasonable for evaluation, there may be situations where the need to find a reasonable interval is a downside for bandit-based methods. We discuss two algorithms.
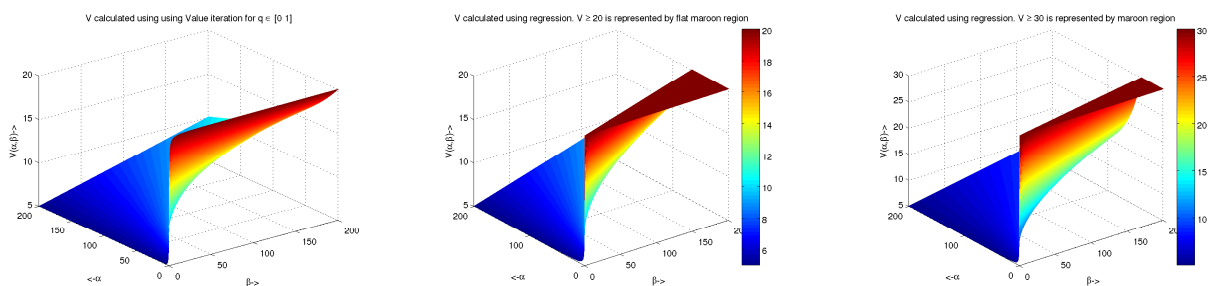
V1: Table-based value function

V2: Value function using regression

V3: Value function extrapolated using regression

Figure 4.2: V1 is the value function computed using table-based value iteration with $q < 1$ (the maximum value of V is 20). V2 is the value function computed using regression (see Equation 4.6), showing the similarity to V1 where the value function is less than 20 (the flat maroon region shows where $V \geq 20$, where the value functions would be expected to differ, and $q > 1$). V3 shows some more of the structure of the value function computed using regression (Equation 4.6) in the region where it attains values between 20 and 30.

A Gittins Index Scheme With Smart Priors We discussed the standard Gittins index scheme in detail in Chapter 3. In this chapter we discuss the modification we propose to the standard Gittins strategy by exploiting the information structure of the problem. The standard approach of initializing all the arms with the same prior is inappropriate in this case because we know that the probability of a buyer buying at a higher price is lower. Thus we arrange the arms in increasing order of their weights and divide them in 4 region. We initialize arms in the region with lowest weight with a Beta $(4, 1)$ prior, the next lowest with a Beta $(3, 2)$ prior, next with $(2, 3)$ and the remaining with $(1, 4)$. As expected, this weighting of the priors significantly outperforms uniform priors on all the arms. Table 4.1 shows the final algorithm in detail.

UCB1 Much work on digital goods auctions has focused on algorithms with good regret bounds. Two of these that are based on algorithms for multi-armed bandit problems have gained particular attention, namely the EXP3 algorithm [2, 15] and the UCB1 algorithm [3, 58]. Kleinberg discusses a "continuum armed" bandit algorithm called CAB1, which is a wrapper around algorithms like UCB1 or EXP3 for continuous spaces [57]. We perform extensive empirical tests on all these algorithms, adapted to our setting. UCB1 and EXP3 discretize the action space and treat each possible price as a unique possible action (or "arm" in bandit language). The EXP3 and UCB1 algorithms are specifically designed for adversarial and i.i.d. scenarios respectively. As expected, we find that EXP3 is outperformed (or equaled in performance) by UCB1 in all our i.i.d. scenarios, so we do not report results from EXP3. While one would expect CAB1 to perform well, since it is designed for continuous action spaces, it is geared more towards producing useful regret bounds, and does not take advantage of the structure of the search space, instead using doubling processes to efficiently scan a

Table 4.1: **A Gittins-Index Based Algorithm**. The K parameter governs the discretization of the space (we use $K = 20$).

Parameters: Price $Q \in [q_{\min}, q_{\max}]^K$, Matrix G of Gittins Indices.

Initialization: $n = 0$ (# buyers so far) , Divide Q in 4 regions in increasing order of magnitude. Initialize state S for each of the K arms according to the region they lie in: from lower to higher: (4,1),(3,2),(2,3),(1,4)

For each k in Buyers do:

1. Price the item at Q_j which maximizes $Q_j.G[S_j]$. Denote the chosen price by Q_{j*}.

2. If the buyer buys, set $S_j(a) = S_j(a) + 1$ else set $S_j(b) = S_j(b) + 1$

---

potentially large continuum. It is outperformed by UCB1. The specific form of the UCB1 algorithm we use is shown in Table 4.2.

4.4 Experimental Results

4.4.1 Synthetic Datasets

We consider various different distributions that generate demand. We restrict ourselves to i.i.d. assumptions rather than considering adversarial scenarios.

Choice of distributions We consider three sets of valuation distributions that generate a wide range of optimal prices:

1. Uniform on $[0, Z]$ where Z is $1.5, 2.5, 4$.

2. Exponential with rate (λ) parameters $1.75, 0.8, 0.5$.

3. Log-normal with location (μ) and scale (σ) parameters $(1, 1)$, $(1, 0.75)$ and $(1, 0.5)$.

Table 4.2: **Algorithm UCB1**, adapted to our setting. The K parameter governs the discretization of the space (we use $K = 20$).

Parameters: Price $Q \in [q_{\min}, q_{\text{high}}]^K$, Number of buyers: *nob*.
Initialization: $n = 0$ (# buyers so far)

For each k in first K buyers do:

1. Price the item at Q_k

2. $n_k = 1; n = n + 1$

3. If the buyer buys then $x_k = Q_k$ else $x_k = 0$

For the remaining buyers at each time instant t do:

1. Price the item at Q_j which maximizes $\frac{x_j}{n_j} + \sqrt{\frac{2 \ln n}{n_j}}$. Denote the chosen price by Q_{j*}.

2. $n_{j*} = n_{j*} + 1; n = n + 1$

3. If the buyer buys, set $x_{j*} = x_{j*} + Q_{j*}$ and update total profit

---

Experimental settings   For these artificial datasets while using multi-armed bandit based algorithms (Gittins and UCB1) we discretize the interval from $[0.5, 2q^*]$ uniformly into 20 steps, where q^* is the analytically computed optimal price for the specific valuation distribution. The initial parameters for the Beta prior in LLVD algorithm are set to $\alpha = 1.5, \beta = 1.5$. For the Pareto-prior in the KG strategy, the starting parameters are chosen to be $a = 2.1$ and $b = 2.3$. We chose these values so that both the KG and the LLVD choose the same first price; this implies that we inject the same initial information into both strategies and enable us to compare them fairly.

Analysis of Results   Each simulation consists of a stream of n buyers, arriving one after the other, each buyer having a valuation v that is sampled at random from the valuation distribution. The seller chooses a price q to offer, and if $v \geq q$ the buyer goes through with

U1: Uniform ($B = 1.5$)   U2: Uniform ($B = 2.5$)   U3: Uniform ($B = 4$)

E1: Exponential ($\lambda = 1.75$)   E2: Exponential ($\lambda = 0.8$)   E3: Exponential ($\lambda = 0.5$)

L1: Log-normal ($\mu = 1, \sigma = 1$)   L2: Log-normal ($\mu = 1, \sigma = 0.75$)   L3: Log-normal ($\mu = 1, \sigma = 0.5$)
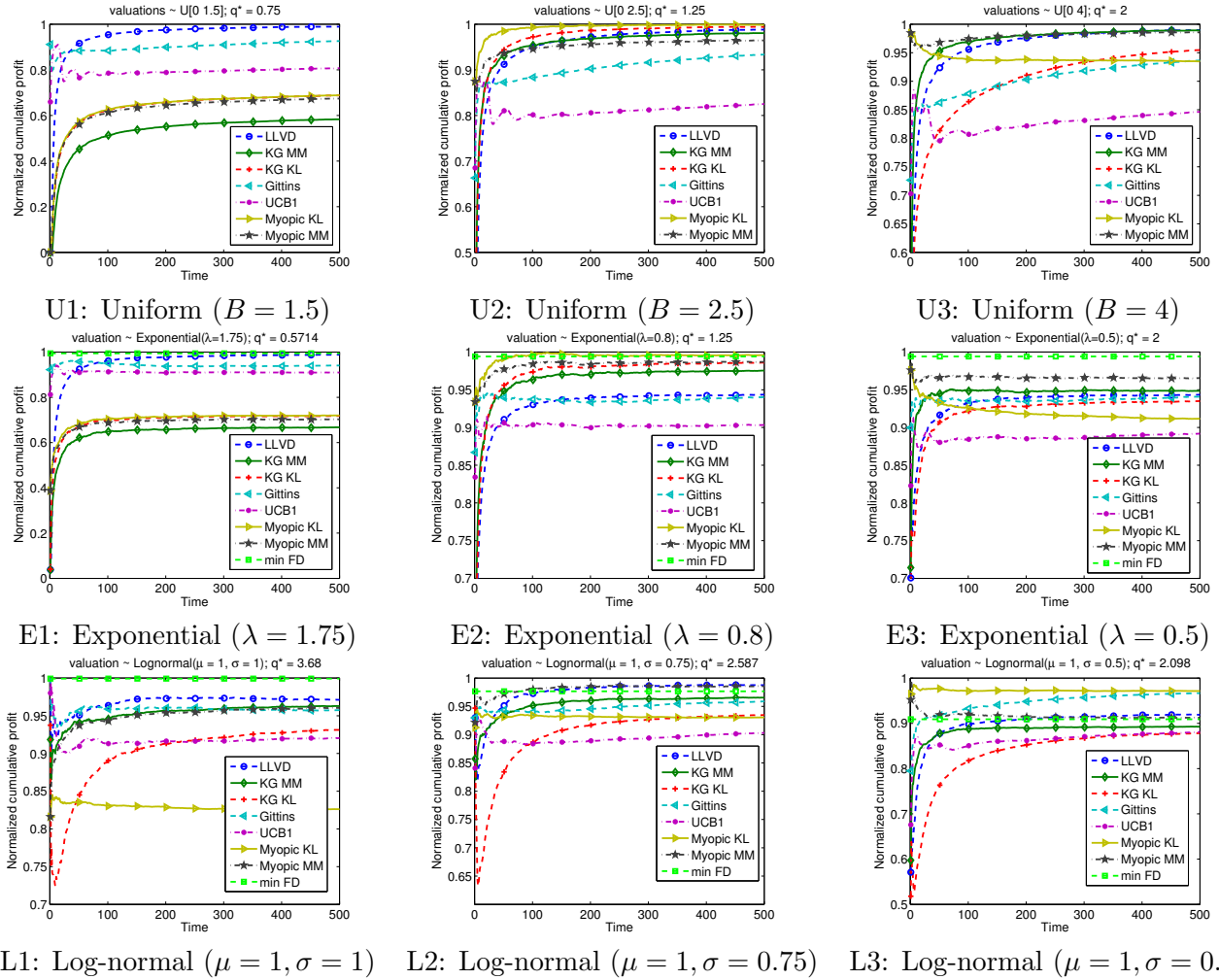
Figure 4.3: In the graphs above, the suffix MM and KL represents that the moment-matching and the KL divergence based approximation has been used respectively. Each graph shows the time-averaged profit received at any time-step, averaged over 1000 simulations. The top row shows uniform valuation distributions, corresponding to the model used by LLVD, KG based and the myopic algorithms. The second row shows exponential valuation distributions, and the bottom row log-normal distributions. All values are represented as fraction of optimal profit.

the purchase, otherwise she turns down the offer. In Figure 4.3 we report results averaged over 1000 simulations of the process, each consisting of 500 time steps.

In addition to comparing the algorithms, in cases where the linearity assumption is violated (exponential and log-normal valuation distributions), we are interested in quantifying how much of the regret of the algorithm can be attributed to the linearity assumption itself,

and how much may be due to not learning the best possible linear function. In order to study this, we also report the analytical profit that would be achieved by using the linear function of the form $1 - \gamma q$ to model the probability of buying, when γ is chosen so that the functional distance between the uniform distribution on $[0, 1/\gamma]$ and the true target valuation distribution is minimized. We evaluate functional distance between the two distributions as the squared L^2 metric between their cumulative distribution function (cdf). Let $F(x)$ and $G(x)$ be the cdf of two distributions

$$f_d = L^2 \text{ metric between } F \text{ and } G = \sqrt{\int_0^\infty (F(x) - G(x))^2 \, dx}$$

In our case where $F(x)$ is the uniform distribution in the interval $[0, Z]$ where $Z = 1/\gamma$.

$$D = f_d{}^2 = \int_0^Z \left(\frac{x}{Z} - G(x)\right)^2 \, dx + \int_Z^\infty (1 - G(x))^2 \, dx$$

Further details on the computation are in Appendix A.2.1. In Figure 4.3, the expected profit corresponding to the minimum functional distance is shown by a green line marked 'min-FD'.

Uniform valuation distributions (linear demand) As expected, the LLVD algorithm performs very well by learning the true distribution rapidly in this case because it is based on a uniform valuation distribution assumption. The Pareto-prior based algorithms (KG and Myopic) perform well when the valuation distributions are $U[0, 2.5]$ or $U[0, 4]$, but not when it is $U[0, 1.5]$. The reason behind this is perhaps unsuitable choice of initial priors as discussed in Chapter 3 (Section 3.4.1). We set the prior parameters to $a_0 = 2.1$ and $b_0 = 2.3$; the b_0 parameter depicts the minimum possible value of Z because the pdf of the Pareto distribution is zero for any value less than b_0. By setting this prior, we are informing our algorithm that $Z \geq 2.3$, therefore the KG-based algorithm does not charge a price less than $b_0/2 = 1.15$. However, the optimal price in this case is 0.75 and thus the algorithms based on the Pareto-prior do not perform as well as they performed in Figure 4.3 (U2) and (U3). The other surprising result is that myopic pricing algorithms have outperforms both LLVD and the KG based strategy in some of the cases.

Exponential valuation distributions In this case, $\Pr(\text{Buyer Buys}|q) = e^{-\lambda q}$, where λ is the rate parameter. LLVD performs either better than or as well as the Gittins-index based scheme in these cases, and significantly outperforms UCB1. The knowledge gradient based algorithm which uses moment-matching based approximation clearly outperformed all the algorithms for $\lambda = 0.8, 0.5$. However, for $\lambda = 1.75$, the knowledge gradient based algorithms do not perform well. The reason is similar to the one we discussed for the uniform distribution with $Z = 1.5$. The optimal price in this case is 0.5714, and the b_0 parameter is set too high so it cannot learn the correct price. Again, in this case, the myopic strategy sometimes outperforms LLVD and KG based algorithms.

Log-normal valuation distributions For the log-normal distribution, $\Pr(\text{Buyer Buys}|q) = 1 - \Phi(\frac{\ln q - \mu}{\sigma})$, where μ and σ are the location and scale parameters for the said distribution.

While LLVD dominates UCB1, the Gittins-index based scheme is competitive, sometimes performing better and sometimes worse. The KG based strategy using moment-matching based approximation performs slightly worse than LLVD but always dominates UCB1. In the beginning, UCB1 outperforms the other KG based strategy; however, the performance of the other KG based strategy increases rapidly with time and after a while, either performs better than or as well as the UCB1 strategy. The myopic algorithm based on moment-matching based posterior approximation either outperforms or performs as well as its KG counterpart. Surprisingly, in some cases the linear demand based algorithms even significantly outperform the "best" linear function (indicating that the fit over the entire distribution is not necessarily the best measure when profit-seeking behavior is determined by only a portion of the distribution).

4.4.2 Real World Datasets

Now, we want to test how these approaches fare in case of a real-world dataset. In order to do so we use three datasets (eBay auction dataset, Jester dataset, Yahoo! keyword bids data) which have been described in detail in Chapter 3. Figure 4.4 shows the results for the three different distributions; the x and y axes are similar to those in Figure 4.3. For the eBay auction dataset and the Yahoo! keyword bids, we generate 1000 random permutations and average results over them. For the jokes rating dataset from Jester we have more than 16000 ratings, so we instead generate 1000 samples of 500 unique users in each case. For the bandit-based algorithms we select 20 prices corresponding to 20 arms of the bandit between the minimum and the maximum values of the the specific dataset uniformly.

In all three cases, the KG based technique using moment-matching based approximation either dominates or performs almost as well as other algorithms. On the other hand, the other KG based strategy does not perform well in most cases. LLVD performs similarly or slightly better than the KG based technique (with moment-matching) for the Jester and the Yahoo! keywords data but is slow to converge for the eBay auctions data. Performance of the Gittins-index based strategy using smart priors is good for the eBay auction dataset and the jokes rating dataset from Jester, but surprisingly, for the Yahoo! advertisers' bids data is very unimpressive. The performance of UCB1 keeps improving over time and converges to that of other better strategies in all three cases. The myopic strategy based on approximating the posterior by minimizing the KL divergence performs the worst, although its performance (both in relative and absolute terms) was better in the synthetic data. The reason behind this is that while updating the beliefs using the myopic price, the b parameter never gets updated. When the buyer decides to buy, the updated $b' = \max(q, b) = b$ as $q = \frac{b(a+1)}{2a} < b$, and when the buyer decides to not buy, the updated $b' = b$. Therefore, the performance of this strategy depends on the initial value of the b parameter which is set by the seller at the beginning of the algorithm, and is not very adaptive to the environment.
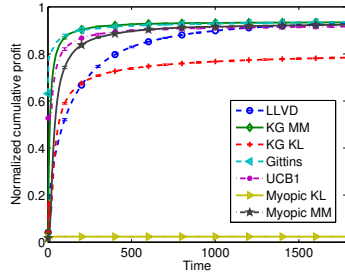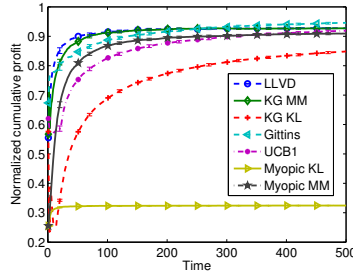
(a) eBay auction dataset   (b) Jester rating dataset   (c) Yahoo! keyword bidding data
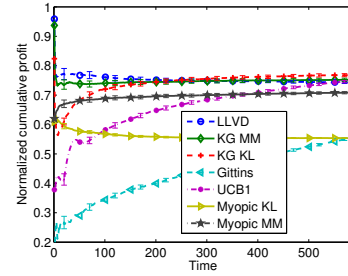
Figure 4.4: Comparison of various strategies on 3 real-world datasets. LLVD and a knowledge gradient based algorithm using moment-matching based approximation perform consistently well in these cases even when the underlying uniform distribution assumption on the valuation data has been violated. We constrain $a \geq 1.01$ in case of Pareto distribution, so the seller's belief distribution always has a finite mean. The initial parameters for the Pareto distribution are $a_0 = 2.1, b_0 = 2.3$ and discount factor $\delta = 0.95$ and for LLVD $\alpha_0 = \beta_0 = 1.5$.

Aggregate analysis Figure 4.4 illustrates the performance of various strategies for 3 specific examples which helps in understanding the pros and cons of these strategies. To get a deeper understanding, we now perform large-scale analysis similar to the one in Chapter 3 by comparing the average discounted and undiscounted profits to evaluate these pricing algorithms. We use the same ratings datasets from Jester and keyword bids from Yahoo! for this purpose and similar setup which we used in Chapter 3.

For the sake of clarity, the experimental setup has been summarized. We have ratings available for 100 jokes; the number of ratings for each joke depends on the number of users out of a total of 24983 who have rated that joke. For each joke, we first calculate the average discounted and the undiscounted profit by averaging the respective profits in various iterations. In each iteration, a sample of 500 unique users is picked uniformly at random. We count the number of times each algorithm has been the best depending on the goal (discounted profit or undiscounted profit) on this set of 100 jokes. We then calculate the normalized discounted and undiscounted profit for all the jokes by dividing them with the respective expected optimal profit that can be achieved by a single fixed pricing strategy using the prior knowledge of the true distribution. We also maintain a count for each algorithm indicating the number of times it performed best. Similarly, for the Yahoo! dataset we select the top hundred keywords containing the highest number of unique bidders. To avoid outliers we winsorize the bids at the 1st and the 99th percentile for each keyword. The number of bidders is small (between 126 - 588) so we generate 100 random permuations of bids for each keyword instead of sampling (like in the Jester dataset) and average the profits over them.

Table 4.3 shows the average of the normalized discounted profit and the number of times

Table 4.3: Comparison of the average discounted profit achieved by different strategies and the number of times each strategy performs best (#wins) on the ratings of 100 jokes from Jester and the bidding data for 100 keywords from Yahoo! sponsored search auctions.

Strategy	Jester dataset	Yahoo! dataset
	Normalized discounted profit, #wins	Normalized discounted profit, #wins
LLVD	0.8646 ± 0.0037, 74	0.6656 ± 0.0245, 19
KG-MM	0.8322 ± 0.0027, 0	0.6680 ± 0.0219, 22
KG-KL	0.4769 ± 0.0113, 0	0.5914 ± 0.0412, 21
Myopic-MM	0.7195 ± 0.0034, 0	0.6094 ± 0.0321, 6
Myopic-KL	0.3332 ± 0.0101, 0	0.5443 ± 0.0501, 9
Gittins	0.8099 ± 0.0149, 26	0.4778 ± 0.0355, 14
UCB1	0.6792 ± 0.0038, 0	0.5560 ± 0.0192, 9

that algorithm performed the best in terms of the discounted profit using the two datasets for various algorithms; Table 4.4 similarly shows results corresponding to the undiscounted profit. The linear demand based algorithms have outperformed all the algorithms in both discounted and undiscounted settings except in case of the undiscounted profit in the Jester dataset. In that case, Gittins based strategy has outperformed these algorithms by a small margin but it has performed worst of all algorithms both in terms of discounted profit and undiscounted profit for the Yahoo! dataset. In general, the bandit based algorithms do not perform well in terms of average discounted profits compared to the linear demand based strategies in these experiments. The UCB1 strategy has been outperformed by even the Myopic pricing strategy (using moment-matching approximation for the posterior) in both datasets. The Gittins-index based strategy fare no better.

The pricing strategy using KG with moment-matching based approximation for the posterior has been competitive with LLVD. The KG based strategy using KL divergence based approximation does not perform as well as its other linear-demand based counterparts in most cases, and there are two possibilities for this algorithm to not perform well. First is the linear demand assumption is restrictive and possibly does not provide enough flexibility to model real-world valuation distributions. Second is the posterior is not being approximated efficiently leading to undesirable outcomes. Since LLVD and the another KG based strategy perform well which are also based on the linear-demand assumption, it indicates that the latter is the reason for the underperformance. This extensive empirical analysis indicate the efficiency of linear-demand models in estimating the demand even when the true demand is not linear.

Comparing the performance of the algorithms in the completely censored and the partially censored settings using Figures 4.4 and 3.2 (from Chapter 3), we find that surprisingly the linear demand based algorithms in case of the Yahoo! dataset in the completely censored

Table 4.4: Comparison of the average undiscounted profit achieved by different strategies and the number of times each strategy performs best (#wins) on the user ratings for 100 jokes from Jester and the bidding data for 100 keywords from Yahoo! sponsored search auctions.

Strategy	Jester dataset	Yahoo! dataset
	Normalized undiscounted profit, #wins	Normalized undiscounted profit, #wins
LLVD	0.9254 ± 0.0064, 37	0.6626 ± 0.0270, 14
KG-MM	0.9232 ± 0.0051, 0	0.6830 ± 0.0230, 15
KG-KL	0.8369 ± 0.0119, 0	0.6994 ± 0.0356, 31
Myopic-MM	0.9072 ± 0.0064, 0	0.6313 ± 0.0301, 9
Myopic-KL	0.3399 ± 0.01, 0	0.5311 ± 0.0409, 6
Gittins	0.9365 ± 0.0038, 58	0.5465 ± 0.0298, 4
UCB1	0.9200 ± 0.0031, 5	0.6345 ± 0.0227, 21

setting perform better in the longer run and are less affected by the fat-tailed behavior of the bids. In fact, LLVD and KG-MM perform better even in terms of discounted profit on the Yahoo! dataset in the completely censored setting than the myopic algorithm (which is also KG) in the partially censored setting. This is because the absence of valuation information prevents the seller's belief to be significantly affected by the valuations from outliers or the heavy tails of the keyword distributions when the model is misspecified. This makes completely censored information updates more robust to violations of the modeling assumptions.

4.5 Discussion

As dynamic pricing becomes a reality with intelligent agents making rapid pricing decisions on the Internet, the field of algorithmic pricing is developing rapidly. While there has been continuing work on revenue management and inventory issues in operations research, the study of posted price mechanisms for digital goods auctions has mostly been confined to theoretical computer science, inspired by developments from computational learning theory. As a result, the focus has mostly been on deriving regret bounds rather than developing and analyzing algorithms that could prove useful in practice. In the spirit of Vermorel and Mohri's empirical analysis of algorithms for bandit problems [89], we believe that it is important to test algorithms in simulation, and ideally in the real world environments, or at least using real-world data. Since, it is hard to get access the buyers' valuation data in real-world, we chose datasets which very closely resemble the valuation data and thus are reasonable to be used for this problem.

We find that the UCB1 algorithm, which has some desirable theoretical properties for posted price auctions with unlimited supply, can be slow to learn in simple simulated environments. Theoretical extensions to spaces with a continuum of actions, like CAB1, do not perform well either. However, there are three promising directions: (1) algorithms based on making a linearity assumption about the demand curve perform well (LLVD and KG based algorithm using moment-matching based approximation), even when the true model is not linear. We test this claim rigorously on synthetic data as well as on multiple real-world datasets. Our experimental results and theoretical analysis of the linearity assumption indicate that it may be a very useful approximation, far beyond just for truly linear models. (2) Using simple but appropriate priors in a Gittins-index based scheme also shows promise. There is still scope to further improve the performance by enabling better information sharing among arms. (3) Completely censored information can be sometimes useful in cases where the true distribution is far from being uniform (e.g., a heavy tailed distribution) by acting as a regularizer on the posterior so it is not significantly affected by the outliers.

Chapter 5

Pricing Information Services in One-sided Search

5.1 Introduction

In *one-sided search*, agents face a stream of opportunities that arise sequentially, and the process terminates when the agent picks one of those opportunities [67, 70, 47]. A classic example is a consumer looking to buy a used car discussed in Chapter 1. She will typically investigate cars one at a time until deciding upon one she wants. Similar settings can be found in job-search, house search, technology R&D, and other applications [63, 75, 11, 88]. Sequential search theory has also been applied extensively in mathematical biology to model the process of mate search [54, 91]. Search is becoming increasingly important in modern electronic marketplaces, both because of the lowering of search costs, as well as the ability of consumers to turn to price search intermediaries, often artificial agents, to find the best prices or values for items they are interested in acquiring [5, 18, 51].

The decision making complexity in one-sided search arises from the fact that there is a cost incurred in finding out the true value of any opportunity encountered [5, 56]. For example, there is a cost to arranging a meeting to test drive a car you are considering purchasing. The searcher needs to trade off the potential benefit of continuing to search and seeing a possibly more valuable opportunity with the costs incurred in doing so. The optimal stopping rule for such search problems has been widely studied, and is often a *reservation strategy*, where the searcher should terminate search once she encounters an opportunity which has a value above a certain reservation value or threshold [90, 70]. Most models assume that upon encountering an opportunity the searcher obtains its exact true value. However, in many realistic settings, search is inherently noisy and searchers may only obtain a noisy signal of the true value. For example, the drivetrain of a used car may not be in good condition, even if the body of the car looks impressive. The relaxation of the assumption of perfect values not

only changes the optimal strategy for a searcher, it also leads to a niche in the marketplace for new knowledge-brokers. The knowledge brokers, or *experts*, are service providers whose main role is to inform consumers or searchers about the values of opportunities.

An expert offers the searcher the option to obtain a more precise estimate of the value of an opportunity in question, in exchange for the payment of a fee (which covers the cost of providing the service as well as the profit of the expert). To continue with the used-car example, when the agent is intrigued by a particular car and wants to learn more about it, she could take the car to a mechanic who could investigate the car in more detail, or make a repeat visit to see the car, possibly bringing an experienced friend for a more thorough examination. In essence, the searcher obtains more accurate information for an additional cost (either monetary or equivalent).

Sequential-search markets with experts are particularly interesting in the context of modern electronic marketplaces for several reasons. First, there is now a proliferation of not just sellers, as mentioned above, but competing marketplaces that attempt to match buyers and sellers (ranging from highly organized marketplaces like `autotrader.com` to bulletin boards like `craigslist.com`), and design choices affect how these marketplaces perform and serve the needs of customers. Second, the transition to electronic marketplaces has greatly lowered search costs (the monetary, time, and opportunity costs of examining new possibilities), which significantly affects market dynamics. Third, this transition has led to the emergence of new kinds of experts; for example, independent agencies like Carfax now monitor the recorded history of transactions, repairs, claims, etc. on cars, and can, to a certain extent, replace traditional experts like mechanics, performing similar services at a much lower marginal cost of production of an expert report. Fourth, many decisions, including setting prices and deciding when to terminate search, can be taken by automated agents operating in the best interests of different users.

In this chapter we investigate optimal strategies for both searcher and expert in environments where searchers observe noisy signals and can obtain (i.e., query the expert for) the actual values for a fee. There has been some previous work on search with noisy signals in the theoretical biology literature, in the context of mate search [91]. However, this existing work only considers restricted strategy spaces (an opportunity cannot be accepted by the searcher until its exact true value is known; the option of accepting an uncertain offer without paying an additional revelation cost is ignored), and also models the cost of revelation of actual values as fixed, rather than as set by a strategic, self-interested agent. In contrast, we formulate the problem as a Stackelberg game, where the expert makes the first move by setting the terms for its services (e.g., price) and the searcher responds. By characterizing optimal strategies for searchers, we can then solve for the price that maximizes the expert's profit. We can use this analysis to understand the effects of different market interventions, for example, subsidizing the purchase of expert services.

5.2   The Model

Our model considers an agent or *searcher* facing an infinite stream of opportunities from which she needs to choose one. The value v of each opportunity is a priori unknown, however the searcher can receive, at cost c_s, a noisy signal s, correlated with the true value according to a known probability density function $f_v(v|s)$. In addition, the searcher may query and obtain from a third party (the *expert*), for a fee c_e, the true value v of an opportunity, for which signal s was received. It is assumed that rejected opportunities cannot be recalled and that the searcher is acquainted with the (stationary) probability density function from which signals are drawn, denoted $f_s(x)$ ($x \in \mathbb{R}$). The searcher's goal is to maximize her total utility, defined as the value of the opportunity eventually picked minus the overall cost incurred during the search.

The assumptions above are standard in the literature. In particular the above model is the same as the one used by MacQueen [66] and used as a basis by many others [91, 62]. When signals are not noisy, then the model is identical to the standard one used in the one-sided search literature in economics [70, 63].

The searcher's strategy is a mapping from the signal received and the number of queries remaining to the set of actions {`terminate`, `resume`, `query`}, which have the following interpretations: `terminate` means taking the current opportunity and ending search; `resume` means incurring a cost c_s and moving on to the next opportunity; `query` means engaging the services of the expert, paying a cost c_e and then making a decision. Unlike prior work, in our setting the expert is also a self-interested agent that sets its price strategically.

We assume that the provider of expert services only pays a marginal cost d_e per query and that her reports are truthful. It is justifiable to assume that the expert only pays a marginal cost per query whenever she has already performed the "startup work" necessary, which is usually the case with information services (since fixed costs are constant, they do not affect the price setting decision at the margin). Lastly, truthful reporting can be assumed for reputational or regulatory reasons.

The first question that arises is how to characterize the optimal strategy for the searcher, given the query cost c_e and search cost c_s. We consider a monopolist provider of expert services (e.g., part of the platform, or approved by the platform). The searcher's optimal strategy is directly influenced by the expert's fees, and therefore implicitly determines the expected number of times the services of the expert are required, and thus the expert's revenue. The expert's per-searcher profit π_e is a function of the expected number of queries purchased by the searcher, denoted η_{c_e}, the cost of the service c_e and her cost of producing the service d_e. Formally, $\pi_e = (c_e - d_e)\eta_{c_e}$. The problem can be thought of as a Stackelberg game where the expert is the first mover, and wants to maximize her profits with respect to the service terms she sets.

5.3   Optimal Policies

In this section, we analyze the searcher's optimal search strategy and her expected use of the expert's services, given the fee c_e set by the expert. The analysis builds on the trivial non-noisy model and gradually adds the complexities of signals and having the expert option. From the searcher's optimal search strategy we derive the expert's expected benefits as a function of the fee she sets, enabling maximization of the expert's revenue.

One-Sided Search   The optimal search strategy for the standard model, where the actual value of an opportunity can be obtained at cost c_s, can be found in the extensive literature of search theory [70, 43]. In this case, the searcher follows a reservation-value rule: she reviews opportunities sequentially (in random order) and terminates the search once a value greater than a reservation value x^* is revealed, where the reservation value x^* satisfies:

$$c_s = \int_{y=x^*}^{\infty} (y - x^*) f_v(y) dy \tag{5.1}$$

Intuitively, x^* is the value where the searcher is precisely indifferent: the expected marginal benefit from continuing search and obtaining the value of the next opportunity exactly equals the cost of obtaining that additional value. The reservation property of the optimal strategy derives from the stationarity of the problem — resuming the search places the searcher at the same position as at the beginning of the search [70]. Consequently, a searcher that follows a reservation value strategy will never decide to accept an opportunity she has once rejected and the optimal search strategy is the same whether or not recall is permitted. The expected number of search iterations is simply the inverse of the success probability, $\frac{1}{1-F_v(x^*)}$, since this becomes a Bernoulli sampling process, as opportunities arise independently at each iteration.

One-Sided Search with Noisy Signals   Before beginning the analysis of search with noisy signals, we emphasize that, given $f_v(x)$ and $f_s(s|v)$, we can also derive the distribution of the signal received from a random opportunity, $f_s(x)$, and the distribution of true values conditional on signals, $f_v(v|s)$ (the conditionals are interchangeable by Bayes' law). In many domains, it may be easier to assess/learn $f_s(s)$ than $f_v(v|s)$ as most past experience involves signals, with the actual value revealed for only a subset of these signals.

When the searcher receives a noisy signal rather than the actual value of an opportunity, there is no guarantee that the optimal strategy is reservation-value based as in the case where values obtained are certain. Indeed, the stationarity of the problem still holds, and an opportunity that has been rejected will never be recalled. Yet, in the absence of any restriction over $f_s(s|v)$, the optimal strategy is based on a set S of signal-value intervals for which the searcher terminates the search. The expected value in this case, denoted $V(S)$, is

given by:

$$V(S) = -c_s + \Pr(s \notin S)V(S) + Pr(s \in S)E[v|s \in S]$$

$$= -c_s + V(S) \int_{s \notin S} f_s(s) \, ds + \int_{s \in S} f_s(s) E[v|s] \, ds \tag{5.2}$$

The fact that the optimal strategy may not be reservation-value based in this case is because there may be no correlation between the signal and the true value of the opportunity. Nevertheless, in most real-life cases, there is a natural correlation between signals and true values. In particular, a fairly weak and commonly used restriction on the conditional distribution of the true value given the signal goes a long way towards allowing us to recapture a simple space of optimal strategies. This is the restriction that higher signal values are "good news" in the sense that when $s_1 > s_2$, the conditional distribution of v given s_1 first-order stochastically dominates that of v given s_2 [92, 72]. The condition requires that given two signals s_1 and s_2 where $s_1 > s_2$, the probability that the actual value is greater than any particular value v is greater for the case where the searcher receives signal s_1. This condition of stochastic dominance is satisfied if the conditional distribution $f_v(y|s)$ satisfies monotone likelihood propery [72]:

Definition 1 *Monotone likelihood ratio property (MLRP): A distribution $f(y|s)$ satisfies MLRP if the ratio $\frac{f(y|s')}{f(y|s)}$ is non-decreasing in y for $s' > s$ i.e., higher values are more likely drawn from $f(y|s')$ that $f(y|s)$. This also implies that $f(y|s')$ first-order, second-order and third-order stochastically dominates $f(y|s)$ for $s' > s$.*

This enables us to prove the following theorem.

Theorem 4 *For any probability density function $f_v(v|s)$ satisfying the MLRP assumption, the optimal search strategy is a reservation-value rule, where the reservation value, t^*, satisfies:*

$$c_s = \int_{s=t^*}^{\infty} \left(E[v|s] - E[v|t^*] \right) ds \tag{5.3}$$

Proof: The proof is based on showing that, if according to the optimal search strategy the searcher should resume her search given a signal s, then she must necessarily also do so given any other signal $s' < s$. The detailed proof is given in Appendix B. Note, the expected number of search iterations is $\frac{1}{1-F_s(t^*)}$, since this is a Bernoulli sampling process.   □

The Expert Option   The introduction of an expert extends the number of decision alternatives available to the searcher. When receiving a noisy signal of the true value, she can choose to (1) reject the offer without querying the expert, paying search cost c_s to reveal

the signal for the next offer; (2) query the expert to obtain the true value, paying a cost c_e, and then make a decision; or (3) accept the offer without querying the expert, receiving the (unknown) true value of the offer. In case (2), there is an additional decision to be made, whether to resume search or not, after the true value v is revealed.

As in the no-expert case, a solution for a general density function $f_v(v|s)$ dictates an optimal strategy of a complex structure. In our case, the optimal strategy will have the form of (S', S'', V), where: (a) S' is a set of signal intervals for which the searcher should resume her search without querying the expert; (b) S'' is a set of signal intervals for which the searcher should terminate her search without querying the expert (and pick the opportunity associated with this signal); and (c) for any signal that is not in S' or S'' the searcher should query the expert, and terminate the search if the value obtained is above a threshold V, and resume otherwise. The value V is the expected benefit from resuming the search and is given by the following modification of Equation 5.2:

$$V = -c_s + V \int_{s \in S'} f_s(s)\, ds - c_e \int_{s \notin \{S', S''\}} f_s(s)\, ds + \int_{s \notin \{S', S''\}} f_s(s) \left(V \int_{-\infty}^{V} f_v(x|s)\, dx + \int_{x=V}^{\infty} x f_v(x|s)\, dx \right) ds + \int_{s \in S''} f_s(s) E[v|s]\, ds \tag{5.4}$$

The first element on the right hand side of the equation applies to the case of resuming search, in which case the searcher continues with an expected benefit V. The second element is the expected payment to the expert. The next elements relate to the case where the search is resumed based on the value received from the expert (in which case the expected revenue is once again V) and where the search is terminated (with the value $E[v|s]$ obtained as the revenue), respectively. Finally, the last element applies to the case where the searcher terminates the search without querying the expert.

MacQueen [66]'s work can be reduced to show that under the MLRP assumption (first order stochastic dominance condition), each of the sets S' and S'' actually contains a single interval of signals, as illustrated in Figure 5.1.

Formally speaking,

> For $f_v(y|s)$ satisfying the MLRP assumption (Definition 1), the optimal search strategy can be described by the tuple (t_l, t_u, V), where: (a) t_l is a signal threshold below which the search should be resumed; (b) t_u is a signal threshold above which the search should be terminated and the current opportunity picked; and (c) the expert should be queried given any signal $t_l < s < t_u$ and the opportunity should be accepted (and search terminated) if the value obtained from the expert is above the expected value of resuming the search, V, otherwise search should resume (see Figure 5.1). The values t_l, t_u and V can be calculated from solving the set of Equations 5.5-5.7:

$$V = -c_s + V \int_{s=-\infty}^{t_l} f_s(s)\,ds - c_e \int_{s=t_l}^{t_u} f_s(s)\,ds + + \int_{s=t_u}^{\infty} f_s(s) E[v|s]\,ds$$

$$\int_{s=t_l}^{t_u} f_s(s) \left(V \int_{x=-\infty}^{V} f_v(x|s)\,dx\,ds + \int_{x=V}^{\infty} x f_v(x|s)\,dx\,ds \right) \tag{5.5}$$

$$c_e = \int_{y=V}^{\infty} (y - V) f_v(y|t_l)\,dy \tag{5.6}$$

$$c_e = \int_{-\infty}^{V} (V - y) f_v(y|t_u)\,dy \tag{5.7}$$

Intuitively, t_l is the point at which a searcher is indifferent between either resuming the search or querying the expert and t_u is the point at which a searcher is indifferent between either terminating the search or querying the expert. The cost of purchasing the expert's services must equal two different things: (1) the expected savings from resuming the search when the actual utility from the current opportunity (which is not known) turns out to be greater than what can be gained from resuming the search (once it is revealed) (this is the condition for t_l); (2) the expected savings from terminating the search in those cases where the actual utility from the current opportunity (once revealed) is less than what can be gained from resuming the search (for t_u).
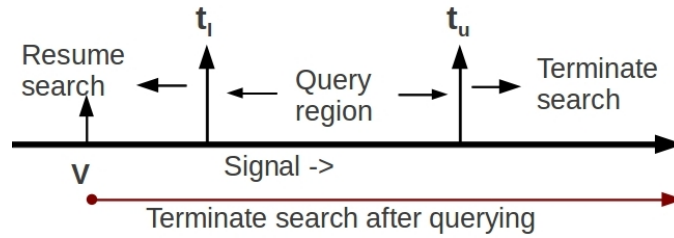
Figure 5.1: Characterization of the optimal strategy for noisy search with an expert. The searcher queries the expert if $s \in [t_l, t_u]$ and accepts the offer if the worth is greater than the value of resuming the search V. The searcher rejects and resumes search if $s < t_l$ and accepts and terminates search if $s > t_u$, both without querying the expert.

It is notable that there is also a reasonable degenerate case where $t_l = t_u(= t)$. This happens when the cost of querying is so high that it never makes sense to engage the expert's services. In this case, a direct indifference constraint exists at the threshold t, where accepting the offer yields the same expected value as continuing search, so $V = E[v|t]$. This can be solved in combination with Equation 5.3, since there are now only two relevant variables.

Expected number of queries: The search strategy (t_l, t_u, V) defines how many times the expert's services are consulted. In order to compute the expected number of queries, we consider four different types of transitions in the system. Let A be the probability that the searcher queries the expert and then does not accept, resuming search, B be the probability

that the searcher resumes search without querying, C be the probability that the searcher terminates without querying, and D be the probability that the searcher queries the expert and terminates search. Then:

$$A = \Pr(t_l \leq s \leq t_u \text{ and } v < V) \tag{5.8}$$
$$B = \Pr(s < t_l) \tag{5.9}$$
$$C = \Pr(s > t_u) \tag{5.10}$$
$$D = \Pr(t_l \leq s \leq t_u \text{ and } v \geq V) \tag{5.11}$$

Expected number of opportunities examined: Using the same notation as above, we see that the probability of terminating the search at any iteration is $C + D$, and these are independent Bernoulli draws at each opportunity. Therefore the expectation of the number of opportunities examined is simply $\eta_s = 1/(C + D)$.

Expected number of queries: The probability that the searcher queries the expert at any time she searches is $A + D$. Therefore $\eta_{c_e} = (A + D)\eta_s = \frac{A+D}{C+D}$.

Expected profit of the expert: Let d_e denote the marginal cost of the service the expert is providing. The expected profit of the expert is then simply

$$\pi_e = \mathbb{E}(\text{Profit}) = (c_e - d_e)\eta_{c_e}$$

The expert can maximize the above expression with respect to c_e (η_{c_e} decreases as c_e increases) to find the profit maximizing price to charge searchers.

5.4　A Specific Example

In this section we illustrate the theoretical analysis given in the former section for a particular plausible distribution of signals and values.

We consider a case where the signal is an upper bound on the true value. Going back to the used car example, sellers and dealers offering cars for sale usually make cosmetic improvements to the cars in question, and proceed to advertise them in the most appealing manner possible, hiding defects using temporary fixes. Specifically, we assume signals s are uniformly distributed on $[0, 1]$, and the conditional density of true values is linear on $[0, s]$. Thus

$$f_s(s) = \begin{cases} 1 & \text{if } 0 < s < 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_v(y|s) = \begin{cases} \frac{2y}{s^2} & \text{for } 0 \leq y \leq s \\ 0 & \text{Otherwise} \end{cases}$$

We can substitute in these distributions in Equations 5.5 through 5.7 and simplify. From

(a)   (b)   (c)
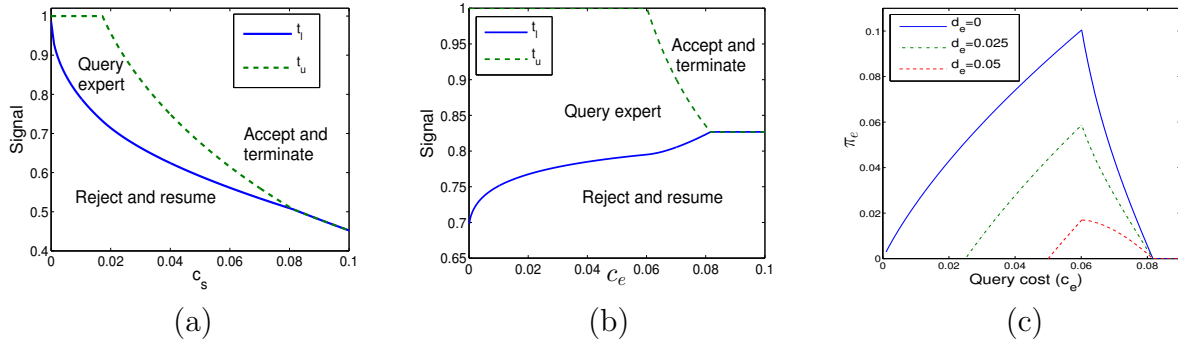
Figure 5.2: (a) Effect of c_s on the signal thresholds (t_l, t_u) for $c_e = 0.05$. (b) Effect of c_e on the signal thresholds (t_l, t_u) for $c_s = 0.01$. (c) Expert's profit as a function of c_e and d_e for $c_s = 0.01$. The model uses the a la carte pricing.

Equation 5.5:

$$V = -c_s + Vt_l - c_e(t_u - t_l) + \frac{V^3(t_u - t_l)}{3t_u t_l} + \frac{1 - t_l^2}{3}$$

$$V = \frac{2t_l}{3} + \frac{V^3}{3t_l^2} - c_e \quad \text{(from Equation 5.6)}$$

$$c_e = \int_0^V (V - y) f_v(y|t_u)\, dy = \frac{V^3}{3t_u^2} \quad \text{(from Equation 5.7)}$$

We can find feasible solutions of this system for different parameter values, as long as the condition $t_l < t_u$ holds. Otherwise, when c_e is high enough that querying never makes sense, a single threshold serves as the optimal strategy, as in the case with no expert. In the latter case, we obtain the optimal reservation value to be used by the searcher from Equation 5.3, yielding $t^* = 1 - \sqrt{3c_s}$.

The other thing to note here is that Equation 5.7 above is for the case when the support on signal s is unbounded. When there is an upper limit on s i.e $s \leq m$ for some m (as is the case here, where signals are bounded in $[0, 1]$), once t_u reaches m (we never buy without querying), Equation 5.7 does not hold. Now the system rejects if the signal is below t_l or queries if it is above.

Figure 5.2(a) illustrates how the reservation values t_l and t_u change as a function of c_s for $c_e = 0.05$ when using a la carte pricing. The vertical axis is the interval of signals. As can be seen from the graph, in this specific case for very small search costs (c_s), the searcher never terminates search without querying the expert.[1] Due to the low search cost the searcher should only query the expert when she receives a high signal, because the cost of finding

---

[1]When search costs are 0 the problem is ill-defined. The first point on the graph shows an extremely low, but non-zero search cost. In this case $t_u = 1$ and t_l is almost 1, but not exactly, and the expert is again always queried before search terminates.

opportunities with high signals is relatively low. As the search cost c_s increases, there is some behavior that is not immediately intuitive. The reservation values t_l and t_u become closer to each other until coinciding at $c_s = 0.08$, at which point the expert is never queried anymore. The reason for this is that the overall utility of continuing search goes down significantly as c_s increases, therefore the cost of querying the expert becomes a more significant fraction of the total cost, making it comparatively less desirable.

It is now easy to calculate numerically the value of c_e^* that maximizes π_e, trading off a decrease in the number of queries η_{c_e} in exchange for an increase in the revenue per query c_e. The expected number of queries can be found by substituting in the signal and noise distributions into Equations 5.8-5.11, yielding

$$A = V^2\left(\frac{1}{t_l} - \frac{1}{t_u}\right); \quad B = t_l; \quad C = (1 - t_u); \quad D = t_u - t_l - V^2\left(\frac{1}{t_l} - \frac{1}{t_u}\right)$$

which give the final expressions:

$$\eta_{c_e} = \frac{t_l t_u(t_u - t_l)}{t_u t_l^2 - t_l V^2 - t_u t_l + t_u V^2}; \quad \eta_{c_s} = \frac{1}{1 - t_l - V^2(\frac{1}{t_l} - \frac{1}{t_u})}$$

Figure 5.2(c) shows examples of the expected profit of the expert as a function of the expert's fee, c_e, for different values of d_e, the marginal cost to the expert of producing an extra report (fixing $c_s = 0.01$). Perhaps most interestingly, the shapes of the curves are the same and peak in the same region for significantly different values of d_e. It is also worth noting that the additional marginal costs are typically not being simply passed on to consumers – the expert suffers when the marginal cost of producing an additional report increases despite her monopoly position. This is explained, to some extent, by the decrease in searchers' tendency to use the expert as the cost of service increases (as illustrated in Figure 5.2(b)). An important implication of this phenomenon is that there may be welfare gains from additional efficiency in producing reports.

5.5　Heterogenous Customer Types

The above techniques can also be applied to solve for optimal expert strategies when there are multiple types of customers in the market. Assume there is a continuum of agent types distributed according to the pdf $f_w(w)$. Agents of different types differ in the values they assign to any subset of the parameters c_s, $f_s(s)$ and $f_v(y|s)$, such that every agent of type w is characterized by c_s^w, $f_s^w(s)$ and $f_v^w(y|s)$. Since each agent uses its best-response strategy based on the parameters c_s^w, $f_s^w(s)$, $f_v^w(y|s)$ and c_e, it is not affected by the search strategy set by the other agents. Therefore, given c_e, the strategy of agents of each agent type w can be calculated as described in Section 5.3, and consequently, the expected profit made by the

expert if the agent type w is the only agent in the market, π_e^w. The expected profit of the expert in this case is given by $\pi_e = \int_w \pi_e^w f_w(w) dw$.

We illustrate through a case where the population consists of a proportion θ of "low" type agents, who have the same distribution for $f_v(y|s)(= 2y/s^2)$ discussed above. The remaining proportion $1 - \theta$ of the population consists of "high" type agents who have a different conditional distribution for valuations given signals, $f_v(y|s) = 3y^2/s^3 (0 \le y \le s)$. Obviously the distribution of valuations of high-type agents stochastically dominates that of low-type agents.

For $c_s = 0.01$

For $c_s = 0.05$
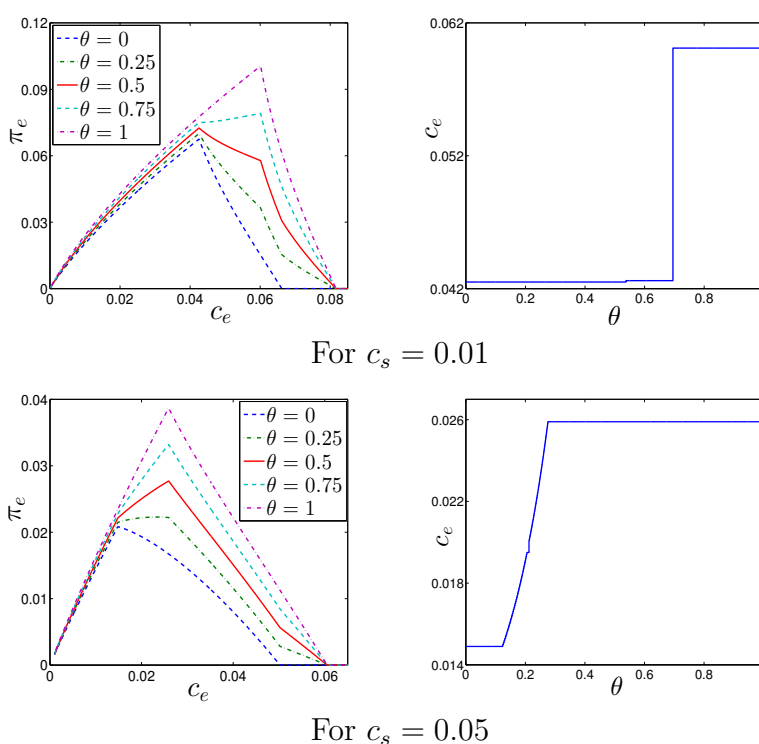
Figure 5.3: The expert's profit as a function of c_e for different proportions of the low type (θ) in the population, and the expert's profit-maximizing c_e as a function of θ, for two different costs of search c_s.

Figure 5.3 shows the profit of an expert for different values of θ vs. the query cost c_e, and the expert's profit-maximizing c_e as a function of θ, for two different costs of search c_s. In the expected profit graphs, each data point is a convex combination of the uppermost ($\theta = 1$) and lowest ($\theta = 0$) curves (weighted according to θ). We observe that the optimal query cost is a non-decreasing function of θ, the fraction of low type agents in the population. This is natural: low-type agents have more to gain from expert services since they receive less accurate signals. The expert charges a lower price for her services when the population is composed only of high-types as opposed to when it is composed only of low-types.
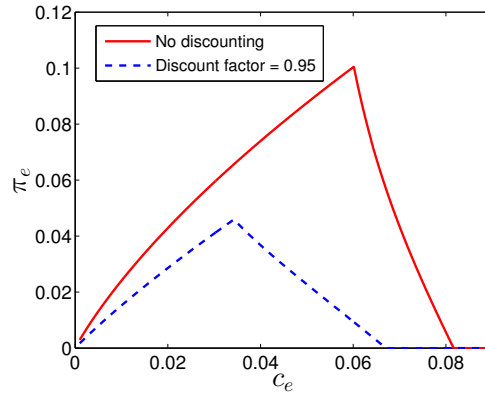
Figure 5.4: Expert's profit with and without discount factor for $c_s = 0.01$.

Interestingly, as θ increases, the change in the expert's strategy from the strategy that is optimal for an entire population of high-types (a higher-volume, lower-margin strategy) to a strategy that is optimal for an entire population of low-types (a lower-volume, higher-margin strategy) is rather sharp (a phase-transition-like pattern). This is because of the asymmetric behavior of the expert's profit around the optimal price: it increases to the peak slowly and decreases rapidly once past the peak. This effect is more apparent for a lower value of c_s, which would typically characterize electronic markets.

5.6   Robustness: Adding Discounting

We note here that, while particulars may vary, the form of our results is robust when we add a discount factor δ into the utility functions.[2] The changes for Equations 5.5-5.7 are easy; Equation 5.5 is updated to Equation 5.12 and Equations 5.6-5.7 remain the same.

$$V = -c_s + \delta \left(VF_s(t_l) + \int_{s=t_u}^{\infty} f_s(s) E[v|s]\, ds + \int_{s=t_l}^{t_u} f_s(s) \left(VF_v(V|s) + \int_{y=V}^{\infty} y f_v(y|s)\, dy - c_e \right) ds \right) \tag{5.12}$$

The only minor technical change is that one cannot use the simple formulation for the expert's strategy that we outline in Section 5.3 because simply multiplying the expected number of queries by the cost per query is insufficient – the timing of queries matters as well. Let P_s and P_q be the probabilities that the searcher resumes the search (given by $A + B$) and that the searcher queries the expert (given by $A + D$), respectively. The expert's profit π_e is given by:

---

[2]Although, in many sequential search environments, discounting can be neglected if it is assumed that it is a short-term search where search costs dominate, as in most online markets, discount factors may be important in more traditional search settings.

$$\pi_e = (c_e - d_e) \sum_{t=0}^{\infty} \delta^t \Pr(\text{Search is not terminated by time } t-1) \times \Pr(\text{Searcher queries at time } t)$$

$$= (c_e - d_e) \sum_{t=0}^{\infty} \delta^t P_s^t P_q = \frac{(c_e - d_e)P_q}{1 - \delta P_s} = \frac{(c_e - d_e)(A + D)}{1 - \delta(A + B)} \tag{5.13}$$

Figure 5.4 shows a comparison of the expert's profit as a function of query cost for the same environment as in the previous section for a discounted case (with discount factor of 0.95) and the non-discounted case. We find that the introduction of discounting does not significantly affect the nature of our results, so we do not consider it further.

5.7 Conclusion

The power of modeling markets using search theory is well established in the literature on economics and social science [63, 19, *inter alia*]. It has led to breakthroughs in understanding many domains, ranging from basic bilateral trade [82] to labor markets [75]. While knowledge has always been an economically valuable commodity, its role continues to grow in the Internet age. The ubiquity of electronic records and communications means that there is an increasing role for knowledge brokers in today's marketplaces. For example, it is now feasible for agencies like Carfax to collect the available records of every recorded accident, insurance claim, oil change, inspection, and so on for every car. The presence of such knowledge brokers necessitates that we take them into account in modeling the search process of consumers.

This chapter takes the first step in this direction. We study how these knowledge brokers (whom we term *experts*) set prices optimally in a search setting in which searchers receive noisy signals of the true value of an object, and can pay an expert to reveal more information.

There are several directions for future research. In the current chapter, we study the optimal pricing for the monopolist expert when the expert provides pay-per-use (a la carte) services. An interesting problem for the monopolist expert is the optimal pricing of a package of queries (non-linear pricing), where the consumer must purchase a package, instead of individual reports, or offer subscription-based pricing. For an example, as of January 2014, Carfax offers one report for $39.99, 5 reports for $44.99 and unlimited subscription for $54.99.

This new search model also raises interesting new questions about market design. We can analyze questions like whether there is scope for an authority like a market designer or regulator to improve social welfare by subsidizing the cost of querying the expert. Can the benefit to the searcher of having less friction in the process potentially more than offset the cost to the authority? The downside could be that if the authority provided too much subsidy, this could lead to inefficient overconsumption of costly (to produce) expert services.

Another possible future direction is to relax the assumption that the expert has access to perfect information about the true values and study the competition dynamics when experts compete not

just in price but also in quality of information provided. Note, that the analysis in this chapter still holds if the monopolist expert provides noisy information, as long as the MLRP assumption holds [66].

Chapter 6

Non-linear Pricing

6.1 Introduction

In Chapter 5, we introduce a model of search markets with information intermediaries (whom we termed experts) and used this model to study the strategic problem faced by a monopolist expert who needs to decide how to set the price per query (pay per use or a la carte pricing) in order to maximize her revenue.

In the real world, however, experts may not only offer pay-per-use query but also sell their services in *packages*, or as contracts over periods of time. For example, as of January 2014, Carfax offers five reports for $44.99, a single report for $39.99 and unlimited reports for one month for $54.99. Motivated by this observation, we extend our framework to allow the expert to choose from the three most common pricing models: (a) a la carte (pay-per-use) pricing; (b) fixed-fee pricing (subscriptions to services); (c) non-linear package pricing where the expert sells packages of a fixed number of uses of her services for a given price (the first two pricing models are actually special cases of the third, but worthy of special consideration because of their ubiquity).

We find that fixed-fee or subscription models are likely to be preferred by the expert in settings with low search costs and low marginal costs of producing extra reports, which are both key features of online markets.

6.2 The Model

We generalize the pricing strategy for the expert, discussed in previous chapter by allowing the expert to sell a package of $k > 0$ queries instead of one query.

In this case, for a cost c_e^k, the searcher obtains, upon purchasing a package, the right to use the expert's service k times along her search path with no additional cost. The package service offer generalizes two other common service pricing schemes: a la carte(pay-per-use) pricing (setting $k = 1$, charging c_e for each requested query, which we discussed in the previous chapter) and fixed-fee (sub-
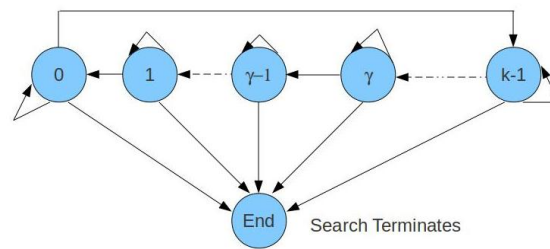
Figure 6.1: MDP representation of the searcher's problem when the expert offers packages of her services (state variable is the number of queries the searcher has remaining).

scriptions to services) pricing (setting $k \to \infty$ and charging cost c_e when the expert is first queried and none for each additional requested query). When $k > 1$, the searcher does not necessarily need to make use of all the k queries, and similarly, if she uses them all she can purchase additional packages as required. The goal of the searcher is to find an optimal strategy as before (mapping from a signal received and the number of queries remaining to $\{terminate, resume, query\}$). When choosing to query the expert the agent needs to pay a cost c_e^k if she has not purchased a package yet or if she has already used the expert's services k times since the last package was purchased.

The first question that arises in the different model variants is how to characterize the optimal strategy for the searcher, given the package terms (k, c_e^k) set by the expert (or the cost c_e set in the pay-per-use and subscription models). A second question is how the expert sets her package terms k and c_e^k in order to maximize her (per-searcher) expected profit, denoted by π_e. We consider a monopolist provider of expert services (e.g., part of the platform, or approved by the platform). The searcher's optimal strategy is directly influenced by the service terms, and therefore implicitly determines the expected number of times the services of the expert are required, and thus the expert's revenue. The problem can be thought of as a Stackelberg game [40] where the expert is the first mover, and wants to maximize her profits with respect to the service terms she sets to searchers. The goal of the searcher is to maximize the total utility received i.e., the expected value of the opportunity eventually picked minus the expected cost of search and expert fees paid along the way.

6.3 Optimal Policies

In this section, we derive optimal strategies for searchers and the resulting per-searcher profit for monopolist experts.

The Expert Option The introduction of an expert extends the number of decision alternatives available to the searcher. Now, when receiving a noisy signal she can, in addition to rejecting or accepting, also query the expert in order to obtain the true value of an opportunity for an additional fee. Since the searcher cannot recall previous opportunities, her state depends only on the number of remaining pre-paid queries, denoted by γ. The system can thus be modeled as a Markov Decision

Process (MDP) with k search states ($\gamma = 0, 1, 2, \ldots, k-1$) and one termination state, as illustrated in Figure 6.1. Let V_γ denote the expected utility-to-go of following the optimal search strategy starting from state γ. The process begins when the searcher is in state $\gamma = 0$. In this state, upon receiving a signal s, the searcher can either: (a) reject the current opportunity and continue search, starting from the same state $\gamma = 0$; (b) accept the current opportunity and terminate search; (c) query the expert for the true value of the current opportunity, incurring a cost c_e^k, and, based on the value received, either accept the current opportunity (terminating the search) or reject the current opportunity and continue search from state $k-1$. When in state $\gamma > 0$, the searcher has the same options when receiving a signal s, except that when querying the expert the agent does not incur any cost, and if, based on the value received, she decides to continue the search, it is resumed from state $\gamma - 1$.

While terminating without querying the expert is a legitimate option in any states $\gamma > 0$, we can show that it is never preferred (see Lemma 1 in Appendix B). Intuitively, if the searcher believes that the opportunity is valuable based on the initial signal and there are still queries remaining from the purchased package, the queries would be worthless if the searcher terminates without using them. On the other hand, if the signals are very weak, the user prefers to resume search without exhausting a query. For any signal s received in state $\gamma > 0$, the expected utility if querying the expert, denoted by $M(s, V_{\gamma-1})$, is given by: $M(s, V_{\gamma-1}) = \int_{-\infty}^{\infty} \max(x, V_{\gamma-1}) f_v(x|s)\, dx$. Resuming search without querying the expert yields V_γ. Therefore, the optimal strategy is to query the expert if $M(s, V_{\gamma-1}) > V_\gamma$. The expected utility of using the optimal strategy, when starting from state $\gamma > 0$, is thus given by:

$$V_{\gamma>0} = -c_s + V_\gamma \int_{M(s,V_{\gamma-1})<V_\gamma} f_s(s)\, ds + \int_{M(s,V_{\gamma-1})>V_\gamma} f_s(s) M(s, V_{\gamma-1})\, ds \tag{6.1}$$

Similarly, when in state $\gamma = 0$, the expected utility when obtaining a signal s is: (a) $E[v|s]$ if terminating the search without querying; (b) V_0 if resuming the search without querying the expert; and (c) $M(s, V_{k-1}) - c_e^k$ on querying the expert. For any signal s, the choice which yields the maximum among the three should be made. Let $\Lambda = \max(E[v|s], V_0, M(s, V_{k-1}) - c_e^k)$ and let ζ_1, ζ_2 and ζ_3 define the sets of signal support such that Λ equals V_0, $E[v|s]$, and $M(s, V_{k-1}) - c_e^k$, respectively. The expected utility is then:

$$V_0 = -c_s + V_0 \int_{\zeta_1} f_s(s)\, ds + \int_{\zeta_2} E[v|s] f_s(s)\, ds + \int_{\zeta_3} f_s(s) \left(M(s, V_{k-1}) - c_e^k \right) ds \tag{6.2}$$

While the structure of the optimal strategy depends on $f_v(y|s)$, under the MLRP assumption the optimal strategy has a simple representation in the form of reservation values.

Theorem 5 *For $f_v(y|s)$ satisfying the MLRP assumption the optimal strategy for a searcher can be described as (see Figure 6.2):*
(1) a tuple (t_l, t_u, V_{k-1}), corresponding to state $\gamma = 0$, such that for any signal s: (a) the search should resume if $s < t_l$; (b) the opportunity should be accepted if $s > t_u$; and (c) the expert should be queried if $t_l < s < t_u$ and the opportunity accepted (and search terminated) if the value obtained from the expert is above the expected utility of resuming the search, V_{k-1}, otherwise search should resume; and
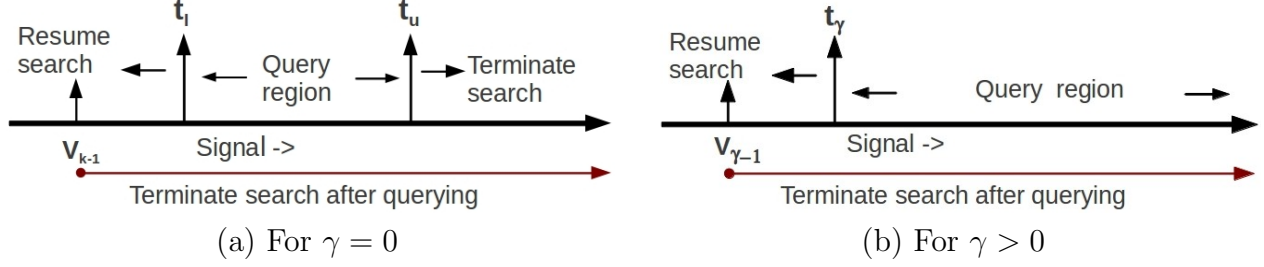
Figure 6.2: Characterization of the optimal strategy for noisy search with an expert offering packages.

(2) a set of $(k-1)$ tuples $(t_\gamma, V_{\gamma-1})$ corresponding to states $\gamma \in 1, 2 \ldots, k-1$ such that: (a) the search should resume if $s < t_\gamma$; and (b) the expert should be queried if $s > t_\gamma$ and the opportunity accepted if the value obtained from the expert is above the expected utility of resuming the search, $V_{\gamma-1}$, otherwise search should resume.
The values V_0, t_l, t_u, t_γ and V_γ (for $1 \le \gamma < k$) can be calculated from solving the set of Equations:

$$V_{\gamma>0} = V_\gamma F_s(t_\gamma) + \int_{s=t_\gamma}^{\infty} f_s(s)\left(\int_{y=V_{\gamma-1}}^{\infty} y f_v(y|s)\,dy + V_{\gamma-1}F_v(V_{\gamma-1}|s)\right)ds - c_s \tag{6.3}$$

$$V_0 = V_0 F_s(t_l) + \int_{s=t_u}^{\infty} f_s(s)E[v|s]\,ds + \int_{s=t_l}^{t_u} f_s(s)\left(V_{k-1}F_v(V_{k-1}|s) + \int_{y=V_{k-1}}^{\infty} y f_v(y|s)\,dy - c_e^k\right)ds - c_s \tag{6.4}$$

$$c_e^k = -V_0 + V_{k-1} + \int_{y=V_{k-1}}^{\infty} (y - V_{k-1})f_v(y|t_l)\,dy \tag{6.5}$$

$$c_e^k = \int_{y=-\infty}^{V_{k-1}} (V_{k-1} - y)f_v(y|t_u)\,dy \tag{6.6}$$

$$V_\gamma = V_{\gamma-1}F_v(V_{\gamma-1}|t_\gamma) + \int_{y=V_{\gamma-1}}^{\infty} y f_v(y|t_\gamma)\,dy \tag{6.7}$$

Proof: The proof is given in Appendix B.   □

Equations 6.3 and 6.4 are the appropriate modifications of Equations 6.1 and 6.2 for the MLRP case (when the searcher's optimal decisions are made based on the thresholds specified in the theorem). Equations 6.5-6.7, that characterize the searcher's optimal strategy, can also be derived from the searcher's indifference conditions: t_l is the signal at which a searcher is indifferent between either resuming t_l is the signal at which a searcher is indifferent between either resuming the search or querying the expert, i.e., $V_0 = \int_{y=V_{k-1}}^{\infty} y f_v(y|t_l)\,dy + V_{k-1}F_v(V|t_l) - c_e^k$, which transforms into Equation 6.5. t_u is the signal at which the searcher is indifferent between either terminating without querying the expert or purchasing a package of queries, i.e., $\int_{y=V_{k-1}}^{\infty} y f_v(y|t_u)\,dy + V_{k-1}F_v(V|t_u) - c_e^k = \int_{y=-\infty}^{\infty} y f_v(y|t_u)dy$, which transforms into Equation 6.6. Consequently, the two thresholds are obtained at the points where the cost of purchasing the expert's services equal: (1) the expected increase in utility from consulting the expert when you would otherwise reject and resume search

(for t_l, according to (6.5)); and (2) the expected increase in utility from consulting the expert when you would otherwise accept the offer, terminating search (for t_u, according to (6.6)). t_γ is the signal at which a searcher is indifferent between either resuming the search or querying the expert when she is in state $\gamma > 0$, which is essentially what Equation 6.7 represents.

There is also a reasonable degenerate case where $t_l = t_u(= t)$, i.e., a single threshold serves as the optimal strategy, as in the case with no expert. This happens when the cost of querying is so high that it never makes sense to engage the expert's services. In this case, a solution to the set of equations specified in Theorem 5 that satisfies $t_l < t_u$ does not exist. Instead, we obtain a single reservation value to be used by the searcher using Theorem 4.[1]

Finally, we note that the above analysis is applicable also for the case where the experts supplies a noisy signal rather than a definite value, whenever the expert's signal complies with MLRP. This is achieved using a transformation that can be found in [66].

A Subscription Model: Fixed Fee Pricing A specific case of importance is when the searcher can use as many queries as she would like upon payment of a fee to the expert (i.e., purchasing a package of size $k \to \infty$). In this case, the model reduces to two states. The searcher starts in state $\gamma = 0$ and continues in this state until she either terminates search or transitions to state $\gamma = \infty$. In the latter case the searcher can keep querying the expert for any reasonable opportunity until she finally finds a sufficiently good opportunity and terminates search (the existence of a search cost ensures that this querying process does not go on forever). Being in state $\gamma = \infty$ is equivalent to being in the world of perfect signals. The optimal reservation value when in state $\gamma = \infty$ can thus be extracted from (e.g., [70]):

$$V_\infty = -c_s + V_\infty F_v(V_\infty) + \int_{y=V_\infty}^{\infty} y f_v(y)\, dy \tag{6.8}$$

For state $\gamma = 0$, we can use appropriate modifications of Equations 6.4-6.6 replacing V_{k-1} with V_∞ (realizing that the searcher transitions to state $\gamma = \infty$). The optimal strategy can be extracted from the solution of this set of four equations.

6.3.1 The Expert's Perspective

A monopolist expert moves first in the implied Stackelberg game by setting the size and price (k, c_e^k) of the package (or the price c_e in the other two pricing schemes). The searcher responds by following her optimal search strategy described above. Therefore, the expert can solve for the searcher's behavior, given knowledge of the size and price (k, c_e^k) of the package and the signal and value distributions. The optimal package from the expert's perspective is thus the one that maximizes her expected overall profit. The expert's profit as a function of the size and price (k, c_e^k) of the package that she sets is given by $\pi_e = c_e^k \eta_b - \eta_{c_e^k} d_e$ (see Section 6.2). Given a way of

---

[1]Another technical point worth noting is that Equation 6.6 is for the case when the support on signal s is unbounded. When there is an upper limit on s, i.e., $s \le m$ for some m, once t_u reaches m (we never buy without querying), Equation 6.6 does not hold. Now the searcher rejects if the signal is below t_l or queries if it is above.

computing η_b and $\eta_{c_e^k}$, the profit-maximizing package (or the profit-maximizing query/subscription price) characteristics can be determined. The details of how to calculate η_b and $\eta_{c_e^k}$ is given below:

Expected Number of Packages Purchased (η_b) and Expected Number of Queries used $\left(\eta_{c_e^k}\right)$ In order to calculate η_b and $\eta_{c_e^k}$, we define and compute the following probabilities:

	Represents	General formulation **MLRP formulation**
$P_{\gamma \to \gamma-1}$	Pr (querying and resuming, moving from state $\gamma > 0$ to $\gamma - 1$)	$\int_{M(s,V_{\gamma-1})>V_\gamma} f_s(s)F_v(V_{\gamma-1}|s)\,ds$ $\Pr(\mathbf{s} > \mathbf{t}_\gamma$ and $\mathbf{v} < \mathbf{V_{\gamma-1}})$
$P_{\gamma \to \gamma}$	Pr (resuming without querying, staying in state $\gamma > 0$)	$\int_{M(s,V_{\gamma-1})<V_\gamma} f_s(s)\,ds$ $\Pr(\mathbf{s} < \mathbf{t}_\gamma)$
$P_{\gamma \to ter}$	Pr (querying and then terminating from state $\gamma > 0$)	$\int_{M(s,V_{\gamma-1})>V_\gamma} f_s(s)(1 - F_v(V_{\gamma-1}|s))\,ds$ $\Pr(\mathbf{s} \geq \mathbf{t}_\gamma$ and $\mathbf{v} \geq \mathbf{V_{\gamma-1}})$
$P_{0 \to k-1}$	Pr (querying and resuming, moving from state $\gamma = 0$ to $k - 1$)	$\int_{\zeta_3} f_s(s)F_v(V_{k-1}|s)\,ds$ $\Pr(\mathbf{t_l} \leq \mathbf{s} \leq \mathbf{t_u}$ and $\mathbf{v} < \mathbf{V_{\gamma-1}})$
$P_{0 \to 0}$	Pr (resuming without querying when $\gamma = 0$)	$\int_{\zeta_1} f_s(s)\,ds$ $\Pr(\mathbf{s} < \mathbf{t_l})$
$P_{0 \to ter}^{\neg query}$	Pr (terminating without querying when $\gamma = 0$)	$\int_{\zeta_2} f_s(s)\,ds$ $\Pr(\mathbf{s} > \mathbf{t_u})$
$P_{0 \to ter}^{query}$	Pr (querying and terminating when $\gamma = 0$)	$\int_{\zeta_3} f_s(s)(1 - F_v(V_{\gamma-1}|s))\,ds$ $\Pr(\mathbf{t_l} \leq \mathbf{s} \leq \mathbf{t_u}$ and $\mathbf{v} \geq \mathbf{V_{\gamma-1}})$

Notice that $P_{\gamma \to \gamma-1} + P_{\gamma \to \gamma} + P_{\gamma \to ter} = 1$, and similarly $P_{0 \to k-1} + P_{0 \to 0} + P_{0 \to ter}^{\neg query} + P_{0 \to ter}^{query} = 1$. Let $P_\gamma(T)$ denote the probability of eventually transitioning, when in state γ, to state $\gamma - 1$ (for $\gamma = 0$ it is the probability of eventually transitioning to state $k-1$). Let P_{cycle} be the probability of starting at a given state and getting back to it after going through all other states (excluding search termination state). The values of $P_\gamma(T)$ and P_{cycle} can be calculated as $P_\gamma(T) = \sum_{j=0}^{\infty}(P_{\gamma \to \gamma})^j P_{\gamma \to \gamma-1} = \frac{P_{\gamma \to \gamma-1}}{1-P_{\gamma \to \gamma}}$; $P_{\text{cycle}} = \prod_{i=0}^{k-1} P_i(T)$. Now let $P_\gamma(\text{Term})$ denote the probability of terminating the search in current state γ without transitioning to state $\gamma - 1$, and $P_0(\text{Buy} \wedge \text{Term})$ denote the probability of eventually purchasing the package and then terminating the search when starting from state $\gamma = 0$. Then:

$$P_\gamma(\text{Term}) = 1 - P_\gamma(T) = \frac{P_{\gamma \to ter}}{1 - P_{\gamma \to \gamma}}; \quad P_0(\text{Buy} \wedge \text{Term}) = \frac{P_{0 \to ter}^{query}}{1 - P_{0 \to 0}};$$

$$P_0(\neg\,\text{Buy} \wedge \text{Term}) = \frac{P_{0 \to ter}^{\neg query}}{1 - P_{0 \to 0}}$$

Expected number of Packages purchased (η_b)   In order for the searcher to purchase the jth package, she first needs to purchase $j - 1$ packages and use them fully (with probability $(P_{\text{cycle}})^{j-1}$), then purchase exactly one additional package. There are three ways for this to occur:

(1) the searcher purchases the package and terminates the search in state $\gamma = 0$ (with probability $P_0(\text{Buy} \wedge \text{Term})$); (2) the searcher purchases the package and terminates the search in some state $\gamma > 0$ (with probability $P_0(T) - P_{\text{cycle}}$); and (3) the searcher purchases the package, fully consumes it (i.e., returns to state $\gamma = 0$) and then terminates search without further package purchases (with probability $P_{\text{cycle}}P_0(\neg\,\text{Buy} \wedge \text{Term})$). Therefore,

$$\Pr(\text{purchase exactly } j \text{ packages}) = (P_{\text{cycle}})^{j-1}\Big(P_0(\text{Buy} \wedge \text{Term}) + P_0(T) - P_{\text{cycle}}$$
$$+ P_{\text{cycle}}P_0(\neg\,\text{Buy} \wedge \text{Term})\Big)$$

$P_0(T) = 1 - P_0(\neg\,\text{Buy} \wedge \text{Term}) + P_0(\text{Buy} \wedge \text{Term})$. Therefore, the above expression becomes $P_{\text{cycle}}^{j-1}(P_0(\text{Buy} \wedge \text{Term}) + P_0(T))(1 - P_{\text{cycle}})$. Then the expected number of packages purchased, η_b, is given by:

$$\eta_b = \sum_{j=1}^{\infty} j \Pr(\text{purchase exactly } j \text{ packages})$$
$$= \sum_{j=1}^{\infty} j(P_{\text{cycle}})^{j-1}(P_0(\text{Buy} \wedge \text{Term}) + P_0(T))(1 - P_{\text{cycle}})$$

Expected number of queries used $(\eta_{c_e^k})$ Let $P_m(Q)$ be the probability that exactly m queries are used. If $m < k$ then the only way to exhaust m queries is by transitioning from state $\gamma = 0$ to state $k - m + 1$ and eventually terminating the search without transitioning to the next state. $m = k$ implies that the searcher transitions to state $\gamma = 1$ and either terminates without transitioning to $\gamma = 0$ or terminates after transitioning to $\gamma = 0$ without purchasing a new package. Therefore:

$$P_m(Q) = \begin{cases} P_0(\text{Buy} \wedge \text{Term}) & m = 1 \\ P_0(T)\Big(\prod_{j=k-m+2}^{j=k-1} P_j(T)\Big)P_{k-m+1}(\text{Term}) & m < k \\ P_0(T)\Big(\prod_{j=2}^{j=k-1} P_j(T)\Big)P_1(\text{Term}) + P_{\text{cycle}}P_0(\neg\text{Buy} \wedge \text{Term}) & m = k \end{cases}$$

For $m > k$, we represent m as $jk + i$ where $i = m\%k$ and $j = \lfloor \frac{m}{k} \rfloor$. Here, j represents the number of full cycles completed (when all queries in a package are used) and i represents the number of queries used prior to terminating before finishing the $(j+1)$th round. This cyclic nature gives us the following recurrence:

$$P_{m=jk+i}(Q) = P_{\text{cycle}}P_{((j-1)k+i)}(Q) = \cdots = (P_{\text{cycle}})^j P_i(Q)$$

Therefore the expected number of queries is given by:

$$\eta_{c_e^k} = \sum_{m=0}^{\infty} mP_m(Q) = \sum_{j=0}^{\infty}\sum_{i=1}^{k}(jk+i)(P_{\text{cycle}})^j P_i(Q) = \sum_{j=0}^{\infty}\sum_{i=1}^{k}(jk+i)\left(\prod_{l=0}^{k-1} P_l(T)\right)^j P_i(Q) \quad (6.9)$$

Expected value of an opportunity received   Let ω be the random variable representing the value of an opportunity received, then $E_i(\omega)$ represents the expected value of the opportunity if the search terminates in state i and $P_i(\omega)$ represents the probability of terminating at that state.

$$E_i(\omega) = \begin{cases} \mathbb{E}[v|v > V_{i-1}, s > t_i] & i > 0 \\ \frac{\mathbb{E}[v|v>V_{k-1},t_l<s<t_u]\Pr(v>V_{k-1}\wedge t_l<s<t_u)+\mathbb{E}[v|s>t_u]\Pr(s>t_u)}{\Pr(v>V_{k-1}\wedge t_l<s<t_u)+\Pr(s>t_u)} & i = 0 \end{cases}$$

$$P_i(\omega) = \begin{cases} \sum_{j=0}^{j=\infty} P_{\text{cycle}}^j P_0(T)\left(\prod_{h=i+1}^{h=k-1} P_h(T)\right) P_i(\text{Term}) & i > 0 \\ \sum_{j=0}^{j=\infty} P_{\text{cycle}}^j (1 - P_0(T)) & i = 0 \end{cases}$$

$$\text{Expected value of opportunity} = \mathbb{E}(\omega) = \sum_{i=0}^{i=k-1} E_i(\omega) P_i(\omega)$$

Expected number of opportunities examined   By definition V_0 is the value of the opportunity accepted minus the total search and query cost incurred ($V_0 = \mathbb{E}(\omega) - \eta_s c_s - \eta_b c_e^k$). Therefore, $\eta_s = \frac{\mathbb{E}(\omega) - \eta_b c_e^k - V_0}{c_s}$.

6.4   An Example

The expert's optimal decision is affected by various parameters like distribution of values, correlation between signals and values, search frictions, query cost and the pricing model which makes difficult to analyse theoretically. Instead, we illustrate the properties of the model using a particular, plausible distribution of signals and values which is also used in Chapter 5:

$$f_s(s) = \begin{cases} 1 & \text{if } 0 < s < 1 \\ 0 & \text{otherwise} \end{cases} \qquad\qquad f_v(y|s) = \begin{cases} \frac{2y}{s^2} & \text{for } 0 \le y \le s \\ 0 & \text{Otherwise} \end{cases}$$

6.4.1   The Dependency of t_l, t_u and t_γ in c_s and c_e

A pattern similar to the a la carte case reported in Chapter 5 holds when using packages of more than a single query: The dependency of t_l and t_u in c_s and c_e is similar to the one illustrated in Figure 5.2. Figure 6.3(a) shows t_γ decreases as c_s increases because due to low search cost, the searcher is more likely to search and therefore query; as expected in Figure 6.3(b), t_γ increases as c_e increases.

6.4.2   Comparison of Different Pricing Plans

We now turn our attention to another aspect of the expert's strategy, whether she should engage in non-linear pricing.
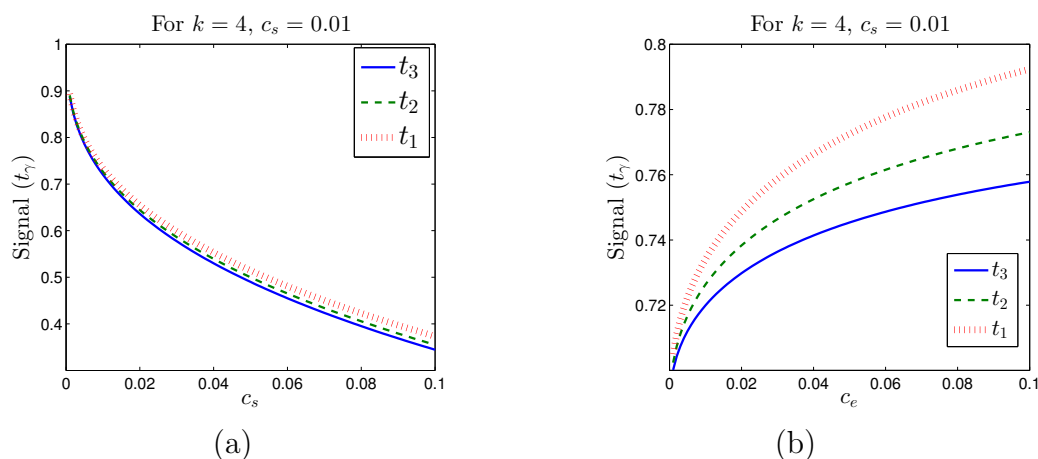
Figure 6.3: (a) Effect of c_s on the signal thresholds (t_γ) for $\gamma = 1, 2, 3$, $c_e^k = 0.01$ and package size $k = 4$. (b) Effect of c_e^k on the signal thresholds (t_γ) for states $\gamma = 1, 2, 3$, package size $k = 4$ and $c_s = 0.01$.

As a first question, we study whether offering a fixed fee, unlimited use scheme makes sense for the expert and when this is the case. Figure 6.4(a) shows the difference in the expected profit of the expert between offering a fixed-fee plan as opposed to per-use pricing as a function of search cost c_s for the digital goods case ($d_e = 0$). We see that fixed-fee pricing can yield significantly higher profits for the expert when search costs are low, but that per-use pricing is preferred when search costs are high (when they become too high consumers no longer use the expert and the profits of both go to 0). In a world of high search costs, users do not expect to keep searching for more than a few opportunities, so they are unlikely to be willing to purchase a fixed-fee plan at a price that would make sense for the expert.

Figure 6.4(b) examines the threshold value of c_s for different marginal costs of producing expert reports. At lower search costs, the expert expects higher profits from fixed-fee pricing, and at higher search costs, she expects higher profits from per-use pricing. We can see that the threshold decreases rapidly as a function of d_e (the X axis is on a log scale), indicating that as the marginal cost of expert services increases, per-use pricing quickly becomes preferred by the expert.

Both these observations, that increasing either the marginal cost of producing expert services and/or the cost incurred by the buyer in searching lead to unit size or smaller package sizes being preferred, correspond well to the real world and to recent observations in the literature. Unlimited subscription models are usually available in online services, where vanishing marginal costs and low search costs dominate. Sundararajan [86] provides a model of why subscription services are often preferred in the world of information goods, focusing on the vanishing marginal costs mentioned above. He includes transaction costs for administering pricing as a key part of his model (and these costs are obviously more relevant when the costs of producing the goods are becoming smaller and smaller). Our result is a parallel one in the realm of sequential search rather than direct pricing of products: we show the effects of search costs, which are borne by the consumer, and how they affect the optimality of subscription pricing models for information services (rather than information goods).
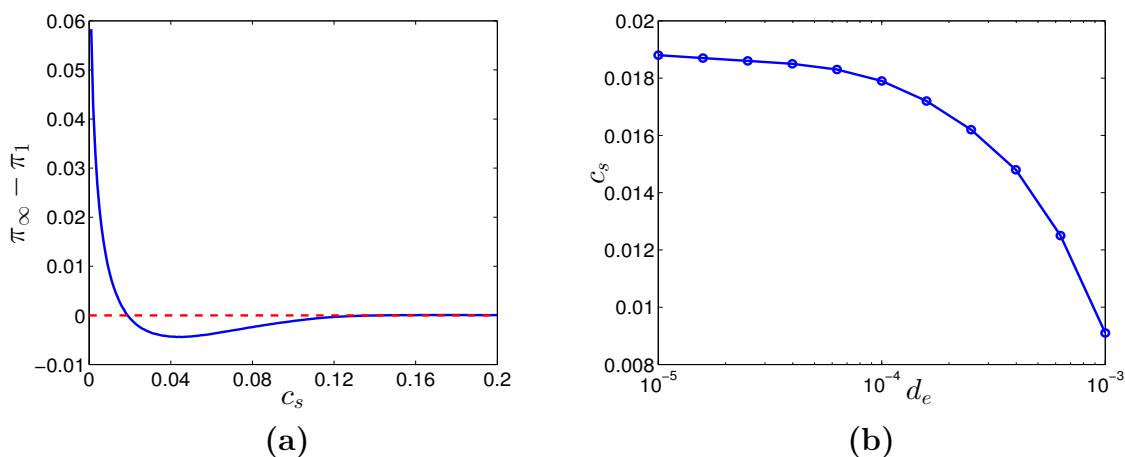
Figure 6.4: **(a)** Difference in expected profit of the expert for fixed-fee and per-use pricing for digital services ($d_e = 0$) vs. search cost. Fixed-fee pricing is preferred when search costs are low. **(b)** The search cost c_s at which the expert is indifferent between fixed-fee and per-use pricing, as a function of d_e, the marginal cost of producing an expert report. For lower search costs the expert prefers fixed-fee pricing, and for higher ones she prefers per-use pricing (semi-log scale).

In a different framework, that of bundling complementary information goods, Geng et al. [42], show that if consumers' values for future goods do not decrease too quickly, then bundling is (approximately) optimal for a monopolist, whereas if the values for future goods do decrease quickly bundling may not be optimal. In our framework, in the presence of high search costs, consumers are unlikely to want to use too many queries to the experts, so the marginal utility of an extra query is rapidly decreasing compared to the case where there are lower search costs.

Finite-size Packages   We can also solve numerically for finite k. Figure 6.5 depicts the expert's expected profit as a function of the package size[2] for $c_s = 0.01$ and different d_e values. One interesting observation is that when the marginal cost of producing an expert report is significantly higher than zero, the optimal number of queries to sell as a package tends to be small. The optimal package size increases as the service production cost d_e increases. Surprisingly, the expert's increased profit from non-linearly priced packages of greater size does not come at the cost of the searcher – searcher utility also increases, so packaging is pareto-improving. The search overall is becoming more efficient, and the expert and the buyer can split the additional gain. When the marginal cost is zero (as in the "digital services" case: an extra Carfax report can be produced essentially for free), the expert continues to achieve high profits with very high k, corresponding with the results above for when fixed-fee pricing is preferred.

---

[2]For each package size, the appropriate optimal c_e^k is used.
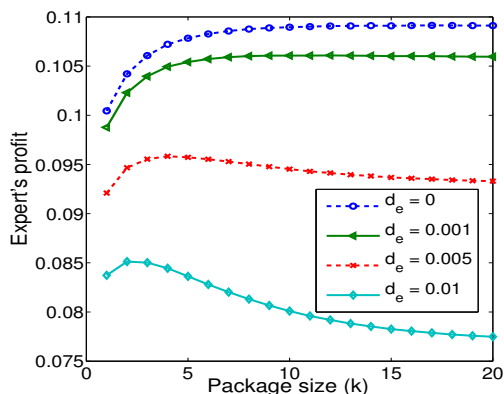
Figure 6.5: Expert profit as a function of the number of expert queries offered as a package for $c_s = 0.01$.

6.5   Conclusion

In this chapter, we study markets with informational intermediaries in the context of sequential search theory. We investigate the interplay between costly sequential search and costly information acquisition, and are able to gain new insights into how a monopolist information broker (an *expert*) should price her services. In doing so, we also characterize the optimal strategy for searchers to follow; this state-dependent "double reservation value"/ "single reservation value" strategy is a novel result for the type of information signals (non-additive) we consider here.

Generalizing ideas from the study of monopolistic setting to understand the dynamics of competition between experts (who could, for example, compete not just in price but also in quality of information provided, as in the third-party certification literature [36]) is an important extension of this work. Our analysis still holds if the monopolist expert provides a noisy signal associated with the true value (and the expected noise level could be a measure of quality), as long as the MLRP assumption holds; however, the extension to oligopolistic markets is non-trivial.

Another worthwhile extension would be to incorporate sellers' responses to the market dynamics into the equilibrium analysis, instead of modeling opportunities as exogenous. Finally, in some cases the platform may be in a position to *charge* experts for the privilege of connecting them with buyers; this poses some intriguing questions that would entail analysis of the dynamics of multi-platform/multi-expert competition. The model presented in this chapter should provide a good foundation for these extensions.

Chapter 7

Market Design - Can Subsidies Increase Social Welfare?

7.1 Introduction

In this chapter, we focus on the strategic implications of the expert-mediated search model developed in Chapter 5 for the platform that connects consumers and experts (to recapitulate, an expert is an intermediary who sells information to bridge the knowledge gap between the consumer and the seller e.g., Carfax).

Consider the design of a large scale Internet website like `AutoTrader.com`. The listings for cars that users see are signals, and users may be unsure of a car's true worth. `AutoTrader` (the platform) can partner with a provider of reports like Carfax (the expert), to make it easy for users to look up a car's worth. As another example, an online mortgage broker may choose to provide, as a premium product, tools that allow customers to learn more about the good and bad features of an offered mortgage. In this case the platform itself would be serving the role of the expert. The platform wishes to attract customers by providing a high value shopping experience. The expert wishes to maximize its profits. Since the platform and the expert both have significant power, it is reasonable to imagine them coming up with different models of the kinds of relationships they may have.

We focus here on one aspect of platform design, where the platform intervenes in the market by subsidizing expert queries. The platform can make payments to the expert in exchange for the expert providing services to users at a lower cost. A typical problem with subsidization is that it often decreases social welfare because the true cost of whatever is being subsidized is hidden from the consumer, leading to overconsumption of the resource. In this instance, however, the natural existence of monopolies in expert services, combined with the importance of search frictions, makes it quite possible that subsidies will in fact increase social welfare.

We first define social welfare in our context and then describe the effects of subsidization on social welfare. For simplicity of exposition, and because it captures all the relevant intuition, we focus

this discussion on the *a la carte* pricing model discussed in Chapter 5. It is also worth noting that this discussion is equally relevant to both private platforms as well as to regulators of large markets.

7.2   The Model

Recapitulation: One-sided search with noisy signals with expert   The notations in this chapter are consistent with previous chapters related to pricing information services. In a one-sided noisy search problem, a searcher faces an infinite stream of opportunities from which she needs to choose one. Each opportunity provides a specific value v to the searcher. The searcher can choose to pay search cost c_s to receive a noisy signal s, correlated with the true value according to a known probability distribution $f_v(v|s)$. We assume that the searcher is also acquainted with the (stationary) probability distribution from which the signal is drawn, denoted $f_s(x)$ ($x \in \mathbb{R}$). After observing a noisy signal s associated with an opportunity, the searcher can choose to: (a) reject it and resume search; (b) accept it and terminate search; (c) consult an expert, for a fee c_e, to know the true value, v, of the opportunity and then decide to either accept it or reject it.

Suppose a monopolist provider of expert services maximizes her profits by setting the query cost to c_e^*, yielding an expected profit $\pi_e = (c_e^* - d_e)\eta_{c^*}$ where $\eta_{c_e^*}$ represents the number of times a single searcher queries her. The platform can step in and negotiate a reduction of the fee c_e charged by the expert to a value c_e', for the benefit of the searchers. In return for the expert's agreement, the platform can offer a per-consumer payment β to the expert, which fully compensates the expert for the decreased revenue, leaving her total profit unchanged. Since $c_e' < c_e^*$, $\eta_{c_e'} > \eta_{c_e^*}$ (the consumer queries more often because she has to pay less). The compensation for a requested decrease in the expert's fee from c_e^* to c_e' is thus $\beta = (c_e^* - d_e)\eta_{c_e^*} - (c_e' - d_e)\eta_{c_e'}$. The overall welfare per agent in this case increases by $V_{c_e'} - V_{c_e^*}$, where $V_{c_e'}$ and $V_{c_e^*}$ are the expected utilities of searchers according to Equation 5.5-5.7 given in Chapter 5, when the expert uses a fee c_e' and c_e^* respectively, at a cost β to the platform.

The social welfare is given by the sum of utilities of all parties involved. While this function can have any form, here we study utilitarian or additive social welfare, which is appropriate especially in the context of possible transfers from the platform that connects experts and searchers. We consider two representative agents, the searcher and the expert, and note that this generalizes to multiple searchers (each search process is independent, and social welfare scales up linearly in the number of searchers). Then: $W = V_{c_e^*} + \pi_e$. When the platform enters the picture by subsidizing expert queries, the social welfare must also take the subsidy into account. Since the expert is fully compensated for her loss due to the decrease in her fee, the change in the overall social welfare is $V_{c_e'} - V_{c_e^*} - \beta$. Under the new pricing scheme c_e', and given the subsidy β, the social welfare is given by $W' = V_{c_e'} + \pi_e - \beta$.

7.2.1   Welfare-Maximizing Subsidies

The first important question that arises is what level the platform should set the subsidy at in order to maximize social welfare. Here we prove that, in the model of subsidization described

above, social welfare is maximized at the point where the searcher pays exactly d_e per query, thus fully internalizing the cost to the expert of producing the service. If the searcher had to pay less, it would lead to inefficient overconsumption of expert services, whereas if she had to pay more, the expected decline in the utility she receives from participating in the search process would outweigh the savings to the platform from having to pay less subsidy.

Theorem 6 *Suppose the platform pays the expert a flat subsidy β (per-customer) in exchange for the expert reducing her per-query fee from c_e^* to c_e' such that $\beta = (c_e^* - d_e)\eta_{c_e^*} - (c_e' - d_e)\eta_{c_e'}$ then social welfare is maximized when β is set such that $c_e' = d_e$.*

Proof: The searcher acts to maximize her utility. For convenience, we mildly abuse notation and assume that $\mathbb{E}[\omega()]$, $\eta_s()$ and $\eta_{c_e}()$ are functions defined by the searcher's optimal behavior from solving the Bellman equation. Here $\mathbb{E}[\omega()]$ is the expected value of the final opportunity that is taken (which is different from the expected utility of the searcher since it does not factor in costs), $\eta_s()$ is the expected number of opportunities examined by the searcher, analogous to η_s, and $\eta_{c_e}()$, analogous to η_{c_e}, the expected number of times the expert is queried.

Then

$$W = \mathbb{E}[\omega(c_e)] - \eta_s(c_e)c_s - \eta_{c_e}c_e + \eta_{c_e}(c_e - d_e) = \mathbb{E}[\omega(c_e)] - \eta_s(c_e)c_s - \eta_{c_e}(c_e)d_e \qquad (7.1)$$

We claim that social welfare is maximized when $c_e = d_e$. We show this by appealing to the optimal solution of a single searcher's decision problem. The functions $\mathbb{E}[\omega()]$, $\eta_s()$ and $\eta_{c_e}()$ in this case are such that the expression in Equation 7.2 is maximized.

$$\mathbb{E}[\omega(d_e)] - \eta_s(d_e)c_s - \eta_{c_e}(d_e)d_e \qquad (7.2)$$

Suppose there exists some $c_{e'} \neq d_e$ that maximizes the social utility (Equation 7.1). Let us now define an alternative strategy for the single searcher above. The searcher pretends that the cost of querying the expert is $c_{e'}$ and follows the strategy given by that belief. Then the expected utility of the searcher can be found by assuming the searcher pays cost $c_{e'}$ but receives a "kickback" of $c_{e'} - d_e$ every time. The expected utility of the searcher is then $(\mathbb{E}[\omega(c_{e'})] - \eta_s(c_{e'})c_s - \eta_{c_e}(c_{e'})c_{e'}) + \eta_{c_e}(c_{e'})(c_{e'} - d_e)$. By the Bellman Optimality Principle, this cannot be greater than Equation 7.2. Therefore, such a $c_{e'} \neq d_e$ does not exist. In addition, note that this is actually just the right hand side of Equation 7.1. Therefore, it must be the case that Equation 7.1 is maximized at $c_e = d_e$. $\square$

An immediate implication of this theorem is for the "digital services" case, where $d_e = 0$ (producing an extra Carfax report, for example, typically has zero marginal cost). In this case, there is no societal cost to higher utilization of the expert's services, so subsidy is welfare improving right up to making the service free. These are the cases where it could make sense for the platform to take over offering the service itself (for example, the online mortgage broker providing mortgage comparison and evaluation tools), and making it free, potentially leveraging the increased welfare of consumers by attracting more consumers to their market, or by increasing fees for the use of platform.

for $d_e = 0$ for $d_e = 0.025$ for $d_e = 0.05$
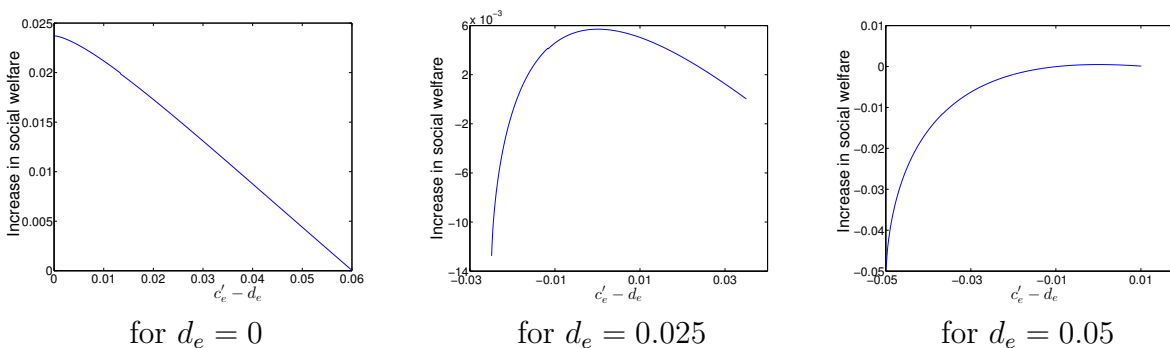
Figure 7.1: Increase in social welfare vs. the difference between the subsidized query cost $(c'_e < c^*_e)$ due to subsidy and the marginal cost d_e. When $d_e = 0$, it is best for the platform to make the service available for free, but when there is some marginal cost involved, then the increase in social welfare is concave, peaking at zero, i.e., when the subsidized price is equal to the marginal cost.

Table 7.1: The different components of social welfare with and without subsidy for $c_s = 0.01$. $\mathbb{E}(\omega)$ is the expected value of the opportunity eventually picked. Initially the decrease in query cost contributes more to the increase in social welfare, but as d_e increases, this contribution becomes less significant. Note that the first two rows in the case without subsidy are similar because the profit-maximizing c_e is the same and the searcher's cost depends only on the value of the selected c_e, not d_e.

	Without Subsidy						With Subsidy						
d_e	c_e	$\eta_{c_e}c_e$	$c_s\eta_s$	$\mathbb{E}(\omega)$	V	W	β	c_e	$\eta_{c_e}c_e$	$c_s\eta_s$	$\mathbb{E}(\omega)$	V	W
0	0.06	0.10	0.08	0.7472	0.5653	0.6656	0.1003	0	0	0.1036	0.7928	0.6893	0.6893
0.025	0.06	0.10	0.08	0.7472	0.5653	0.6238	0.0585	0.025	0.05	0.09	0.7712	0.6295	0.6295
0.05	0.061	0.09	0.08	0.7365	0.5637	0.5806	0.0169	0.05	0.09	0.08	0.7536	0.5824	0.5824

7.2.2   Numerical Simulations

Simulations using the example distributions in Chapter 5 serve to both confirm our theoretical results and provide insight into where the improvements come from. Figure 7.1 shows the improvement in social welfare as a function of the difference between the subsidized query cost c'_e and the marginal cost d_e for various d_e values (where $c_s = 0.01$). From the graphs, we do indeed find that subsidization can lead to substantial increases in social welfare, even when there is a significant marginal cost of producing an expert report. While this could be from a reduction in search and query costs or an increase in the expected value of the opportunity finally taken, the data in Table 7.1 indicates that the latter explanation is the dominant factor in this case. Also, in accordance with Theorem 6, social welfare is maximized at the point where the searcher pays exactly d_e per query, thus fully internalizing the cost to the expert of producing the extra report.

7.3 Discussion

In this chapter, we analyze the issues faced by platforms that bring searchers and experts together. In many search domains, the expert has a special relationship with the platform (e.g., an auto-trading website contracts with Carfax) or may even be embedded within the platform (e.g., an online mortgage broker could develop and offer a tool for understanding terms of different mortgages better). We show that the platform can create surplus by subsidizing expert queries to the point that searchers pay exactly the cost faced by the expert. The platform can capture some of the surplus through commissions or advertising, or through increased market share if it provides a better experience to consumers. In fact, if the platform can quantify its indirect benefit from the increase in the users' social welfare, it can use the analysis given here to determine the optimal subsidy, considering the tradeoff between the payment to the expert and the additional benefit deriving from the resulting increase in users' welfare. The same holds for a platform that is also offering the expert-like premium service — the tradeoff between the expected loss of direct income (i.e., query fees) and the expected gain in indirect benefits dictates the pricing of the service. In the important case of digital services, where the marginal cost of producing an extra unit of services is zero, our analysis suggests that expert services should be free to consumers, and therefore the platform may benefit particularly from a special relationship with, or ownership of, the expert.

Chapter 8

Noisy and Competitive Experts

8.1 Introduction

Consider the example of a searcher looking to a buy a used car discussed in previous chapters. She will examine advertisements for car sale sequentially until she finds a car she likes and buys it. When she looks at an advertisement, she may not interpret the quality of the car correctly, as the seller may not reveal the true condition of the car. The searcher may choose to consult an agency like Carfax, which provides information about the car's history. In the previous chapters (Chapters 5, 6 and 7), we modeled the impact of the presence of monopolistic information brokers or experts like Carfax on search markets like `autotrader.com` which serve as platforms for such consumer search [24, 25]. In that work we assume that these experts have access to perfect information while in reality, experts' signals are themselves likely to be noisy, and in this chapter, we relax this contraint. The noise in the expert's signal is a direct measure of the quality of the expert: the less the noise the higher the quality. By relaxing this constraint on the expert providing perfect information, we can also study oligopolistic information services markets as competition among multiple experts of varying quality can lead to non-trivial equilibrium market dynamics. Recognizing that quality may be determined exogenously or it might be a choice made by the experts themselves, we model the dynamics of competition between experts who can compete on both price and quality. For example, the used car information services Carfax and AutoCheck compete in both the price space (as of this writing, AutoCheck charges \$29.99 for a single report and Carfax charges \$39.99) and anecdotally, at least, are of different qualities.

We introduce a new model of competitive information provision in search markets. Within this model, we characterize the searcher's strategy when there are two competitive experts with different levels of information quality, and show that, under certain conditions, the optimal strategy of the searcher with imperfect experts in a duopoly is reservation-based: there are lower and upper thresholds between which an expert is consulted, as in the monopolistic expert case. Now, the region between the lower and upper thresholds is itself divided into three, with the higher quality expert being consulted in the middle of the region and the lower quality expert in the upper and lower parts of that region. We then compute the optimal price for experts as a function of the quality of the

information they are providing and the marginal cost related to it, both when information quality is controlled by the expert and when it is exogenous, and use it to study equilibrium dynamics.

8.2   The Model

The model is very similar to that in Chapter 5, but we summarize it again for clarity since the experts actions and how the searcher can use them are different. We consider a one-sided noisy search environment which is similar to that of in Chapter 5 where a searcher observes a stream of opportunities sequentially, from which she eventually needs to choose one. We summarize it again for clarity since the experts actions and how the searcher can use them are different. The value v of each opportunity is a priori unknown, however the searcher receives, upon paying a cost c_s, a noisy signal s which is correlated with the true value v according to a known probability density function $f_s(s|v)$. Let $f_v(x)(x \in \mathbb{R})$ denote the (stationary) probability density function for the true values (we assume the searcher knows $f_v(x)$). Even with the signal, the true value of the opportunity remains unknown and revealed only if it is accepted. The searcher thus has to decide whether to accept the opportunity and terminate the search or reject and pay to receive the signal for the next opportunity. The searcher also has a third option. She can consult an expert for some fee in order to get more information about the opportunity. An expert provides a noisy signal s_e. The conditional density function of s_e given the true value v is given by $f_e(s_e|v)$. The searcher's signal and the expert's signal are assumed to be conditionally independent given the true value, i.e., for each expert $f(s, s_e|v) = f_s(s|v)f_e(s_e|v)$; thus one can easily calculate the joint density $f(s, s_e, v)$ and other conditional densities using Bayes' rule. The searcher is a risk-neutral utility maximizer.

There are two competing experts the searcher can turn to. Each expert offers a different signal quality, captured by its unique conditional density $f_e(s_e|v)$, and the searcher, who is aware of the qualities of the two experts, can only use one of them. Specifically, we assume that signals are corrupted by zero-mean noise, ϵ, therefore $s_e = v + \epsilon$, (i.e., the mean of $f_e(s_e|v)$ is v). The variance σ_e^2 of the distribution is thus a measure for the expert's signal quality — the greater the variance, the lower is the quality and vice versa. Signals received from a zero variance expert represent perfect information.

The two experts are self-interested, risk-neutral, expected utility maximizers and act strategically (the expected utility of an expert is the expected payments received from searchers minus the cost of producing its signals in response to queries). We consider both exogenous and endogenous quality. In the first case each expert needs to set merely its fee, denoted by c_{e_i} for Expert $i \in \{1, 2\}$ while in the second case the decision involves both the quality and the fee to be charged. We assume the marginal cost of provision of services for each expert, denoted d_{e_i}, may vary for different qualities of information, and therefore represent it as a function of its variance $\sigma_{e_i}^2$ respectively for Expert $i \in \{1, 2\}$.

We assume that the searcher can only query once per signal, so if she chooses to consult an expert she also has to decide which expert to select. This restriction, where only one expert can be queried is applicable whenever the nature of the opportunity, or constraints imposed by the provider of the opportunity, preclude obtaining a second opinion. For example, a car owner might not be willing to let a potential buyer take his car to be inspected by more than one mechanic.

8.3　Optimal Policies

In this section, we derive the optimal strategy for the searcher and then use it to characterize the strategies for both experts. We first derive the form of the optimal strategy when multiple experts are available. Using the searcher's optimal strategy, we can calculate the expected demand for each expert's services, and thus their profit. This enables us to characterize equilibrium conditions and to study the phenomenological properties of the market in equilibrium.

8.3.1　Optimal Searcher Strategy

We have two competing experts E_1 and E_2, each providing a different degree of expertise such that their conditional distributions of the signals given the values are $f_{e_1}(s_{e_1}|v)$ and $f_{e_2}(s_{e_2}|v)$ respectively, and they charge c_{e_1} and c_{e_2} per query respectively. Without loss of generality, let E_1 be the lower quality (higher variance) expert and E_2 be the higher quality (lower variance) expert. For convenience, we refer to E_1 and E_2 as E_{low} and E_{high}, and their corresponding signals as $s_{e_{\text{low}}}$ and $s_{e_{\text{high}}}$, respectively. If the searcher decides to query and get more information about the current opportunity, she also has to decide which expert to query. Theorem 7 characterizes the structure of the searcher's optimal strategy for the duopoly problem under the MLRP.

Theorem 7 *If the conditional distribution of the true value v given the searcher's signal, s, ($f_v(v|s)$) and the conditional distribution of the true value v given either expert's signal $s_{e_{low}}$ or $s_{e_{high}}$ (i.e., $f_v(v|s_{e_{low}})$ and $f_v(v|s_{e_{high}})$) satisfy MLRP, the searcher's signal s and the experts' signal $s_{e_{low}}$ and $s_{e_{high}}$ are conditionally independent given the true value v, and the conditional distribution of Z given s is differentiable, then the optimal search strategy can be described by the tuple (V, t_l, t_u), where: (a) t_l is a signal threshold below which the search should be resumed; (b) t_u is a signal threshold above which the current opportunity should be accepted and the search should be terminated; (c) V is the expected utility of the searcher; if the signal s, she observes, lies in between t_l and t_u then the searcher should query one of the experts depending on whichever yields higher expected value from querying. The expected value of querying the experts E_{low} and E_{high} is given by Equations 8.1 and 8.2 respectively given the searcher's signal s:*

$$Z_{low} = \mathbb{E}(v|s, s_{e_{low}}) \quad Z_{high} = \mathbb{E}(v|s, s_{e_{high}})$$

$$U_{low}(s) = -c_{e_{low}} + V F_{z_{low}}(V|s) + \int_V^\infty z f_{z_{low}}(z|s)\, dz \tag{8.1}$$

$$U_{high}(s) = -c_{e_{high}} + V F_{z_{high}}(V|s) + \int_V^\infty y f_{z_{high}}(z|s)\, dz \tag{8.2}$$

On receiving the signal s_e, from the expert she chooses, she should terminate search if $\mathbb{E}(v|s, s_e) \geq V$ otherwise resume search. The values t_l, t_u, and V can be calculated using Equations 8.3-8.5:

$$V = -c_s + V F_s(t_l) + \int_{t_u}^\infty \mathbb{E}(v|s) f_s(s)\, ds + \int_{t_l}^{t_u} \max(U_{high}(s), U_{low}(s))\, ds \tag{8.3}$$

$$V = \max(U_{high}(t_l), U_{low}(t_l)) \tag{8.4}$$

$$\mathbb{E}(v|t_u) = \max(U_{high}(t_u), U_{low}(t_u)) \tag{8.5}$$

Proof: This is a simple corollary to MacQueen's result showing the optimality of a double reservation strategy for the searcher when there is only one expert under a stochastic dominance assumption on the distribution $F_Z(v|s)$ [65]. Since the MLRP implies stochastic dominance, MacQueen's proof can be extended by replacing the terms related to the utility of querying an expert in a monopolistic world with the maximum of the utilities that can be achieved by querying each expert separately. The detailed proof is in Appendix B.4.   □

Equation 8.3 is the main Bellman equation which summarizes the searcher's expected utility from continuing search, in which case she incurs a cost c_s for obtaining the signal of one additional opportunity. If the signal obtained is below t_l (with probability $F_s(t_l)$) then the search resumes, yielding expected utility V. Otherwise, if the signal obtained is above t_u, the search terminates and the agent's utility is $\mathbb{E}(v|s)$. Finally, if the signal is between t_l and t_u, the expert who provides higher expected utility to the searcher is used. Equations 8.4 and 8.5 can be obtained by setting the derivative of 8.3 w.r.t t_l and t_u respectively to zero; they can also be seen as indifference conditions at signal t_l, and t_u; for example, Equation 8.4 represents the fact that when the signal value is equal to t_l, the expert is indifferent between querying an expert (whichever one would be more useful given signal value t_l), and rejecting the opportunity, thus resuming search. There is also a degenerate case when neither expert is being queried. In that case, a single threshold serves as an optimal strategy for the searcher.

What happens between t_l and t_u? Theorem 7 tells us that there is a region (between t_l and t_u), where one of the experts will be queried. However, it is in general non-trivial to determine how the searcher's strategy behaves in $[t_l, t_u]$. However, for one of the most common and well-known value-signal structures in the literature [31, 52], we can show that it is optimal for the searcher to use a double reservation-value strategy to partition the region $[t_l, t_u]$ itself.

Theorem 8 *Assume the true value is normally distributed ($v \sim \mathcal{N}(\mu, \sigma_v^2)$) and signals are corrupted by additive white Gaussian noise (AWGN), i.e., $s = v + \epsilon_s$, $s_{e_{high}} = v + \epsilon_{e_{high}}$ and $s_{e_{low}} = v + \epsilon_{e_{low}}$ where ϵ_s, $\epsilon_{e_{high}}$ and $\epsilon_{e_{low}}$ represent independent draws from zero-mean independent Gaussians. Then the optimal strategy of the searcher can be characterized by a tuple (V, t_l, t_1, t_2, t_u) such that a rational searcher should: (1) reject all the signals $s < t_l$; (2) accept all the signals $s > t_u$ without querying the expert; (3) query the low quality expert if $t_l \leq s \leq t_1$ and $t_2 \leq s \leq t_u$. On consulting the expert, if $Z_{low} = \mathbb{E}(v|s, s_{e_{low}}) \geq V$, then accept the opportunity and terminate the search, otherwise resume; (4) query the high quality expert if the signal s lies in between t_1 and t_2. On consulting the expert if $Z_{high} = \mathbb{E}(v|s, s_{e_{high}}) \geq V$, then accept the opportunity and terminate the search, otherwise resume;*

Proof: Since $v \sim \mathcal{N}(\mu_v, \sigma_v^2)$, $f_{e_{\text{high}}}(s_e|v) \sim \mathcal{N}(v, \sigma_{e_{\text{high}}}^2)$ and $f_{e_{\text{low}}}(s_e|v) \sim \mathcal{N}(v, \sigma_{e_{\text{low}}}^2)$, the following holds:

$$f_{Z_{\text{high}}}(z|s) \sim \mathcal{N}\left(\frac{s\sigma_v^2 + \mu\sigma_s^2}{\sigma_v^2 + \sigma_s^2}, \frac{\sigma_s^4\sigma_v^4(\sigma_v^2 + \sigma_s^2)^{-1}}{\sigma_s^2\sigma_v^2 + \sigma_s^2\sigma_{e_{\text{high}}}^2 + \sigma_v^2\sigma_{e_{\text{high}}}^2} \right)$$

$$f_{Z_{\text{low}}}(z|s) \sim \mathcal{N}\left(\frac{s\sigma_v^2 + \mu\sigma_s^2}{\sigma_v^2 + \sigma_s^2}, \frac{\sigma_s^4\sigma_v^4(\sigma_v^2 + \sigma_s^2)^{-1}}{\sigma_s^2\sigma_v^2 + \sigma_s^2\sigma_{e_{\text{low}}}^2 + \sigma_v^2\sigma_{e_{\text{low}}}^2} \right)$$

This enables the calculation of $F_{Z_{\text{high}}}(z|s)$ and $F_{Z_{\text{low}}}(z|s)$. Since the mean is equal for the two conditional distributions above, and is expressed as a function of the variable s, it can be represented as $\mu(s)$. Let $\sigma_{Z_{\text{high}}}$ and $\sigma_{Z_{\text{low}}}$ represent the respective standard deviations. The difference in the utility from querying either expert is:
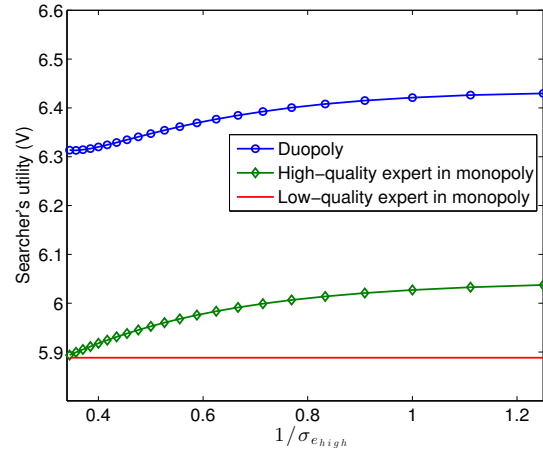
$$U_{\text{low}}(s) - U_{\text{high}}(s) = -c_{e_{\text{low}}} + c_{e_{\text{high}}} + \int_V^\infty (\Phi(z, \mu(s), \sigma_{Z_{\text{high}}}^2) - \Phi(z, \mu(s), \sigma_{Z_{\text{low}}}^2))\, dz$$

where $\Phi(x; \mu, \sigma^2)$ represents the normal cdf with mean μ and variance σ^2 evaluated at x. We want to find the region in which each expert is queried. In order to do so, we check how many times the sign of the difference of utilities received by querying the two experts changes. Note, that we only need to consider cases where $c_{e_{\text{high}}} - c_{e_{\text{low}}} > 0$, as otherwise the strategy is trivial — consult the high quality expert. We claim that $U_{\text{low}}(s) - U_{\text{high}}(s)$ is single peaked function, and therefore can change sign twice (have at most two roots). In order to do so we first find the points at which the first derivative is 0, and then perform higher order derivative test.

$$\frac{d(U_{\text{low}}(s) - U_{\text{high}}(s))}{ds} = \frac{dU_{\text{low}}(s)}{ds} - \frac{dU_{\text{high}}(s)}{ds}$$

$$= -\frac{\sigma_v^2}{\sigma_v^2 + \sigma_s^2}\left(\Phi(V; \mu(s), \sigma_{Z_{\text{low}}}^2) - \Phi(V; \mu(s), \sigma_{Z_{\text{high}}}^2) \right)$$

Putting the derivative of $U_{\text{low}}(s) - U_{\text{high}}(s)$ equal to zero, we get: $\Phi(V; \mu(s), \sigma_{Z_{\text{low}}}^2) = \Phi(V; \mu(s), \sigma_{Z_{\text{high}}}^2)$. The cdf for two normal distribution with same mean and different variance are equal at the extremes $\{-\infty, \infty\}$ or at the mean because cdf of any normal distribution at the mean is 0.5. We know V is finite, therefore, $\mu(s) = V \Rightarrow s = \frac{V(\sigma_v^2 + \sigma_s^2)}{\sigma_v^2}$. To check if this is an extremum, we calculate the second derivative of $U_{\text{low}}(s) - U_{\text{high}}(s)$:

$$\frac{d^2(U_{\text{low}}(s) - U_{\text{high}}(s))}{ds^2} = \frac{d}{ds}\left(\frac{dU_{\text{low}}(s)}{ds} - \frac{dU_{\text{high}}(s)}{ds} \right)$$

$$= \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_s^2} \right)^2 \left(\phi(V; \mu(s), \sigma_{Z_{\text{low}}}^2) - \phi(V; \mu(s), \sigma_{Z_{\text{high}}}^2) \right)$$

$$\left.\frac{d^2(U_{\text{low}}(s) - U_{\text{high}}(s))}{ds^2}\right|_{s=\frac{V(\sigma_v^2 + \sigma_s^2)}{\sigma_v^2}} = \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_s^2} \right)^2 \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma_{Z_{\text{low}}}} - \frac{1}{\sigma_{Z_{\text{high}}}} \right) > 0$$

$$\text{as } \sigma_{e_{\text{high}}} < \sigma_{e_{\text{low}}} \Rightarrow \sigma_{Z_{\text{high}}} > \sigma_{Z_{\text{low}}} \Rightarrow \frac{1}{\sigma_{Z_{\text{high}}}} < \frac{1}{\sigma_{Z_{\text{low}}}}$$
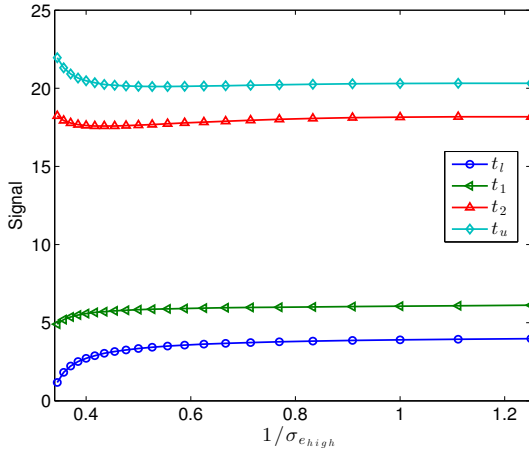
(a) Searcher's strategy vs. $\frac{1}{\sigma_{e_{\text{high}}}}$ (b) Effect of duopoly on searcher's utility

Figure 8.1: (a) The optimal strategy for the searcher (b) Searcher's utility as a function of the quality of the high-quality expert ($\frac{1}{\sigma_{e_{\text{high}}}}$). For comparison, the green and the red line show the searcher's utility when the high-quality expert and the low-quality expert are the monopolist respectively. For both graphs, we keep the low-quality expert's standard deviation, $\sigma_{e_{\text{low}}}$, constant at 3, $d_{e_{\text{low}}} = \frac{0.01}{\sigma_{e_{\text{low}}}^2}$, $d_{e_{\text{high}}} = \frac{0.01}{\sigma_{e_{\text{high}}}^2}$, $c_s = 0.1$, $\mu_v = 5$, $\sigma_s = 4$, $\sigma_v = 2$.

where $\phi(x; \mu, \sigma^2)$ represent the normal probability density function with mean μ and variance σ^2. As the function is single peaked it can at most change sign twice (intersect with x-axis). Note, if the maximum of $(\int_V^\infty (\Phi(z; \mu(s), \sigma_{Z_{\text{high}}}^2) - \Phi(z; \mu(s), \sigma_{Z_{\text{low}}}^2)) \, dz)$ is less than $c_{e_{\text{low}}} - c_{e_{\text{high}}}$ then the high quality expert will never be queried. If there are two roots of $U_{\text{low}}(s) - U_{\text{high}}(s) = 0$, say t_1 and t_2, then as the point of extremum calculated is a minima (the second derivative is positive and the first derivative is zero) then for signals less than t_1 or greater than t_2, the low-quality expert will be preferred, and in between the high-quality expert. If $t_1 < t_l$ and $t_2 > t_u$ then the low quality expert is never queried in practice because then resuming the search or accepting without querying are better alternatives for signals $s < t_1$ and $s > t_u$ respectively.   □

8.3.2   The Experts' Strategies

While the strategies described above are optimal for the searcher no matter what prices the experts set, the non-degenerate cases only hold when the experts pursue sensible pricing strategies, given their respective qualities. In a duopoly, for both E_{low} and E_{high} to survive, it must be the case that the lower quality expert charges less in equilibrium ($c_{e_{\text{low}}} < c_{e_{\text{high}}}$), otherwise it will never be queried.

Under the white Gaussian noise assumption, the query region of the two experts is nicely partitioned, therefore, we can calculate each expert's profit analytically (similar to the monopolist profit calculation discussed in [24]) and we use that to solve for equilibrium using best-response dynamics.

Let A, B, C, D, E and F represent the probability of (1) rejecting and resuming the search; (2) accepting the search without querying the expert; (3) querying the low quality expert and terminating the search; (4) querying the low-quality expert and resuming the search; (5) querying the high-quality expert and terminating the search; (6) querying the high-quality expert and resuming the search respectively.

The expected number of opportunities examined $\eta_s = \frac{1}{B+C+E}$ (because termination is a Bernoulli trial with probability $B + C + E$). The expected number of times an expert is queried, denoted by η_{c_e}, is given by:

$$\eta_{c_e} = \Pr(\text{Expert is queried})\eta_s$$

Therefore, the expected number of times E_{low} and E_{high} are queried is $\eta_{c_{e_{\text{low}}}} = \frac{C+D}{B+C+E}$ and $\eta_{c_{e_{\text{high}}}} = \frac{E+F}{B+C+E}$ respectively. We now turn to a specific example to understand the properties of equilibrium.

8.3.3   An Example

Suppose we have two experts E_{high} and E_{low} such that the conditional distributions of their signals, $s_{e_{\text{high}}}$ and $s_{e_{\text{low}}}$ respectively, given the true value v are both normally distributed with mean v and variance $\sigma^2_{e_{\text{high}}}$ and $\sigma^2_{e_{\text{low}}}$ (i.e., $f_{e_{\text{high}}}(s_e|v) \sim \mathcal{N}(v, \sigma^2_{e_{\text{high}}})$ and $f_{e_{\text{low}}}(s_e|v) \sim \mathcal{N}(v, \sigma^2_{e_{\text{low}}})$, and $\sigma_{e_{\text{high}}} < \sigma_{e_{\text{low}}}$). The distribution of the true value $v \sim \mathcal{N}(\mu_v, \sigma^2_v)$. This example satisfies the assumptions for Theorem 8.

Exogenous expert quality

First, we assume that the quality of both experts is fixed and different, but they compete on price. We use a best-response dynamic to find equilibrium prices for the experts to set, as mentioned above.

Searcher strategy and utilities: The searcher's optimal strategy is characterized by a 5-tuple (V, t_l, t_1, t_2, t_u) as per Theorem 8. Figure 8.1(a) shows the searcher's strategy as a function of the quality of the higher-quality expert ($\frac{1}{\sigma_{e_{\text{high}}}}$), holding the quality of the lower-quality expert constant. Therefore, the quality of the higher-quality expert increases as we go right. We can see that the lower-quality expert is only utilized for signals that are "close to the edge" in terms of whether or not the searcher wishes to consult an expert at all. The higher quality expert is consulted far more, since it is consulted for all signals that fall in the intermediate range. Figure 8.1(b) shows the expected utility of the searcher as the quality of the better expert increases, but we can see that the rate of increase is certainly no faster than it would be if the better expert was a monopolist, and, as expected, the searcher is much better off in a world with competition than a world where the expert is a monopolist. The extent to which this effect holds is interesting. Note that the searcher potentially has access to much better information at the rightmost point of the green line (a monopolist with $\frac{1}{\sigma_{e_{\text{high}}}} \approx 1.25 \Rightarrow \sigma_{e_{\text{high}}} \approx 0.8$), than at the leftmost point of the blue line (duopolistic experts with $\sigma_{e_{\text{low}}} = 3.0$ and $\frac{1}{\sigma_{e_{\text{high}}}} \approx 0.35 \Rightarrow \sigma_{e_{\text{high}}} \approx 2.85$ respectively).
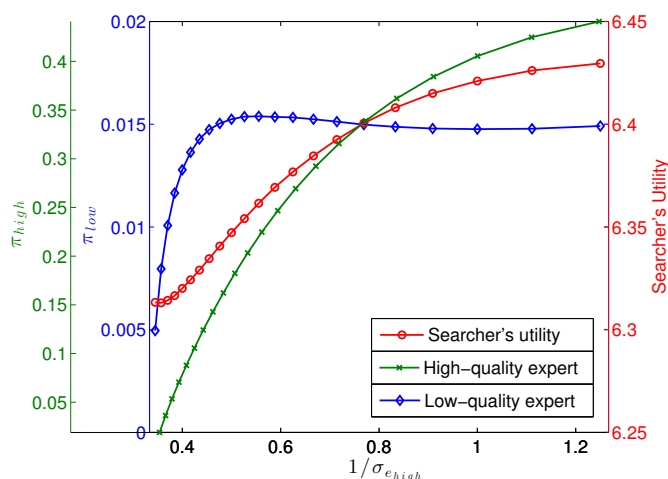
Figure 8.2: Equilibrium profits of the two experts and the searcher's utility as a function of quality of the high-quality expert $(\frac{1}{\sigma_{e_{\text{high}}}})$, keeping the quality of the low-quality expert constant $(\sigma_{e_{\text{low}}} = 3)$. For $\frac{1}{\sigma_{e_{\text{high}}}} < 0.5 \Rightarrow \sigma_{e_{\text{high}}} > 2$, we see that the profit of both high-quality and low-quality expert and the utility of the searcher increases as the quality of the high-quality expert increases, therefore, the overall social welfare also increases. The settings are same as in Figure 8.1.

However, the searcher's expected utility is much higher for the leftmost point of the blue line than the rightmost point of the green line, showing that the high expert prices charged in monopoly are much worse for the searcher than having potentially worse information on the basis of which to make a decision.

Expert behavior and profits: Figure 8.2 shows the profits accruing to each expert (along with the searcher's utility) as a function of the quality of the high-quality expert $\left(\frac{1}{\sigma_{e_{\text{high}}}}\right)$, holding $\sigma_{e_{\text{low}}}$ constant ($= 3$). For this figure, the marginal cost of production of expert reports is a function of the quality provided (specifically, $d_e = \frac{0.01}{\sigma_e^2}$) for both experts. Several interesting things jump out from this picture. First, the profit of the higher quality expert is an order of magnitude higher than the profit of the lower quality expert (the Y axes for the two curves are different, the higher quality expert always makes more profit than the lower quality one). The profit of the higher quality expert increases significantly as her quality increases. Despite the probability of the higher quality expert getting queried in each round ($\Pr(t_1 \leq s \leq t_2)$) decreases the amount it is able to charge ($c_{e_{\text{high}}}^{\text{eq}}$) increases at a much faster rate (see Figure 8.3), and hence the profit. The behavior of the profit of low quality expert is more complex — it initially increases rapidly when the difference in the quality of the high-quality expert increases and then starts decreasing very slowly. This initial increase is particularly surprising because one would imagine that enhancing the quality of the better expert would hurt the low quality expert. This effect can be teased out by again examining the probabilities of each expert being consulted and their fees in Figure 8.3. Interestingly, although the probability of consulting the low quality expert decreases as the quality of the high quality expert increases, the price she charges also increases. The higher quality expert charges more when its quality is improved in order to offset its increasing marginal cost associated with increasing

(a) Probability of querying either expert   (b) Prices charged by the two experts
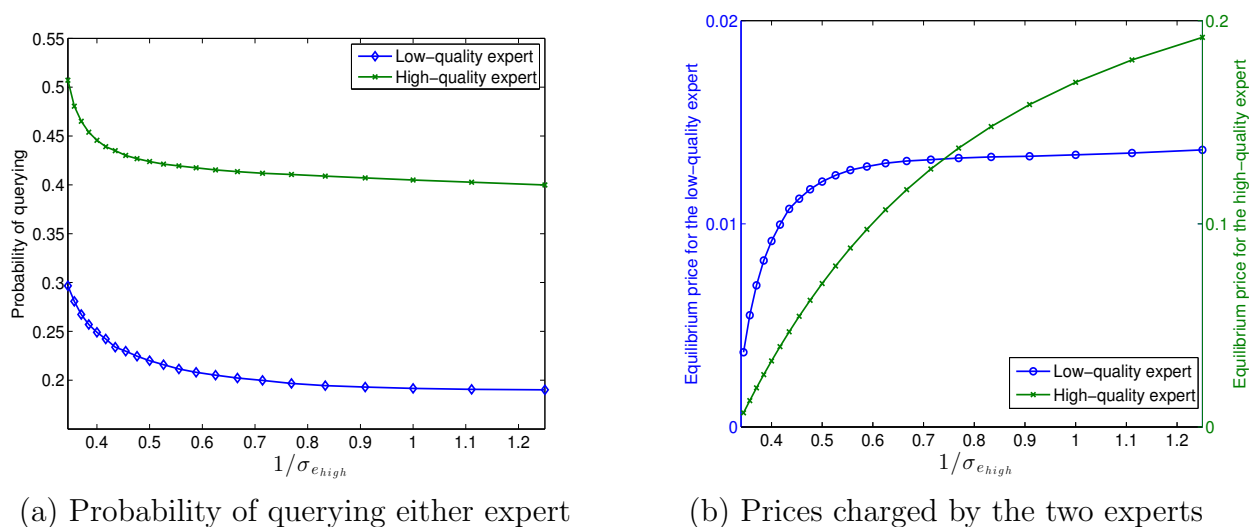
Figure 8.3: Probability that an expert will be queried (left) and price set by each expert (right) as a function of the quality of the high-quality expert ($\frac{1}{\sigma_{e_{\text{high}}}}$). The definitions and settings are the same as in Figure 8.1

quality. It is possible that in this situation, the low-quality expert is also able to take advantage of the higher price charged by high-quality expert by increasing its own fee and thus gaining more profit. However, eventually the rate of increase of the query price at equilibrium for the low quality expert decreases, and the declining probability that she is used becomes more important.

It is worth noting that the searcher's utility (the red line in Figure 8.2) also increases with the increase in the quality of the high-quality expert, mainly because better information is available to the searcher. Clearly, the searcher and the high-quality expert are doing better as the quality of the high-quality expert improves. The fact that the low-quality expert's profit also increases until $\sigma_{e_{\text{high}}} > 2$ (or $\frac{1}{\sigma_{e_{\text{high}}}} < 0.5$) therefore demonstrates the surprising result that improving the quality of the high-quality expert can be pareto-improving.

Endogenous expert quality

When the expert not only has control over the price it charges but also over the quality it offers, additional equilibrium dynamics must be concerned. One option in this case is that the experts have already done enough research on their part and are able to provide any quality and choose the one which is in their best interest, i.e., in order to maximize their expected profit. Alternatively, it is possible that the experts incur some switching cost whenever changing their quality. We analyze equilibrium conditions for these two situations for two experts with identical marginal costs of producing expert reports.

No switching costs   It is a fairly standard result on differentiated Bertrand competition that the following conditions must hold in equilibrium (cf. [37]):

1. The two experts should make identical profits, otherwise one expert can slightly undercut the other expert in price at the same quality and increase her profit, taking over the entire market share.

2. At equilibrium the price charged by each expert should be equal to the marginal cost, otherwise they will tend to undercut each other and capture the complete market. Hence, their profit at equilibrium is zero. However, not all marginal cost pricing strategies are in equilibrium.

In order to find an equilibrium in this case one needs to find a quality (σ_e) (the price will be equal to the marginal cost because of the condition discussed above), such that the best response of the second expert is to charge a price equal to the marginal cost for any quality. For example, if the marginal cost of production is not dependent on the quality and is a fixed identical constant for both experts, then in equilibrium both experts provide perfect information at a that is equal to their marginal cost (e.g., in the case of informational goods, that are characterized with zero marginal cost of production, the service should be provided for free at equilibrium).

In the presence of switching costs   Consider two experts with quality and price combinations (σ_1, c_{e_1}) and (σ_2, c_{e_2}) respectively such that c_{e_1} and c_{e_2} are equilibrium prices given that qualities are fixed. If the experts are also allowed to change their quality and either of them find a strategy (σ'_e, c'_e) to be more beneficial, taking the strategy of the other expert to be fixed, then the solution no longer complies with the equilibrium stability condition. However, if changing one's quality incurs a cost, which is greater than the benefit achieved by deviating to any new quality (and its corresponding price), the equilibrium conditions will still hold. It is natural to assume zero costs for decreasing quality but positive costs for improving quality, so we model switching costs using a hinge function. In this case, any quality pairs (σ_1, σ_2) for which neither of the agents finds a benefit in switching to a lower quality can be stabilized, and thus remain in equilibrium, using an appropriate switching cost (that applies only to increasing quality). Consider the settings depicted in Figure 8.2 for example. From the figure we observe that the curve corresponding to the higher quality expert's profit increases as the quality it provides increases. Therefore in all the potential solutions depicted in the figure that agent has an incentive to improve its quality and not to worsen it. Checking the lower quality expert's benefit from using a different quality in all the solutions captured by Figure 8.2 reveals that this expert, also, has an incentive to increase its quality rather than worsen it. Therefore, if the switching cost is high enough then all the solutions represented in this figure can be guaranteed to be in equilibrium, even with endogenous quality.

In order to understand how the switching cost affects the nature of equilibria, consider an example where the switching cost depends linearly upon the amount of improvement required to improve the profit. Formally, $S = \max(A(\frac{1}{\sigma'_{e_i}} - \frac{1}{\sigma_{e_i}}), 0)$ where $A > 0$ and σ'_{e_i} is the new standard deviation for Expert $i \in \{1, 2\}$ (the higher the standard deviation of the noise, the lower the quality). In order for a pair of standard deviations to be in equilibrium, the proportionality constant A must be sufficiently high to incentivize the two experts to not deviate from the qualities they are offering. Figure 8.4 shows how the minimum value of the proportionality constant A required to keep the two experts in equilibrium varies as a function of the quality of the high-quality expert ($\frac{1}{\sigma_{e_{\text{high}}}}$), keeping the quality of the low quality expert constant. While the figure shows the minimum value

of constant A needed for both experts separately, under the assumption that there is a universal switching cost function for both experts, A would effectively be the max of the two at each point. Note that in this example the proportionality constant for the low-quality expert is always higher than that of the high-quality expert. One reason for this could be that the low-quality expert can gain more by improving her quality. Also, the minimum value needed for the proportionality constant decreases as the quality of the high-quality expert increases. Therefore, if we choose any value of A (say $A = 1$), then only the points lying on the right of $\frac{1}{\sigma_{e_{\text{high}}}} \approxeq 0.7$ will be in equilibrium.
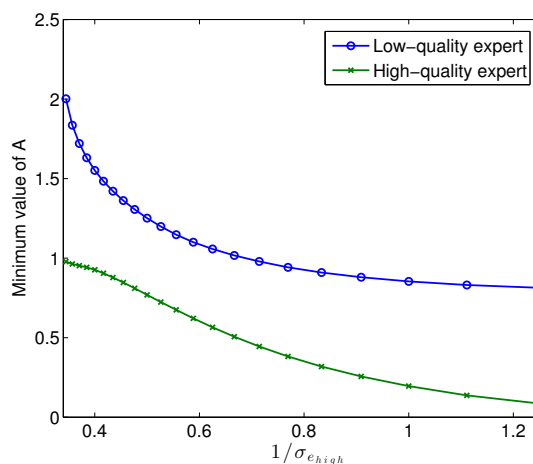
Figure 8.4: Minimum value of A, the proportionality constant for the switching cost, required for the two experts to be in equilibrium as a function of the quality of the high-quality expert ($\frac{1}{\sigma_{e_{\text{high}}}}$) keeping $\sigma_{e_{\text{low}}}$ constant. The switching cost $S = \max\left(A(\frac{1}{\sigma'_{e_i}} - \frac{1}{\sigma_{e_i}}), 0\right)$ where σ'_{e_i} and σ_{e_i} are the new and the current standard deviations respectively for Expert $i \in \{1, 2\}$. We find that the minimum value of this constant decreases as the quality of the high quality expert increases further, and the value is always higher for the low quality expert. Therefore, if A is equal to any point on the blue curve, then all the points to its right will be in equilibrium but none of those to its left. The settings are same as Figure 8.1.

8.4 Discussion

We have introduced a model for studying competitive information provision in search markets, focusing especially on the duopolistic case. Our model allows us to study the impact of experts (information providers) of different quality, and strategic issues in how these experts should price their services. We demonstrate some basic results – e.g, that, with normally distributed noise for both experts, the higher quality expert will be consulted for "more ambiguous" signals, while the lower quality expert will be consulted for ones that are closer to the searcher's original decision thresholds. Using those results, we discuss the characteristics of equilibria under both exogenous and endogenous assumptions on expert quality. Surprisingly, we show that improvements in the quality of the higher quality expert can be Pareto-improving, increasing not just her profit and

the utility of the searcher, but also the profit of the lower quality expert (whose quality does not change). We also characterize the level of switching cost needed in order to "stabilize" the (price, quality) pairs offered by the two experts so that they are in equilibrium.

These results are of much importance to MAS designers, platform owners and regulators that often have control of or can impose constraints on experts and the quality they offer. For example, if the quality of the expert relies on the amount of information it has or its access to databases that are under the control of the MAS designer, then the latter can dictate the set of expert qualities that maximizes user's benefit. Alternatively, the higher quality expert can be paid to improve its quality, in a way that benefits all players.

The plausibility and tractability of our model, with normally distributed noise of different variances as the defining features of experts of different qualities, opens up many interesting questions for further study. For example, extending the model to multiple types of consumers/searchers may lead to some interesting insights. Suppose there were some consumers with noisier signals than others, or some who were pickier than others. Would the dynamics of competition lead to market segmentation where different qualities of information were provided to different types of searchers? Another extension can be to study non-linear pricing in the context of competition. In the endogenous quality case similar conditions of equilibrium should hold. However, in case of predetermined qualities, the searcher's double threshold based strategy (given in Theorem 7) should hold but the complexity of computing the expert strategy increases as the experts have to decide not only the price but also the package size. An important question here is if improving the high quality will still be a Pareto improvment. Moreover, can it also help bridge the revenue gap between the two experts? We believe our model can provide a useful foundation for studying such questions.

Chapter 9

Conclusion

In this thesis we discuss two specific problems related to the algorithmic pricing of digital and information goods which are described below:

The first problem we address is that of maximizing revenue in a setting where the demand of a good being sold is unknown. It is also known as the problem of learning while earning because of the inherent exploration vs. exploitation dilemma (i.e., the seller has to learn the demand and at the same time maximize her revenue) [50]. Due to the rise of e-Commerce, it is easier for the seller to adjust the price of a good being sold in response to its purchase history. We study the revenue maximization problem in the context of pricing digital goods. These goods have unlimited supply and zero marginal cost of production e.g., iTunes apps/songs. In this setting, time proceeds in rounds and in each round a buyer arrives with a unit demand. The buyers are characterized by their valuation which are i.i.d draws from some (static) distribution unknown to the seller [58]. The research problem in this case is to set the price in each round taking into account the selling history of the item so that the infinite horizon discounted profit is maximized (equivalently, the regret is minimized). Recently, there has also been a rising interest in designing strategies which not only minimize regret but also converge to the true optimal price asymptotically.

Our contributions are as follows: (1) we show that under certain settings the myopic policy is Bayes-optimal for the case where, if a buyer decides to buy, she also reveals her valuation. In fact the myopic strategy also has a finite regret which implies that the learning is indeed very fast. We also demonstrate the effectiveness of this strategy on three different real-world datasets comprising bids from multiple eBay auctions for Palm Pilot M515, bids for a particular keyword from a Yahoo! sponsored search auction, and a jester rating dataset consisting of ratings for 100 jokes, when all of these datasets violate the assumptions required for the theoretical optimality. (2) We propose a knowledge gradient (Bayesian one-step-look-ahead) based algorithm and another reinforcement-learning based near-optimal algorithm for the posted-price setting when the buyer does not reveal her valuation in any case to the seller [58]. This implies that the seller has to learn from completely censored information. We evaluate the performance of both of these strategies rigorously on multiple artificial as well as real-world datasets.

There are several interesting extensions of the proposed work. Firstly, can we bound the regret

for the completely censored information scenario while using knowledge gradient strategy (i.e., the buyers do not disclose their valuation under any case)? Secondly, in Bayesian inferencing problems it is often hard to find conjugate priors. One alternative is to find a way to approximate the true posterior distribution with another distribution which has the same form as the prior. We use two such techniques in this thesis. The first is to find an approximation such that the KL divergence between the two distributions is minimized. The second is to find an approximate distribution such that the first two moments are the same as that of the true posterior. We find that the moment-matching based approximation is more appropriate for our settings as it always leads to higher profits in our experiments. However, it is possible that the KL divergence based approximation is more suitable for some other settings. An important future direction would be to compare these two approximations for various distributions and find out what the best suited settings are for each of the them. Another avenue for extending this work is to explore other algorithms for approximating the posterior distribution.

The second problem we focus on in this thesis is that of pricing information-intermediaries (referred to as experts) in a sequential search setting. These experts sell information in order to bridge the knowledge gap between the buyer and the seller. For example, Carfax sells vehicle history reports and can be a useful resource for an individual looking to buy a used car. In a sequential search environment, a searcher receives a stream of opportunities, and she continues examining one opportunity at a time until she accepts one of them. On observing the opportunity, she has to make a decision whether to accept it or reject it and continue her search. In our setting, due to the presence of an expert, the searcher has an additional option to query the expert in order to procure more information about the opportunity and then decide whether she wants to pursue it and terminate her search, or reject it and continue with the search process. In this setup both the searcher and the expert are risk-averse and rational, and choose their expected utility maximizing strategy. The searcher has to devise a strategy to take an optimal action on observing the opportunity, and the problem in front of the expert is to optimally price her services knowing that the searcher is a rational agent. This scenario can be seen as a Stackelberg game where the expert makes a first move by setting the price for her services and the searcher chooses the best strategy given the expert's fee. Our major contributions are: (1) we formulate and solve a monopolist expert's problem of maximizing her revenue when she can choose to offer packages of her services (non-linear pricing e.g., the experts only offers 6 reports for $10 instead of offering a single report) for sale; (2) we study equilibrium competition dynamics when there are two experts of different qualities competing with each other for marketshare; (3) we show that subsidization can significantly improve social welfare, and we also derive the optimal subsidy.

There are several open questions in this area for which our model can provide a good starting point. Firstly, in this thesis, the expert could only offer one of the three pricing schemes: pay-per-use, non-linear pricing, or subscription-based pricing. However, in the real world we see that there are often multiple options for the searcher to choose from. For example, as of January, 2014 Carfax offers all three options: 1 report for $39.99, 5 reports for $44.99 and unlimited reports for a month for $54.99. The current set-up can be a good basis for modeling this problem. Secondly, while studying competitive experts, we have focused on a duopoly; an obvious question is whether we can generalize our results to oligopoly. Thirdly, in case of a duopoly, we place the restriction that the searcher can only consult one of the experts. What if the searcher is allowed to search both the experts? Perhaps after consulting one of the experts, the searcher still wants another opinion to

make better decisions. For example, someone could search Carfax and after looking at the report could want to take it to a mechanic for a second opinion. Taking multiple opinions before making a decision is very common. Another motivating example is that of a sick person who first visits a general physician but after receiving a diagnosis wants to consult a specialist to acquire more specific information. In fact, even in monopolistic markets when the expert is itself noisy, a searcher may want to consult the same expert again e.g., often in case of a medical diagnosis, a doctor may want to repeat some tests when they are not completely accurate. We believe our model can be useful for such studies.

Chapter 10

Bibliography

[1] Yahoo! Webscope, A1 - Yahoo! Search Marketing Advertiser Bidding Data, version 1.0 (81MB), 2014. URL http://webscope.sandbox.yahoo.com.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, volume 36, pages 322–331. IEEE Computer Society Press, 1995.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.

[4] M. Babaioff, S. Dughmi, R. Kleinberg, and A. Slivkins. Dynamic pricing with limited supply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 74–91. ACM, 2012.

[5] J. Bakos. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, 43(12):1676–1692, 1997.

[6] Y. Bakos and E. Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management Science*, 45(12):1613–1630, 1999. ISSN 0025-1909.

[7] M. Balcan, A. Blum, J. Hartline, and Y. Mansour. Mechanism design via machine learning. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 605–614. IEEE, 2005.

[8] R. J. Balvers and T. F. Cosimano. Actively learning about demand and the dynamics of price adjustment. *Economic Journal*, 100(402):882–98, 1990.

[9] B. Becker and T. Milbourn. *Reputation and competition: Evidence from the credit rating industry*. Harvard Business School, 2008.

[10] B. Becker and T. Milbourn. How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3):493–514, 2011.

[11] B. Bental and D. Peled. Quantitative growth effects of subsidies in a search theoretic R&D model. *Journal of Evolutionary Economics*, 12(4):397–423, 2002.

[12] D. Bertsimas and G. Perakis. Dynamic pricing: A learning approach. *Mathematical and Computational Models for Congestion Charging*, pages 45–79, 2005.

[13] O. Besbes and A. Zeevi. On the minimax complexity of pricing in a changing environment. *Operations Research*, 59(1):66–79, 2011.

[14] O. Besbes and A. Zeevi. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Columbia Business School Research Paper*, (12-5), 2012.

[15] A. Blum, V. Kumar, A. Rudra, and F. Wu. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.

[16] M. Brezzi and T. Lai. Incomplete learning from endogenous data in dynamic allocation. *Econometrica*, 68(6):1511–1516, 2000.

[17] J. Broder and P. Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

[18] E. Brynjolfsson and M. Smith. Frictionless commerce? a comparison of Internet and conventional retailers. *Management Science*, 46(4):563–585, 2000.

[19] K. Burdett and R. Wright. Two-sided search with nontransferable utility. *Review of Economic Dynamics*, 1:220–245, 1998.

[20] K. P. Burnham and D. R. Anderson. *Model selection and multi-model inference: a practical information-theoretic approach.* Springer, 2002.

[21] A. Carvalho and M. Puterman. Learning and pricing in an Internet environment with binomial demands. *Journal of Revenue & Pricing Management*, 3(4):320–336, 2005.

[22] A. Chapman, A. Rogers, and N. Jennings. Learn while you earn: two approaches to learning auction parameters in take-it-or-leave-it auctions. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, pages 1561–1564, 2008.

[23] M. Chhabra and S. Das. Learning the demand curve in posted-price digital goods auctions. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 63–70, 2011.

[24] M. Chhabra, S. Das, and D. Sarne. Expert mediated search. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, pages 415–422, 2011.

[25] M. Chhabra, S. Das, and D. Sarne. Expert-mediated sequential search. *European Journal of Operational Research*, 234(3):861–873, 2014.

[26] M. Chhabra, S. Das, and D. Sarne. Competitive information provision in sequential search markets. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*, page Forthcoming, 2014.

[27] V. Conitzer and N. Garera. Learning algorithms for online principal-agent problems (and selling goods online). In *Proceedings of the 23rd International Conference on Machine Learning*, pages 209–216. ACM, 2006. ISBN 1595933832.

[28] E. Cope. Bayesian strategies for dynamic pricing in e-commerce. *Naval Research Logistics (NRL)*, 54(3):265–281, 2006.

[29] T. M. Cover, J. A. Thomas, and J. Kieffer. Elements of information theory. *SIAM Review*, 36 (3):509–510, 1994.

[30] C. Crampes and A. Hollander. Duopoly and quality standards. *European Economic Review*, 39(1):71–82, 1995.

[31] S. Das. The effects of market-making on price dynamics. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, pages 887–894, Estoril, Portugal, May 2008.

[32] S. Das and M. Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *Advances in Neural Information Processing Systems*, pages 361–368, 2008.

[33] M. H. DeGroot. *Optimal Statistical Decisions*. Wiley-Interscience, 2005.

[34] A. den Boer and B. Zwart. Dynamic pricing and learning with finite inventories. Technical report, Technical report, Centrum Wiskunde en Informatica (CWI), Amsterdam, 2011.

[35] J. M. Dimicco, P. Maes, and A. Greenwald. Learning curve: A simulation-based approach to dynamic pricing. *Electronic Commerce Research*, 3(3-4):245–276, 2003.

[36] D. Dranove and G. Jin. Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–63, 2010.

[37] N. S. Economides. A note on equilibrium in price-quality competition. *Greek Economic Review*, 7(2):179l186, 1985.

[38] E. Farhi, J. Lerner, and J. Tirole. Fear of rejection? Tiered certification and transparency. Technical report, National Bureau of Economic Research, 2008.

[39] V. Farias and B. Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

[40] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[41] J. Gallien. Dynamic mechanism design for online commerce. *Operations Research*, 54(2): 291–310, 2006.

[42] X. Geng, M. Stinchcombe, and A. Whinston. Bundling information goods of decreasing value. *Management Science*, 51(4):662–667, 2005.

[43] J. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, 61:35–73, 1966.

[44] J. Gittins and D. Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

[45] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

[46] A. Greenwald, J. Li, and E. Sodomka. Approximating equilibria in sequential auctions with incomplete information and multi-unit demand. In *Advances in Neural Information Processing Systems*, pages 2330–2338, 2012.

[47] A. Grosfeld-Nir, D. Sarne, and I. Spieglier. Modeling the search for the least costly opportunity. *European Journal of Operational Research*, 197(2):667–674, 2009.

[48] S. Gupta and K. Miescke. Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of statistical planning and inference*, 54(2):229–244, 1996.

[49] J. Harrison, N. Keskin, and A. Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Submitted for publication*, 2013.

[50] J. M. Harrison, N. Keskin, and A. Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586, 2012.

[51] M. Haynes and S. Thompson. Price, price dispersion and number of sellers at a low entry cost shopbot. *International Journal of Industrial Organization*, 26(2):459–472, March 2008.

[52] D. Hennes, D. Claes, W. Meeussen, and K. Tuyls. Multi-robot collision avoidance with localization uncertainty. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 147–154, 2012.

[53] J. Hey and C. McKenna. Consumer search with uncertain product quality. *The Journal of Political Economy*, pages 54–66, 1981.

[54] A. Janetos. Strategies of female mate choice: A theoretical analysis. *Behavioral Ecology and Sociobiology*, 7(2):107–112, 1980.

[55] W. Jank and G. Shmueli. *Modeling online auctions*, volume 91. John Wiley & Sons, 2010. URL http://www.modelingonlineauctions.com/datasets.

[56] J. O. Kephart, J. E. Hanson, and A. R. Greenwald. Dynamic pricing by software agents. *Computer Networks*, 32:731–752, 2000.

[57] R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, volume 17, pages 697–704, 2005.

[58] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for on-line posted-price auctions. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science*, pages 594–605, 2003.

[59] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[60] M. Lariviere and E. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45(3):346–363, 1999.

[61] B. Leloup and L. Deveaux. Dynamic pricing on the internet: Theory and simulations. *Electronic Commerce Research*, 1(3):265–276, 2001.

[62] C. Lim, J. Bearden, and J. Smith. Sequential search with multiattribute options. *Decision Analysis*, 3(1):3–15, 2006.

[63] S. Lippman and J. McCall. The economics of job search: A survey. *Economic Inquiry*, 14(2): 155–189, 1976.

[64] M. S. Lobo and S. Boyd. Pricing and learning with uncertain demand. *working paper*, 2003.

[65] J. MacQueen. Optimal policies for a class of search and evaluation problems. *Management Science*, 10(4):746–759, 1964.

[66] J. B. MacQueen. Optimal policies for a class of search and evaluation problems. *Management Science*, 10(4):746–759, 1964.

[67] J. McCall. Economics of information and job search. *The Quarterly Journal of Economics*, 84 (1):113–126, 1970.

[68] K. McCardle. Information acquisition and the adoption of new technology. *Management Science*, 31(11):1372–1389, 1985.

[69] A. McLennan. Price dispersion and incomplete learning in the long run. *Journal of Economic Dynamics and Control*, 7(3):331–347, 1984.

[70] J. McMillan and M. Rothschild. Search. In *Handbook of Game Theory with Economic Applications*, pages 905–927. North-Holland, 1994.

[71] A. Mersereau, P. Rusmevichientong, and J. Tsitsiklis. A structured multiarmed bandit problem and the greedy policy. *IEEE Transactions on Automatic Control*, 54(12):2787–2802, 2009.

[72] P. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12:380–391, 1981.

[73] G. Monahan. Optimal stopping in a partially observable Markov process with costly information. *Operations Research*, 28(6):1319–1334, 1980.

[74] G. Monahan. Optimal stopping in a partially observable binary-valued Markov chain with costly perfect information. *Journal of Applied Probability*, pages 72–81, 1982.

[75] D. Mortensen and C. Pissarides. New developments in models of search in the labor market. *Handbook of Labor Economics*, 3:2567–2627, 1999.

[76] G. Moscarini and L. Smith. The optimal level of experimentation. *Econometrica*, 69(6): 1629–1644, 2003.

[77] M. Motta. Endogenous quality choice: price vs. quantity competition. *The Journal of Industrial Economics*, pages 113–131, 1993.

[78] W. Powell and I. Ryzhov. *Optimal Learning*, volume 841. Wiley, 2012.

[79] H. Qu, I. O. Ryzhov, and M. C. Fu. Ranking and selection with unknown correlation structures. In *Proceedings of the Winter Simulation Conference*, pages 1–12, 2012.

[80] M. Rothschild. Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy*, 82(4):689–711, 1974.

[81] M. Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9 (2):185–202, 1974.

[82] A. Rubinstein and A. Wolinsky. Middlemen. *The Quarterly Journal of Economics*, 102(3): 581–593, 1987.

[83] I. Ryzhov, P. Frazier, and W. Powell. On the robustness of a one-period look-ahead policy in multi-armed bandit problems. *Procedia Computer Science*, 1(1):1635–1644, 2010.

[84] I. Ryzhov, W. Powell, and P. Frazier. The knowledge gradient algorithm for a general class of online learning problems. *Operations Research*, 60(1):180–195, 2012.

[85] B. C. Smith, D. P. Günther, B. V. Rao, and R. M. Ratlife. E-commerce and operations research in airline planning, marketing, and distribution. *Interfaces*, 31(2):37–55, 2001.

[86] A. Sundararajan. Nonlinear pricing of information goods. *Management Science*, 50(12):1660–1673, 2004.

[87] K. T. Talluri and G. J. Van Ryzin. *The Theory and Practice of Revenue Management*. Springer, 2005.

[88] C. Terwiesch and C. H. Loch. Collaborative prototyping and the pricing of custom-designed products. *Management Science*, 50(2):145–158, 2004.

[89] J. Vermorel and M. Mohri. Multi-armed bandit algorithms and empirical evaluation. In *Proceedings of the 16th European conference on Machine Learning*, pages 437–448. Springer, 2005.

[90] M. Weitzman. Optimal search for the best alternative. *Econometrica*, 47(3):641–654, 1979.

[91] D. Wiegmann, K. Weinersmith, and S. Seubert. Multi-attribute mate choice decisions and uncertainty in the decision process: A generalized sequential search strategy. *Journal of Mathematical Biology*, 60(4):543–572, 2010.

[92] R. Wright. Job search and cyclical unemployment. *The Journal of Political Economy*, 94(1): 38–55, 1986.

Appendix A

Pricing of a Digital Good

A.1 Partially Censored Case

A.1.1 Myopic Strategy

We compute the expected profits for the two possible values of M separately and choose the one which yields higher profits.

Case 1: $M = \max(b, q_t) = b$:

We calculate the myopically optimal price for this setting by equating the first derivative of the expected profit to zero.

$$\mathbb{E}(\pi_t) = q_t(1 - F_B(M; a, b)) - \frac{q_t^2 a(1 - F_B(M; a + 1, b))}{b(a + 1)}$$

$$\mathbb{E}(\pi_t) = q_t - \frac{q_t^2 a}{b(a + 1)}$$

$$\Rightarrow \frac{d\mathbb{E}(\pi_t)}{dq_t} = 1 - \frac{2q_t a}{b(a + 1)}$$

Setting the first derivative to zero we get: $q_t = \dfrac{b(a + 1)}{2a}$

$$\Rightarrow \frac{d^2\mathbb{E}(\pi_t)}{dq_t^2} = -\frac{2a}{b(a + 1)} < 0 \text{ condition for maximum is satisfied}$$

The myopic profit in this setting is $\dfrac{b(a + 1)}{4a}$

Case 2: $M = \max(b, q_t) = q_t$

$$\mathbb{E}(\pi_t) = q_t(1 - F_B(M; a, b)) - \frac{q_t^2 a(1 - F_B(M; a+1, b))}{b(a+1)}$$

$$= q_t(1 - F_B(q_t; a, b)) - \frac{q_t^2 a(1 - F_B(q_t; a+1, b))}{b(a+1)}$$

$$= q_t \left(\frac{b}{q_t}\right)^a - \frac{a q_t^2}{b(a+1)} \left(\frac{b}{q_t}\right)^{a+1}$$

$$= \frac{b^a}{q_t^{a-1}} - \frac{a b^a}{(a+1) q_t^{a-1}}$$

$$= \frac{b^a}{q_t^{a-1}(a+1)}$$

For $a > 1$, $\mathbb{E}(\pi_t)$ decreases with increase in q_t, therefore, the myopic price is b. Since $a < 1$ implies that the mean of the Pareto distribution is infinite and we assume that the expected value of Z is finite, we do not have to consider the case $a < 1$.

The myopically optimal expected profit for $b < q_t$ is $\frac{b}{a+1}$. The profit $\left(\frac{b}{a+1}\right)$ under this condition is less than the maximum expected profit when $M = \max(b, q_t) = b$ (i.e., $\frac{b(a+1)}{4a}$) . Therefore, a myopic seller would always prefer to price the item $q_{t_{\text{myopic}}} = \frac{b(a+1)}{2a}$ which would yield an expected profit $\mathbb{E}(\pi_{t_{\text{myopic}}}) = \frac{b(a+1)}{4a}$.

A.1.2   Knowledge-Gradient Strategy

Similar to the calculation of myopic profit we have to consider two cases $b > q_t$ and $b < q_t$ separately. We already calculated the expected KG profit when $b > q_t$ in Section 3.4. Here we will only consider the case $b \leq q_t$.

We first consider each term in mentioned Equation 3.3 separately:

1. $\mathbb{E}(\pi_t)$: When $q_t > b$

$$\mathbb{E}(\pi_t) = \frac{b^a}{q_t^{a-1}(a+1)} \qquad \text{(see Section A.1.1 for details)} \tag{A.1}$$

   This expression is clearly decreasing in q_t when $a > 1$.

2. $\Pr(\text{Buy}|q_t) \max(\mathbb{E}(\pi_{t+1|\text{Buy}}))$: This expression can be calculated as follows:

$$\Pr(\text{Buy}|q_t) \max(\mathbb{E}(\pi_{t+1|\text{Buy}})) = \int_{q_t}^{\infty} dx\, f_Z(x) \int_{q_t}^{x} dv\, f_v(v|x) \frac{\max(b,v)(a+2)}{4(a+1)}$$

$$= \int_{q_t}^{\infty} dx\, f_Z(x) \int_{q_t}^{x} dv\, \frac{v(a+2)}{4(a+1)x}$$

$$= \int_{q_t}^{\infty} dx\, f_Z(x) \frac{(x^2 - q_t^2)(a+2)}{8(a+1)x}$$

$$= \frac{a+2}{8(a+1)} \left(\int_{q_t}^{\infty} dx\, f_Z(x)(x - q_t^2/x) \right)$$

$$= \frac{a+2}{8(a+1)} \left(\frac{ab}{a-1}(1 - F_Z(q_t; a-1, b)) - \frac{q_t^2 a}{b(a+1)}(1 - F_Z(q_t; a+1, b)) \right)$$

$$= \frac{ab(a+2)}{4(a-1)(a+1)^2} \left(\frac{b}{q_t} \right)^{a-1} \tag{A.2}$$

This is also decreasing in q_t for $a > 1$.

3. $Pr(\neg\text{Buy}|q_t) \max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))$: For $q_t > b$ in case of a no buy the posterior distribution is not Pareto, so we need to calculate the $\mathbb{E}(\pi_{t+1}|\neg\text{Buy})$ according to the new posterior distribution.

Let q_{t+1} represent the price charged at time $t+1$. We calculate $\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))$ for the following three cases:

(a) **Case 1:** $q_{t+1} < b < q_t$

$$\Pr_{t+1}(\text{Buy}|q_t, q_{t+1}) = \int_{b}^{\infty} \frac{x - q_{t+1}}{x} f_B(x|\neg\text{Buy})\, dx$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)} \left(1 - \frac{b^a}{(a+1)q_t^a} + q_{t+1} \left(\frac{ab^a}{q_t^{a+1}(a+1)(a+2)} - \frac{a}{b(a+1)} \right) \right)$$

In the KG technique we assume that from the next period we play the myopic strategy. Therefore we want to calculate the myopically optimal q_{t+1} (i.e, choose a q_{t+1} which maximizes $(q_{t+1}\Pr_{t+1}(\text{Buy}|q_{t+1}, q_t))$).

After some simple calculations, the myopic price at $t+1$ (q_{t+1}) is:

$$q_{t+1}^* = \frac{1 - \frac{b^a}{q_t^a(a+1)}}{2\left(\frac{a}{b(a+1)} - \frac{ab^a}{q_t^{a+1}(a+1)(a+2)} \right)}$$

The expected profit at the myopic price calculated above is:

$$\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}})) = \frac{1}{\Pr(\neg\text{Buy}|q_t)} \frac{\left(1 - \frac{b^a}{q_t^a(a+1)} \right)^2}{4\left(\frac{a}{b(a+1)} - \frac{ab^a}{q_t^{a+1}(a+1)(a+2)} \right)}$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)} \frac{(a+1)b}{4a} \frac{\left(1 - \frac{b^a}{q_t^a(a+1)} \right)^2}{\left(1 - \frac{b^{a+1}}{q_t^{a+1}(a+2)} \right)}$$

Therefore,

$$\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t) = \frac{b(a+1)}{4a}\frac{\left(1-\frac{b^a}{q_t^a(a+1)}\right)^2}{\left(1-\frac{b^{a+1}}{q_t^{a+1}(a+2)}\right)} \tag{A.3}$$

Using Equation A.3 and A.2 we calculate the expected profit at time $t+1$:

$$\mathbb{E}(\pi_{t+1_{\text{Myopic}}}) = \max(\mathbb{E}(\pi_{t+1|\text{Buy}}))\Pr(\text{Buy}|q_t) + \max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t)$$

$$= \frac{ab(a+2)b^{a-1}}{4(a-1)(a+1)^2 q^{a-1}} + \frac{b(a+1)}{4a}\frac{\left(1-\frac{b^a}{q_t^a(a+1)}\right)^2}{\left(1-\frac{b^{a+1}}{q_t^{a+1}(a+2)}\right)}$$

$$\frac{d\mathbb{E}(\pi_{t+1_{\text{KG}}})}{dq_t} = -b\left(\frac{(\frac{b}{q})^a(a+2)\left(b-b(\frac{b}{q})^a+2ab-2aq+a^2b-a^2q\right)^2}{4ab(a+1)^2\left(2q-b(\frac{b}{q})^a+aq\right)^2}\right)$$

$$\leq 0$$

Since the derivative of $\mathbb{E}(\pi_{t+1_{\text{Myopic}}})$ w.r.t to q_t is always negative, $\mathbb{E}(\pi_{t+1_{\text{Myopic}}})$ is decreasing in q_t. Given $\mathbb{E}(\pi_{t+1_{\text{Myopic}}})$ and $\mathbb{E}(\pi_t)$ are both decreasing in q_t, $\mathbb{E}(\pi_{KG})$ is also a decreasing function of q_t and is maximized at $q_t = b$ as $q_t \geq b$.

$$\mathbb{E}(\pi_{KG}) = \max_{q_t}\left(\mathbb{E}(\pi_t) + \frac{\delta}{1-\delta}\mathbb{E}(\pi_{t+1_{KG}})\right)$$

$$= \frac{b}{a+1} + \frac{\delta}{1-\delta}\left(\frac{ab(a+2)}{4(a+1)^2} + \frac{ab(a+2)}{4(a-1)(a+1)^2}\right)$$

$$= \frac{b}{a+1} + \frac{\delta}{1-\delta}\left(\frac{a^2b(a+2)}{4(a-1)(a+1)^2}\right) \tag{A.4}$$

(b) **Case 2:** $b < q_t < q_{t+1}$

In this case the probability of buying a good at time $t+1$ is:

$$\Pr_{t+1}(\text{Buy}|q_t, q_{t+1}) = \int_{q_{t+1}}^{\infty} \frac{x-q_{t+1}}{x}\frac{q_t f_B(x)}{x\Pr(\neg\text{Buy}|q_t)}\,dx$$

$$= \frac{q_t a}{b(a+1)(a+2)\Pr(\neg\text{Buy}|q_t)}\left(\frac{b}{q_{t+1}}\right)^{a+1}$$

$$\mathbb{E}(\pi_{t+1|\neg\text{Buy}}) = q_{t+1}\Pr_{t+1}(\text{Buy}|q_{t+1}, q_t)$$

$$= \frac{q_t a}{(a+1)(a+2)\Pr(\neg\text{Buy}|q_t)}\left(\frac{b}{q_{t+1}}\right)^{a}$$

$$\mathbb{E}(\pi_{t+1|\neg\text{Buy}})\Pr(\neg\text{Buy}|q_t) = \frac{q_t a}{(a+1)(a+2)}\left(\frac{b}{q_{t+1}}\right)^{a} \tag{A.5}$$

As Equation A.5 is decreasing in q_{t+1}, it is maximized when $q_{t+1} = q_t$. Substituting $q_{t+1} = q_t$ we get:

$$\Pr(\neg\text{Buy}|q_t)\max_{q_{t+1}}(\mathbb{E}(\pi_{t+1|\neg\text{Buy}})) = \frac{q_t a}{(a+1)(a+2)}\left(\frac{b}{q_t}\right)^a \tag{A.6}$$

For this case $\mathbb{E}(\pi_{KG})$ can be calculated using Equations A.6, A.2 and A.1:

$$\mathbb{E}(\pi_{KG}) = \max_{q_t}\left(\frac{b^a}{q_t^{a-1}(a+1)} + \frac{\delta}{1-\delta}\left(\frac{q_t a}{(a+1)(a+2)}\left(\frac{b}{q_t}\right)^a + \frac{ab(a+2)}{4(a+1)^2(a-1)}\left(\frac{b}{q_t}\right)^a\right)\right)$$

Since the right hand side of the above expression is decreasing in q_t, it is maximized when $q_t = b$. The KG profit is equal to: $\mathbb{E}(\pi_{KG})$ is decreasing in q_t, hence it is maximized when

$$\mathbb{E}(\pi_{KG}) = \frac{b}{a+1} + \frac{\delta}{1-\delta}\left(\frac{ab}{(a+1)(a+2)} + \frac{ab(a+2)}{4(a-1)(a+1)^2}\right) \tag{A.7}$$

(c) **Case 3:** $b < q_{t+1} < q_t$ The probability of buying a good at time $t+1$ is:

$$\Pr_{t+1}(\text{Buy}|q_t, q_{t+1}) = \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(\int_{q_{t+1}}^{q_t}\frac{x-q_{t+1}}{x}\frac{f_B(x)}{}dx + \int_{q_t}^{\infty}\frac{x-q_{t+1}}{x}\frac{q_t f_B(x)}{x}dx\right)$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(\int_{q_{t+1}}^{q_t} f_B(x)\,dx - \int_{q_{t+1}}^{q_t}\frac{q_{t+1}f_B(x)}{x}dx\right.$$

$$\left. + \int_{q_t}^{\infty}\frac{q_t f_B(x)}{x}dx - \int_{q_t}^{\infty}\frac{q_{t+1}q_t f_B(x)}{x^2}dx\right)$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(F_B(q_t) - F_B(q_{t+1}) - \frac{q_{t+1}a}{b(a+1)}(F_B(q_t;a+1) - F_B(q_{t+1};a+1))\right.$$

$$\left. + \frac{q_t a}{b(a+1)}(1 - F_B(q_t;a+1)) - \frac{q_{t+1}q_t a}{b^2(a+2)}(1 - F_B(q_t;a+2))\right)$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(\left(\frac{b}{q_{t+1}}\right)^a - \left(\frac{b}{q_t}\right)^a - \frac{q_{t+1}a}{b(a+1)}\left(\left(\frac{b}{q_{t+1}}\right)^{a+1} - \left(\frac{b}{q_t}\right)^{a+1}\right)\right.$$

$$\left. + \frac{a}{a+1}\left(\frac{b}{q_t}\right)^a - \frac{q_{t+1}a}{b(a+2)}\left(\frac{b}{q_t}\right)^{a+1}\right)$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(\frac{1}{a+1}\left(\frac{b}{q_{t+1}}\right)^a - \frac{1}{a+1}\left(\frac{b}{q_t}\right)^a + \frac{b^a aq}{q_t^{a+1}}\left(\frac{1}{a+1} - \frac{1}{a+2}\right)\right)$$

$$= \frac{1}{\Pr(\neg\text{Buy}|q_t)}\left(\frac{1}{a+1}\left(\frac{b}{q_{t+1}}\right)^a - \frac{1}{a+1}\left(\frac{b}{q_t}\right)^a + \frac{b^a aq_{t+1}}{q_t^{a+1}(a+1)(a+2)}\right)$$

$$\mathbb{E}(\pi_{t+1|\neg\text{Buy}}) = q_{t+1}\Pr_{t+1}(\text{Buy}|q_{t+1}, q_t)$$

$$= \frac{q_{t+1}}{\Pr(\neg\text{Buy}|q_t)}\left(\frac{1}{a+1}\left(\frac{b}{q_{t+1}}\right)^a - \frac{1}{a+1}\left(\frac{b}{q_t}\right)^a + \frac{b^a aq_{t+1}}{q_t^{a+1}(a+1)(a+2)}\right) \tag{A.8}$$

To calculate the profit maximizing q_{t+1}, we first calculate the derivative of $\mathbb{E}(\pi_{t+1|\neg\text{Buy}})$ (given in Equation A.8) w.r.t q_{t+1}.

$$
\begin{aligned}
\frac{d\mathbb{E}(\pi_{t+1|\neg\text{Buy}})}{dq_{t+1}} &= \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(-\frac{(a-1)}{q_{t+1}^a}-\frac{1}{q_t^a}+\frac{2aq_{t+1}}{q_t^{a+1}(a+2)}\right)\\
&\leq \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(-\frac{a-1}{q_{t+1}^a}-\frac{1}{q_t^a}+\frac{2a}{(a+2)q_t^a}\right) \quad q_{t+1} < q_t\\
&\leq \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(-\frac{a-1}{q_{t+1}^a}+\frac{a-2}{q_t^a(a+2)}\right)\\
&\leq \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(-\frac{a-1}{q_{t+1}^a}+\frac{a-1}{q_t^a(a+2)}\right)\\
&\leq \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(-\frac{a-1}{q_{t+1}^a}+\frac{a-1}{q_t^a}\right)\\
&\leq 0 \quad 1/q_{t+1} > 1/q_t
\end{aligned}
$$

As $\frac{d\mathbb{E}(\pi_{t+1|\neg\text{Buy}})}{dq_{t+1}}$ is always less than 0, $\mathbb{E}(\pi_{t+1|\neg\text{Buy}})$ is a decreasing function of q_{t+1} and hence is maximized when $q_{t+1} = b$:

$$
\begin{aligned}
\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}})) &= \frac{b^a}{\Pr(\neg\text{Buy}|q_t)(a+1)}\left(\frac{1}{b^{a-1}}-\frac{b}{q_t^a}+\frac{ab^2}{q_t^{a+1}(a+2)}\right)\\
\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t) &= \frac{b^a}{(a+1)}\left(\frac{1}{b^{a-1}}-\frac{b}{q_t^a}+\frac{ab^2}{q_t^{a+1}(a+2)}\right)
\end{aligned}
$$

$\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t)$ is an increasing function of q_t, we will evaluate the $\mathbb{E}(\pi_{t+1_{\text{Myopic}}})$ and its derivative w.r.t to q_t.

$$
\begin{aligned}
\mathbb{E}(\pi_{t+1_{\text{Myopic}}}) &= \max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t) + \max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}}))\Pr(\neg\text{Buy}|q_t)\\
&= \frac{b^a}{(a+1)}\left(\frac{1}{b^{a-1}}-\frac{b}{q_t^a}+\frac{ab^2}{q_t^{a+1}(a+2)}\right)+\frac{ab(a+2)}{4(a-1)(a+1)^2}\left(\frac{b}{q_t}\right)^{a-1}\\
\frac{d\mathbb{E}(\pi_{t+1_{\text{Myopic}}})}{dq_t} &= \frac{b^a}{q_t^{a+1}}\left(\frac{ab}{q_t^{a+1}}-\frac{ab^2(a+1)}{q_t^{a+2}(a+2)}-\frac{a(a+2)}{4(a+1))q_t^a}\right)\\
&= -\frac{b^a a}{q_t^a(a+2)}\left(\frac{b^2}{q_t^2}-\frac{(a+2)b}{(a+1)q_t}+\frac{(a+2)^2}{4(a+1)}\right)\\
&= -\frac{ab^a}{(a+2)q_t^a}\left(\frac{b}{q_t}-\frac{a+2}{2(a+1)}\right)^2\\
&\leq 0
\end{aligned}
$$

Since $\frac{d\mathbb{E}(\pi_{t+1_{\text{Myopic}}})}{dq_t}$ is negative, $\mathbb{E}(\pi_{t+1_{\text{Myopic}}})$ is a decreasing function of q_t and is maximized when $q_t = b$. We know from Equation A.1 that $\mathbb{E}(\pi_t)$ is also a decreasing function of q_t when $q_t > b$. Therefore, the right hand side of Equation 3.3 is decreasing in q_t and

is maximized when $q_t = b$. Thus, the KG optimal price $q_{KG} = b$. The value of $\mathbb{E}(\pi_{KG})$ in this case is:

$$\mathbb{E}(\pi_{KG}) = \max_{q_t}(\mathbb{E}(\pi_t) + \frac{\delta}{1-\delta}\mathbb{E}(\pi_{t+1_{KG}}))$$

$$= \frac{b}{a+1} + \frac{\delta}{1-\delta}\left(\frac{ab}{(a+1)(a+2)} + \frac{ab(a+2)}{4(a-1)(a+1)^2}\right) \tag{A.9}$$

Out of the three possible values of q_{t+1}, the seller will choose the one that maximizes the expected myopic profit at time $t+1$ depending on the value of other parameters. If we compare the KG profit for all the three cases given in Equations A.4, A.7 and A.9, A.4 yields the highest revenue, therefore, the seller will always choose the $q_{t+1} < b$.

To find the KG optimal q_t, we compare the expected maximum revenue for both cases — $b > q_t$ and $b \leq q_t$. The maximum expected profit for the case $b \leq q_t$ is given by Equation A.4 and for $b > q_t$ it is given by Equation 3.6. Clearly, the profit is greater when $b > q_t$, therefore, it never makes sense for the seller to price the item greater than b.

A.2 Completely Censored Case

A.2.1 Functional Distance

Let $F(x) = \frac{x}{Z}$ represent the cdf of uniform distribution over the interval $[0, Z]$ and $G(x)$ be the cdf be the actual valuation distribution. L^2-Metric between the two distributions is given by:

$$f_d = \sqrt{\int_0^\infty (F(x) - G(x))^2 \, dx}$$

For convenience we consider $D = f_d{}^2$, written as

$$D = f_d{}^2 = \int_0^\infty ((1 - G(x)) - (1 - F(x)))^2 \, dx$$

Let $F_1(x) = 1 - F(x)$ and $G_1(x) = 1 - G(x)$. Then

$$D = \int_0^Z F_1^2(x) \, dx - 2\int_0^Z G_1(x)F_1(x) \, dx + \int_0^\infty G_1^2(x) \, dx$$

$$= \frac{Z}{3} - 2\int_0^Z G_1(x)F_1(x) \, dx + \int_0^\infty G_1^2(x) \, dx$$

differentiating w.r.t Z and setting to 0 to calculate minima, we find

$$\frac{1}{3} - 2\int_0^Z \frac{xG_1(x)}{Z^2} \, dx = 0$$

This equation can easily be solved numerically for $G(x)$ exponential and lognormal respectively, and the solution obtained should be verified for maximum by checking if $\frac{d^2D}{dZ^2} > 0$.

Appendix B

Pricing Information Services in One-sided Search

B.1   Proof of Theorem 4

We show that, optimally, if the searcher should resume her search given a signal s, then she must necessarily also do so given any signal $s' < s$. Let V denote the expected utility to the searcher upon resuming the search if signal s is obtained. Since the optimal strategy given signal s is to resume search, we know $V > E[v|s]$. From the MLRP assumption, $E[v|s] \geq E[v|s']$ for $s' < s$. Therefore, $V > E[v|s']$, proving that the optimal strategy is reservation-value and search should be resumed for any $s' < s$. Then, the expected utility when using reservation signal t is given by:

$$V(t) = -c_s + V(t) \int_{s=-\infty}^{t} f_s(s)\,ds + \int_{s=t}^{\infty} E[v|s]f_s(s)\,ds = -c_s + V(t)F_s(t) + \int_{s=t}^{\infty} E[v|s]f_s(s)\,ds \quad \text{(B.1)}$$

where $F_s(s)$ is the cumulative distribution function of the signal s. Taking the first derivative with respect to t on both sides and then setting it to 0 at t^*, we get:

$$\frac{dV(t)}{dt} = \frac{dV(t)}{dt}F_s(t) + V(t)f_s(t) - E[v|t]f_s(t) = 0$$

And since $\frac{dV(t)}{dt} = 0$ for $t = t^*$ we obtain:

$$V(t^*) = E[v|t^*] \qquad f_s(t^*) \neq 0 \tag{B.2}$$

The condition $V(t^*) = E[v|t^*]$ implies that the reservation value t^* is the signal for which the searcher's utility of resuming search is equal to the expected value of the opportunity associated with that signal.

To verify that $V(t)$ reaches its maximum for t^*, we calculate the second derivative.

$$\frac{d^2V}{dt^2} = F_s(t)\frac{d^2V(t)}{dt^2} + f_s(t)\frac{dV(t)}{dt} + \frac{d(V(t)f_s(t))}{dt} - \frac{d(E[v|t]f_s(t))}{dt}$$

$$\frac{d^2V}{dt^2}(1 - F_s(t)) = f_s(t)\frac{dV(t)}{dt} + \frac{d((V(t) - E[v|t])f_s(t))}{dt}$$

$$\frac{d^2V}{dt^2}(1 - F_s(t)) = f_s(t)\frac{dV(t)}{dt} + \frac{df_s(t)}{dt}(V(t) - E[v|t]) + f_s(t)\frac{d(V(t) - E[v|t])}{dt}$$

At $t = t^*$, $\frac{d(V(t))}{dt}|t^* = 0$ and $V(t^*) = E(v|t^*)$, therefore substituting these values in the above equation, we get:

$$\frac{d^2V}{dt^2}(1 - F_s(t^*)) = 0 + 0 + 0 - f_s(t^*)\frac{dE[v|t])}{dt}|t^* = \frac{d^2V}{dt^2} = -\frac{f_s(t^*)}{1 - F_s(t^*)}\left(\frac{dE[v|t]}{dt}|t^*\right)$$

Given the MLRP assumption, $E[v|t]$ is an increasing function in t. Therefore $\frac{dE[v|t]}{dt} > 0$, hence the value of the second derivative evaluated at t^* is less than zero, which confirms that t^* indeed yields a maximum.

Finally substituting $V(t^*) = E[V|t^*]$ in (B.1) we obtain Equation 5.3 as follows:

$$E[v|t^*] = -c_s + E[v|t^*]\int_{s=-\infty}^{t^*}f_s(s)\,ds + \int_{s=t^*}^{\infty}E[v|s]f_s(s)\,ds \Rightarrow c_s = \int_{s=t^*}^{\infty}(E[v|s] - E[v|t^*])f_s(s)\,ds$$

The value t^* can be calculated using the above equation. $\square$

B.2 Calculation of $\frac{\partial V_\gamma}{\partial t_l}$ and $\frac{\partial^2 V_\gamma}{\partial t_l^2}$

From Equation 6.3:

$$V_{\gamma>0} = V_\gamma F_s(t_\gamma) + \int_{s=t_\gamma}^{\infty} f_s(s)\left(\int_{y=V_{\gamma-1}}^{\infty} y f_v(y|s)\,dy + V_{\gamma-1}F_v(V_{\gamma-1}|s)\right)ds - c_s$$

$$\frac{\partial V_\gamma}{\partial t_l}(1 - F_s(t_\gamma)) = \frac{\partial V_{\gamma-1}}{\partial t_l}\int_{t_\gamma}^{\infty} f_s(s)F_v(V_{\gamma-1}|s)\,ds$$

$$= \frac{1}{1 - F_s(t_\gamma)}\frac{\partial V_{\gamma-1}}{\partial t_l}\int_{t_\gamma}^{\infty} f_s(s)F_v(V_{\gamma-1}|s)\,ds$$

$$= \frac{\Pr(v \le V_{\gamma-1} \wedge s \ge t_\gamma)}{\Pr(s \ge t_\gamma)}\frac{\partial V_{\gamma-1}}{\partial t_l}$$

$$= \Pr(v \le V_{\gamma-1}|s \ge t_\gamma)\frac{\partial V_{\gamma-1}}{\partial t_l}$$

$$= \prod_{i=1}^{\gamma}\Pr(v \le V_{i-1}|s \ge t_i)\frac{\partial V_0}{\partial t_l}$$

$$\frac{\partial^2 V_\gamma}{\partial t_l^2} = \prod_{i=1}^{\gamma}\Pr(v \le V_{i-1}|s \ge t_i)\frac{\partial^2 V_0}{\partial t_l^2} + \frac{\partial}{\partial t_l}\left(\prod_{i=1}^{\gamma}\Pr(v \le V_{i-1}|s \ge t_i)\right)\frac{\partial V_0}{\partial t_l}$$

B.3 Proof of Theorem 5

Lemma 1 *In state $\gamma > 0$, querying the expert dominates terminating search without querying the expert*

Proof: For any strategy that terminates search upon obtaining a signal s when in state $\gamma > 0$, consider instead a modification of that strategy which queries the expert, and terminates only if the true value v, revealed by the expert, is greater than $V_{\gamma-1}$. This new strategy dominates the one it is based on, since if the true value is less than $V_{\gamma-1}$ the searcher is better off resuming search, and she pays no marginal cost to obtain the expert's service in this instance. $\qquad\square$

Proof of Theorem 5: For Part (1), we first show that if, optimally, the searcher should resume her search given a signal s, then she must also do so given any signal $s' < s$. Then, we show that if, optimally, the searcher should terminate her search given a signal s, then she must also do so given any signal $s'' > s$.

If the optimal strategy given signal s is to resume search then the following two inequalities hold, describing the superiority of resuming search over terminating search (B.3) and querying the expert (B.4):

$$V_0 > E[v|s] \quad \text{(B.3)} \qquad\qquad V_0 > M(s, V_{k-1}) - c_e^k \quad \text{(B.4)}$$

Given the MLRP assumption and since $s' < s$, (B.3) holds also for s'. Similarly, notice that:

$$
\begin{aligned}
M(s, V_{k-1}) &= \int_{y=V_{k-1}}^{\infty} y f_v(y|s)\, dy + V_{k-1} F_v(V_{k-1}|s) \\
&= \left[y F_v(y|s) \right]_{y=V_{k-1}}^{\infty} - \int_{y=V_{k-1}}^{\infty} F_v(y|s)\, dy + V_{k-1} F_v(V_{k-1}|s) \\
&= \int_{y=V_{k-1}}^{\infty} (1 - F_v(y|s))\, dy + V_{k-1} > M(s', V_{k-1}) \quad \forall s' < s, \quad \text{due to MLRP.} \quad \text{(B.5)}
\end{aligned}
$$

Consequently (B.4) holds also for s', proving 1(a).

The proof for $s'' > s$ is exactly equivalent: the expected utility of accepting the current opportunity can be shown to dominate both resuming the search and querying the expert, proving 1(b) and thus (together with 1(a)) also 1(c). The optimal strategy when in state $\gamma = 0$ can thus be described by the tuple (t_l, t_u, V_{k-1}) as stated in the theorem and consequently calculated as the Bellman Equation captured by Equation 6.4 for the case of $\gamma = 0$. Taking the derivative of Equation 6.4 w.r.t. t_l and equating to zero, we obtain a unique t_l which maximizes the expected utility.

$$
\begin{aligned}
\frac{\partial V_0}{\partial t_l} =\,& V_0 f_s(t_l) + F_s(t_l) \frac{\partial V_0}{\partial t_l} - f_s(t_l) V_{k-1} F_v(V_{k-1}|t_l) + \int_{t_l}^{t_u} f_s(s) \frac{\partial (V_{k-1} F_v(V_{k-1}|s))}{\partial t_l} \\
& - f_s(t_l) \int_{V_{k-1}}^{\infty} y f_v(y|t_l)\, dy - \int_{t_l}^{t_u} f_s(s) V_{k-1} f_v(V_{k-1}|s) \frac{\partial V_{k-1}}{\partial t_l}\, ds + c_e^k f_s(t_l)
\end{aligned}
$$

where $F_s(s)$ is the cumulative distribution function of the signal s. After some algebraic manipulations, we get:

$$
\frac{\partial V_0}{\partial t_l}(1 - F_s(t_l)) = f_s(t_l) \left(V_0 - V_{k-1} - \int_{V_{k-1}}^{\infty} (1 - F_v(y|t_l))\, dy + c_e^k \right) + \frac{\partial V_{k-1}}{\partial t_l} \int_{t_l}^{t_u} f_s(s) F_v(V_{k-1}|s)\, ds
$$

$$\text{(B.6)}$$

Substituting $\frac{\partial V_0}{\partial t_l} = 0$ and $\frac{\partial V_{k-1}}{\partial t_l} = 0$[1] in (B.6), we obtain Equation 6.5: $c_e^k = \int_{V_{k-1}}^{\infty} (1 - F_v(y|t_l))\, dy + V_{k-1} - V_0 = V_{k-1} - V_0 + \int_{V_{k-1}}^{\infty} (y - V_{k-1}) f_v(y|t_l)\, dy$.

To confirm that this is a maximum, we compute the second derivative of V_0 w.r.t t_l:

$$
\begin{aligned}
\frac{\partial^2 V_0}{\partial t_l^2}(1 - F_s(t_l)) - f_s(t_l)\frac{\partial V_0}{\partial t_l} =\,& f_s'(t_l)\left(V_0 - V_{k-1} - \int_{V_{k-1}}^{\infty} (1 - F_v(y|t_l))\, dy + c_e^k \right) \\
& + f_s(t_l)\left(\frac{\partial V_0}{\partial t_l} - \frac{\partial V_{k-1}}{\partial t_l} + (1 - F_v(V_{k-1}|t_l))\frac{\partial V_{k-1}}{\partial t_l} + \int_{V_{k-1}}^{\infty} dy\, \frac{\partial F_v(y|t_l)}{\partial t_l} \right) \\
& + \frac{\partial^2 V_{k-1}}{\partial t_l^2} \int_{t_l}^{t_u} f_s(s) F_v(V_{k-1}|s)\, ds + \frac{\partial V_{k-1}}{\partial t_l} \frac{\partial}{\partial t_l}\left(\int_{t_l}^{t_u} f_s(s) F_v(V_{k-1}|s)\, ds \right) \quad \text{(B.7)}
\end{aligned}
$$

---

[1] This holds whenever $\frac{\partial V_0}{\partial t_l} = 0$, see Section B.2

Substituting $\frac{\partial V_0}{\partial t_l} = 0$, $\frac{\partial V_{k-1}}{\partial t_l} = 0$, $c_e^k = \int_{V_{k-1}}^{\infty} (1 - F_v(y|t_l))\, dy + V_{k-1} - V_0$, and $\frac{\partial^2 V_{k-1}}{\partial t_l^2} = \prod_{i=1}^{k-1} \Pr(v \leq V_{i-1}|s \geq t_i) \frac{\partial^2 V_0}{\partial t_l^2} + \frac{\partial}{\partial t_l} \left(\prod_{i=1}^{k-1} \Pr(v \leq V_{i-1}|s \geq t_i) \right) \frac{\partial V_0}{\partial t_l}$ from Section B.2 in (B.7):

$$\frac{\partial^2 V_0}{\partial t_l^2} \left(1 - F_s(t_l) - \left(\prod_{i=1}^{k-1} \Pr(v \leq V_{i-1}|s \geq t_i) \right) \Pr(v \leq V_{k-1} \wedge t_l \leq s \leq t_u) \right)$$
$$= f_s(t_l) \left(\int_{V_{k-1}}^{\infty} dy \frac{\partial F_v(y|t_l)}{\partial t_l} \right)$$

The left hand side is positive, since:

$$1 - F_s(t_l) - \left(\prod_{i=1}^{k-1} \Pr(v \leq V_{i-1}|s \geq t_i) \right) \Pr(v \leq V_{k-1} \wedge t_l \leq s \leq t_u)$$
$$= 1 - F_s(t_l) - \left(\prod_{i=1}^{k-1} \Pr(v \leq V_{i-1}|s \geq t_i) \right) \Pr(v \leq V_{k-1}|t_l \leq s \leq t_u) \Pr(t_l \leq s \leq t_u)$$
$$\geq 1 - F_s(t_l) - \Pr(t_l \leq s \leq t_u) = 1 - F_s(t_l) - (F_s(t_u) - F_s(t_l)) = 1 - F_s(t_u) \geq 0$$

By the MLRP assumption, $F_V(y|t_l)$ is a decreasing function in t_l, therefore, the derivative is negative, and the term inside the integral is negative, meaning the above term is negative, which completes the proof, showing that Equation 6.5 must hold when using the optimal search strategy. The proof for Equation 6.6 is similar.

For Part (2), given Lemma 1, we only need to prove that if, optimally, the searcher should resume search given a signal s, then she must also do so given signal $s' < s$. The proof is similar to the above, ignoring the option to terminate without querying the expert and the query cost, and replacing V_{k-1} by $V_{\gamma-1}$ whenever applicable. If the optimal strategy given signal s is to resume search then the following equality (B.8) holds, describing the superiority of resuming search over querying the expert:

$$V_{\gamma>0} > M(s, V_{\gamma-1}) \tag{B.8}$$

Since $M(s, V_{\gamma-1}) > M(s', V_{\gamma-1})$ as shown in (B.5), we obtain that (B.8) holds also for s'. This, together with Lemma 1 suggests that if, optimally, the searcher should terminate her search given a signal s, then she must also do so given any signal $s'' > s$. Taking the derivative of Equation 6.3 (the Bellman Equation for state $\gamma > 0$) w.r.t. t_γ and equating to zero (similar to the way given above), we obtain a unique t_γ which maximizes the expected utility. We can use Equations 6.4, 6.5, 6.6, and the set of $2k - 2$ instances of Equations 6.3 and 6.7 (one for each γ value, $1 \leq \gamma < k$ i.e., a total of $2k + 1$ equations), to solve for $2k + 1$ unknowns: V_0, t_l, t_u, and for each $\gamma = 1, 2, \ldots, k - 1$ the appropriate V_γ and t_γ. $\square$

We note that all these proofs generalize easily to cases where the value distribution has support $[a, b]$. Furthermore, the above analysis is applicable also for the case where the expert supplies a noisy signal rather than a definite value, whenever the expert's signal complies with MLRP. This is achieved using a simple transformation that can be found in [66].

B.4   Proof of Theorem 7

Corollary 2 *For any probability distribution $f_v(v|s)$ satisfying MLRP then $f_v(v|s, s_e)$ also satisfies MLRP i.e., $\frac{f_v(v|s_1,s_e)}{f_v(v|s_2,s_e)}$ is non-decreasing in v if $s_1 > s_2$ and if s and s_e are conditionally independent given v.*

Proof:

$$
\begin{aligned}
\frac{f_v(v|s_1, s_e)}{f_v(v|s_2, s_e)} &= \frac{f(s_1, s_e|v) f_v(v) f(s_2, s_e)}{f(s_2, s_e|v) f_v(v) f(s_1, s_e)} \quad \text{Bayes' Rule} \\
&= \frac{f(s_1|v) f(s_e|v) f(s_2, s_e)}{f(s_2|v) f(s_e|v) f(s_1, s_e)} \quad \text{Conditional independence of } s_1 \text{ and } s_e \text{ given } v \\
&= \frac{f(v|s_1) f(s_1) f_v(v) f(s_2, s_e)}{f(v|s_2) f(s_2) f_v(v) f(s_1, s_e)} \quad \text{Bayes' Rule} \\
&= \frac{f_v(v|s_1)}{f_v(v|s_2)} \frac{f(s_e|s_2)}{f(s_e|s_1)}
\end{aligned}
$$

As $f_v(v|s)$ satisfies MLRP $\Rightarrow \frac{f_v(v|s_1)}{f_v(v|s_2)}$ is increasing in v if $s_1 > s_2$, therefore, $\frac{f_v(v|s_1,s_e)}{f_v(v|s_2,s_e)}$ is also increasing in v when $s_1 > s_2$. $\square$

Proof for Theorem 7

Proof: We know $Z = \mathbb{E}(v|s, s_e)$ where s_e is expert's signal. The distribution of Z can be calculated $F_Z(z|s) = \Pr(\mathbb{E}_v(v|s, s_e) < z|s)$. From Corollary 2, for $s_1 > s_2$, we can derive $\mathbb{E}(v|s_1, s_e) > \mathbb{E}(v|s_1, s_e)$ (MLRP implies second order stochastic dominance). This implies that $F_Z(z|s)$ is non-increasing in s, thus, the distribution of Z conditional on a higher searcher signal stochastically dominates the distribution conditional on a lower searcher signal. We will use this stochastic dominance property in the remainder of the proof.

The utility of the searcher from querying the more noisyexpert is given by Equation 8.1:

$$
\begin{aligned}
U_{\text{low}}(s) &= -c_{e_{\text{low}}} + V F_{z_{\text{low}}}(V|s) + \int_V^\infty z f_{z_{\text{low}}}(z|s)\, dz \\
&= -c_{e_{\text{low}}} + V + \int_V^\infty (1 - F_{Z\text{low}}(z|s))\, dz \quad \text{using integration-by-parts}
\end{aligned}
$$

Similarly, the utility for querying the less noisy expert is:

$$
U_{\text{high}}(s) = -c_{e_{\text{high}}} + V + \int_V^\infty (1 - F_{Z\text{high}}(z|s))\, dz
$$

We prove the theorem as follows:

1. We first show if it is optimal for the searcher to reject the opportunity and resume search when the signal is s, then it is also optimal to reject if the signal is $s' < s$:

   Since it is optimal to reject the opportunity when the searcher's signal is s, V is greater then the utility received from accepting without querying, therefore:

   $$V > \mathbb{E}_v(v|s)$$
   $$> \mathbb{E}_v(v|s') \quad \text{Due to MLRP}$$

   Also V is greater than the utility received by the searcher from querying:

   $$V > \max(U_{\text{low}}(s), U_{\text{high}}(s))$$
   $$> \max(U_{\text{low}}(s'), U_{\text{high}}(s'))$$

   Therefore, if it is optimal for the searcher to resume search on receiving signal s then it is optimal to resume the search at signal $s' < s$.

2. If it is optimal to accept the opportunity at any signal s, then it is also optimal to accept the opportunity for any signal $s' > s$:

   $$\mathbb{E}_v(v|s) > V$$
   $$\mathbb{E}_v(v|s') > \mathbb{E}_v(v|s) \quad \text{MLRP implies second-order stochastic dominance}$$
   $$\Rightarrow \mathbb{E}_v(v|s') > V$$

   As it is optimal for the searcher to accept the opportunity when the signal received is s — $\mathbb{E}_v(v|s)$ is greater than the utility received by the searcher from querying.

   $$\mathbb{E}_v(v|s) > \max(U_{\text{low}}(s), U_{\text{high}}(s))$$
   $$\mathbb{E}_v(v|s') > \mathbb{E}_v(v|s) \quad \text{MLRP implies second order stochastic dominance}$$
   $$\max(U_{\text{low}}(s'), U_{\text{high}}(s')) > \max(U_{\text{low}}(s), U_{\text{high}}(s)) \quad \text{First-order stochastic dominance}$$

   We need to show $\mathbb{E}_v(v|s') > \max(U_{\text{low}}(s'), U_{\text{high}}(s'))$

   Let $\Delta E = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s)$ and $\Delta Q = U_e(s') - U_e(s)$, where $U_e(s)$ and $U_e(s')$ represent the expected utility received if the searcher queries the best expert at signal s ($\max(U_{\text{low}}(s), U_{\text{high}}(s))$) and s' ($\max(U_{\text{low}}(s'), U_{\text{high}}(s'))$) respectively.

   If $\Delta E - \Delta Q > 0$, then it is better for the searcher to accept the opportunity at any signal $s' > s$, if it is optimal to accept at signal s because there is higher increase in utility from accepting without querying than querying.

   We know that $Z_{\text{high}} = \mathbb{E}(v|s, s_{e_{\text{high}}})$ and $Z_{\text{low}} = \mathbb{E}(v|s, s_{e_{\text{low}}}) \Rightarrow \mathbb{E}(v|s) = \int_{-\infty}^{\infty} z f_{Z_{\text{high}}}(z|s)\, dz = \int_{-\infty}^{\infty} z f_{Z_{\text{low}}}(z|s)\, dz$ (i.e, the mean of $f_{Z_{\text{low}}}(z|s)$ and $f_{Z_{\text{high}}}(z|s)$ is equal)

$$\Delta E - \Delta Q = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_e(s') + U_e(s)$$
$$= \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - \max(U_{\text{low}}(s'), U_{\text{high}}(s'))$$
$$+ \max(U_{\text{high}}(s), U_{\text{low}}(s)) \tag{B.9}$$

We need to consider the following four cases to show that Equation B.9 is greater than zero:

(a) Case 1:

$$\max(U_{\text{low}}(s'), U_{\text{high}}(s')) = U_{\text{low}}(s')$$
$$\max(U_{\text{low}}(s), U_{\text{high}}(s)) = U_{\text{low}}(s)$$

Equation B.9 becomes:

$$\Delta E - \Delta Q = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{low}}(s') + U_{\text{low}}(s)$$
$$= \int_\infty^\infty z f_{Z_{\text{low}}}(z|s')\, dz - \int_\infty^\infty z f_{Z_{\text{low}}}(z|s)\, dz - \int_V^\infty (1 - F_{Z_{\text{low}}}(z|s'))\, dz$$
$$+ \int_V^\infty (1 - F_{Z_{\text{low}}}(z|s))\, dz$$
$$= \int_{-\infty}^V (F_{Z_{\text{low}}}(z|s) - F_{Z_{\text{low}}}(z|s'))\, dz$$
$$> 0 \quad \text{First order stochastic dominance}$$

(b) Case 2:

$$\max(U_{\text{low}}(s'), U_{\text{high}}(s')) = U_{\text{low}}(s')$$
$$\max(U_{\text{low}}(s), U_{\text{high}}(s)) = U_{\text{high}}(s)$$

Equation B.9 becomes:

$$\Delta E - \Delta Q = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{low}}(s') + U_{\text{high}}(s)$$
$$> \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{low}}(s') + U_{\text{low}}(s) \quad \text{as } U_{\text{high}}(s) > U_{\text{low}}(s)$$
$$\int_\infty^\infty z f_{Z_{\text{low}}}(z|s')\, dz - \int_\infty^\infty z f_{Z_{\text{low}}}(z|s)\, dz - \int_V^\infty (1 - F_{Z_{\text{low}}}(z|s'))\, dz$$
$$+ \int_V^\infty (1 - F_{Z_{\text{low}}}(z|s))\, dz$$
$$\int_{-\infty}^V (F_{Z_{\text{low}}}(z|s) - F_{Z_{\text{low}}}(z|s'))\, dz$$
$$> 0 \quad \text{First order stochastic dominance}$$

(c) Case 3:

$$\max(U_{\text{low}}(s'), U_{\text{high}}(s')) = U_{\text{high}}(s')$$
$$\max(U_{\text{low}}(s), U_{\text{high}}(s)) = U_{\text{low}}(s)$$

Equation B.9 becomes:

$$\Delta E - \Delta Q = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{high}}(s') + U_{\text{low}}(s)$$
$$> \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{high}}(s') + U_{\text{high}}(s) \quad \text{as } U_{\text{low}}(s) > U_{\text{high}}(s)$$
$$\int_{\infty}^{\infty} z f_{Z_{\text{high}}}(z|s')\,dz - \int_{\infty}^{\infty} z f_{Z_{\text{high}}}(z|s)\,dz - \int_{V}^{\infty} (1 - F_{Z_{\text{high}}}(z|s'))\,dz$$
$$+ \int_{V}^{\infty} (1 - F_{Z_{\text{high}}}(z|s))\,dz$$
$$\int_{-\infty}^{V} (F_{Z_{\text{high}}}(z|s) - F_{Z_{\text{high}}}(z|s'))\,dz$$
$$> 0 \quad \text{First order stochastic dominance}$$

(d) Case 4:

$$\max(U_{\text{low}}(s'), U_{\text{high}}(s')) = U_{\text{high}}(s')$$
$$\max(U_{\text{low}}(s), U_{\text{high}}(s)) = U_{\text{high}}(s)$$

Equation B.9 becomes:

$$\Delta E - \Delta Q = \mathbb{E}_v(v|s') - \mathbb{E}_v(v|s) - U_{\text{high}}(s') + U_{\text{high}}(s)$$
$$= \int_{\infty}^{\infty} z f_{Z_{\text{high}}}(z|s')\,dz - \int_{\infty}^{\infty} z f_{Z_{\text{high}}}(z|s)\,dz - \int_{V}^{\infty} (1 - F_{Z_{\text{high}}}(z|s'))\,dz$$
$$+ \int_{V}^{\infty} (1 - F_{Z_{\text{high}}}(z|s))\,dz$$
$$= \int_{-\infty}^{V} (F_{Z_{\text{high}}}(z|s) - F_{Z_{\text{high}}}(z|s'))\,dz$$
$$> 0 \quad \text{First order stochastic dominance}$$

Therefore, it is optimal to accept an item at any signal $s' > s$, if it is optimal to accept it at s. The Bellman equation for the searcher is given by:

$$V = -c_s + V F_s(t_l) + \int_{t_u}^{\infty} \mathbb{E}(v|s) f_s(s)\,ds \int_{t_l}^{t_u} ds \max(U_{\text{high}}(s), U_{\text{low}}(s))\,ds$$

We derive the other two equations using the indifference method. We know that at lower threshold i.e., $s = t_l$, the searcher is indifferent between rejecting the item and querying the expert.

$$V = \max(U_{\text{high}}(t_l), U_{\text{low}}(t_l))$$

Similarly, at upper threshold i.e. $s = t_u$, the searcher is indifferent in querying the item vs. accepting it without querying

$$\mathbb{E}(v|t_u) = \max(U_{\text{high}}(t_u), U_{\text{low}}(t_u))$$

This proves our theorem. $\square$