Dynamic Pricing and Learning

A.V. den Boer

Dit proefschrift is goedgekeurd door de promotoren:

prof. dr. R.D. van der Mei
prof. dr. A.P. Zwart

Contents

1 Introduction and contributions **1**
- 1.1 Introduction . . . 1
  - 1.1.1 Background . . . 1
  - 1.1.2 Dynamic pricing and learning . . . 2
  - 1.1.3 Relation to revenue management . . . 2
- 1.2 Contributions of this thesis . . . 3

2 Literature **7**
- 2.1 Historical origins of pricing and demand estimation . . . 7
  - 2.1.1 Demand functions in pricing problems . . . 7
  - 2.1.2 Demand estimation . . . 7
  - 2.1.3 Practical applicability . . . 8
- 2.2 Dynamic pricing . . . 9
  - 2.2.1 Dynamic pricing with dynamic demand . . . 9
  - 2.2.2 Dynamic pricing with inventory effects . . . 10
- 2.3 Dynamic pricing and learning . . . 12
  - 2.3.1 No inventory restrictions . . . 12
  - 2.3.2 Finite inventory . . . 16
  - 2.3.3 Joint pricing and inventory problems . . . 18
- 2.4 Methodologically related areas . . . 19

3 Dynamic pricing and learning for a single product with infinite inventory **21**
- 3.1 Introduction . . . 21
- 3.2 Model, assumptions, and estimation method . . . 23
  - 3.2.1 Model . . . 23
  - 3.2.2 Discussion of model assumptions . . . 24
  - 3.2.3 Estimation of unknown parameters . . . 24
- 3.3 Performance of pricing policies . . . 25
  - 3.3.1 Inconsistency of certainty equivalent pricing . . . 25
  - 3.3.2 Controlled Variance Pricing . . . 26
- 3.4 Discussion . . . 28
  - 3.4.1 Quality of regret bound . . . 28
  - 3.4.2 Generality of result . . . 29
  - 3.4.3 Differences with the multiperiod control problem . . . 29
  - 3.4.4 Applicability to other sequential decision problems . . . 30
- 3.5 Numerical evaluation . . . 30
- 3.6 Proofs . . . 33

4 Dynamic pricing and learning for multiple products with infinite inventory **39**
- 4.1 Introduction . . . 39

4.2 Model, assumptions, and estimation method . . . 40
4.2.1 Model and notation . . . 40
4.2.2 Discussion of model assumptions . . . 42
4.2.3 Estimation of unknown parameters . . . 42
4.3 Adaptive pricing policy . . . 43
4.4 Bounds on the regret . . . 45
4.4.1 General link functions . . . 45
4.4.2 Canonical link functions . . . 47
4.4.3 Auxiliary results . . . 48
4.4.4 Quality of regret bounds . . . 48
4.5 Numerical illustration . . . 49
4.5.1 Two products, Poisson distributed demand . . . 49
4.5.2 Ten products, normally distributed demand . . . 50
4.6 Proofs . . . 52

5 Dynamic pricing and learning for a single product with finite inventory 65
5.1 Introduction . . . 65
5.2 Preliminaries . . . 66
5.2.1 Model formulation . . . 66
5.2.2 Demand distribution . . . 67
5.2.3 Full-information optimal solution . . . 67
5.2.4 Regret measure . . . 68
5.3 Parameter estimation . . . 69
5.3.1 Maximum-likelihood estimation . . . 69
5.3.2 Convergence rates of parameter estimates . . . 69
5.4 Endogenous learning . . . 70
5.5 Pricing strategy . . . 71
5.6 Numerical illustration . . . 72
5.7 Proofs . . . 73
5.8 Proofs of auxiliary lemmas . . . 81

6 Dynamic pricing and learning in a changing environment 85
6.1 Introduction . . . 85
6.2 Model . . . 87
6.3 Estimation of market process . . . 89
6.4 Pricing policy and performance bounds . . . 91
6.5 Hedging against changes . . . 93
6.6 Numerical illustration . . . 97
6.7 Proofs . . . 101

7 Mean square convergence rates for maximum quasi-likelihood estimators 115
7.1 Introduction . . . 115
7.1.1 Motivation . . . 115
7.1.2 Literature . . . 116
7.1.3 Assumptions and contributions . . . 117
7.1.4 Applications . . . 118
7.1.5 Organization of the chapter . . . 118
7.2 Results for general link functions . . . 119
7.3 Results for canonical link functions . . . 120
7.4 Proofs . . . 122
7.5 Appendix: auxiliary results . . . 127

Future directions **135**

Nederlandse samenvatting **137**

Bibliography **141**

Chapter 1

Introduction and contributions

1.1 Introduction

1.1.1 Background

The emergence of the Internet as a sales channel has made it easy for companies to experiment with selling prices. Web shops can adapt their prices with a proverbial flick of the switch, without any additional costs or efforts; the same applies for retail stores that make use of digital price tags (Kalyanam et al., 2006). In contrast, in the past costs and effort were needed to change prices, for example by issuing a new catalog or replacing price tags. This current flexibility in pricing is one of the main drivers for research on *dynamic pricing*: the study of determining optimal selling prices under changing circumstances.

These changing circumstances can, for example, be (perceived or expected) changes in the market, price changes or marketing campaigns by competing firms, technological innovations in the products, shifts in consumer tastes, or market saturation effects related to the life-cycle of a product. Even the weather forecast may influence selling prices, for example in electricity markets, the tourism sector, or in apparel retail stores.

One particularly important example of external factors that drive dynamic pricing is changing inventory levels. This is important in situations where a firm sells a finite number of products during a finite time period, as is the case for airline tickets, hotel room reservations, concert tickets, and perishable products in general. The optimal selling price usually depends on the inventory level and on the remaining length of the selling season, and if one of these changes, a price adjustment is beneficial.

Ideas and techniques from dynamic pricing are nowadays widely applied in various business contexts: airline companies, hotels, restaurants, concert halls and theaters, amusement parks, car rental, vendors of train tickets, e-books, and many retail companies. Several textbooks are already available that discuss various aspects of dynamic pricing that are important in practical applications, cf. Talluri and van Ryzin (2004), Phillips (2005), Özer and Phillips (2012).

Only little is understood, however, about how the large amount of sales data that typically is available should efficiently be used in pricing problems. This data can be used to derive estimates on vital information, such as market trends or price-sensitivity of consumers. Traditional approaches to dynamic pricing have mostly neglected this aspect. In this thesis, we study data-driven pricing policies that incorporate real-time incoming sales data to determine optimal pricing decisions.

1.1.2 Dynamic pricing and learning

An intrinsic property of many price optimization problems is *lack of information*: the seller does not know how consumers respond to different selling prices, and thus does not know the optimal price. Dynamic pricing problems are therefore not merely about optimization, but also about learning the relation between price and market response. Typically, this relation is modeled by a demand model or demand function that depends on a number of unknown parameters. The value of these parameters can be learned by applying appropriate statistical estimation techniques on historical sales data.

The presence of digitally available and frequently updated sales data makes this problem essentially an on-line learning problem: after each sales occurrence, the firm can use the newly obtained sales data to update its knowledge on consumer behavior. If, in addition, selling prices can easily be adapted without much costs or effort - as is the case with web-based sales channels or in brick-and-mortar stores with digital price tags - the firm may immediately exploit this improved knowledge of consumer behavior by appropriately adapting the selling prices.

The decision maker thus faces the task of both estimating the price-demand relation and optimizing selling prices. An important insight from studies on dynamic pricing and learning is that these two tasks should not be separated, but rather conducted simultaneously. The reason is that chosen selling prices generally influence the quality of the parameter estimates, which in turn affect the quality of future pricing decisions. A selling price may for example optimize the expected profit with respect to current parameter estimates, but hardly increase the quality of future parameter estimates. In such cases, it may be beneficial to deviate from this price, and sacrifice some of the short-term earnings in order to improve future profits.

Estimation and pricing are thus two closely related problems, which in general should be considered simultaneously instead of separately. Neglecting the influence of chosen prices on the estimation process can lead to significant revenue losses, as illustrated in Chapter 3. In some other business contexts, however, such a pricing policy may perform very well, as for example shown in Chapter 5. The main goal of this thesis is to investigate which pricing policies optimally balance the two objectives of learning and optimization, for several practically relevant settings.

Dynamic pricing and learning problems fall in the class of sequential decision problems under uncertainty. In these problems, a decision maker has to repeatedly choose an action in order to optimize some objective, without exactly knowing which action is optimal. As time progresses and more actions have been taken, the decision maker may learn the quality of different actions, based on previously observed outcomes. A key question for these type of problems is: should the decision maker always take the action that, according to current knowledge, is best? Or should he sometimes deliberately deviate from the (perceived) optimal action, in order to increase his knowledge about the system and consequently improve his future actions? In this thesis, we address these questions in the context of dynamic pricing problems.

1.1.3 Relation to revenue management

Dynamic pricing is closely related to revenue management, and the terms are sometimes used interchangeably to denote the same business practice. Revenue management mostly refers however to settings where not prices, but capacities are dynamically adjusted, and the term is often connected to models used in the airline industry; cf. page 6 of Talluri and van Ryzin (2004), where the authors write: “Where did R[evenue] M[anagement] come from? In short, the airline industry. There are few business practices whose origins are so intimately connected to a single industry”.

Common practice in airline industry is to sell different "fare classes" for flight tickets. A fare class corresponds to a selling price (a "fare") for a seat on a particular flight, together with a number of conditions, such as the costs for cancellation, or the amount of luggage that one is allowed to carry. A fare class is distinct from the usual division of seats between business and economy class: generally, multiple fare classes correspond to the same economy or business class seat. The fares and fare classes themselves are not changed during the selling period. Instead, the firm dynamically adjusts which fare classes are offered for sale, in order to maximize profit or revenue.

A disadvantage of such decision models with capacity-based controls is the restriction to a finite number of fare classes and corresponding selling prices. The customer reservation systems of many airline companies can not handle more than twenty-six fare classes, mainly because the software uses a coding of fare classes that is based on the letters in the alphabet. In contrast, dynamic pricing models are much more flexible and generally do not restrict the number of prices that can be deployed by the firm.

A further difficulty of capacity-based revenue management is that modeling and estimating the mutual dependence of demand between different fare classes is not an easy task. If the expected demand for each fare class is estimated independently, accumulating sales data for a particular fare class does not provide any information about the expected demand for other fare classes. As a result, much data is needed before accurate decisions can be made. In dynamic pricing models one usually considers a parametric demand model that describes the relation between price and expected demand, and that depends on unknown parameters. Then accumulating sales data improves estimates for the demand at any price, because it improves the estimates for the unknown parameters, and, compared to capacity-based models, considerably less data is needed before accurate pricing decisions can be made.

For these reasons, dynamic pricing is a more flexible approach than capacity-based revenue management methods. These latter methods are tailored to the characteristics of the airline industry, whereas dynamic pricing is suitable for application in various fields and branches.

1.2 Contributions of this thesis

We here provide a summary of the contents of the remainder of this thesis. Chapters 3 to 6 contain our contributions to dynamic pricing problems. Chapter 7 studies the convergence speed of certain statistical estimators; the results from this chapter are applied in Chapters 3, 4, and 5 to characterize the performance of pricing strategies. The thesis closes with a discussion on interesting directions for future research, and with a Dutch summary of our findings.

Chapter 3: pricing a single product with infinite inventory

In Chapter 3, we consider the dynamic pricing problem of a monopolist firm selling a single product with infinite inventory (in practice, this means that stock-outs occur only with a very small probability, or that orders can be backlogged). We assume a parametric form of the relation between selling price and expected demand, but we do not require that the seller has complete knowledge on the demand distribution; only knowledge on the relation between the first two moments of demand and the selling price is assumed. This results in a generic model that includes practically all demand functions used in practice. In addition, the model is more robust to misspecification than models in which a complete demand distribution is assumed. To estimate the unknown parameters of the model, we deploy maximum quasi-likelihood estimation;

this is a natural extension of maximum-likelihood estimation to settings where only the first two moments of the distribution are known.

An intuitive pricing policy is certainty equivalent pricing (CEP), which at each decision moment chooses the price that is optimal with respect to the available parameter estimates. We show that if the seller uses this policy, then with positive probability the true values of the unknown parameters are never learned, and the selling price does not converge to the optimal price. The intuitive reason behind this inconsistency is that CEP puts too much emphasis on instant revenue maximization, and too little emphasis on information collection through price experimentation.

We propose a new dynamic pricing policy, called Controlled Variance Pricing (CVP). The key idea is to always choose the price closest to the certainty equivalent price, but such that a certain amount of price dispersion is guaranteed. The price dispersion, measured by the sample variance of the selling prices, influences the quality of the parameter estimates, and by carefully tuning the growth rate of the price dispersion, we achieve an optimal balance between learning and instant optimization. We characterize the performance of CVP, and show that its performance is arbitrarily close to the best achievable performance of any pricing policy. Two other advantages of CVP compared to alternative pricing policies in the literature are (1) guaranteed convergence of the selling price to the optimal price, and (2) the optimal use of all available sales data to form estimates.

Chapter 3 is based on den Boer and Zwart (2010).

Chapter 4: pricing multiple products with infinite inventory

In Chapter 4, we consider the problem of dynamic pricing and learning with infinite inventory, in a setting with multiple products. This allows the modeling of substitute and complementary products. We deploy a similar type of demand model as in Chapter 3: the seller knows the relation between the selling prices and the first two moments of the demand distribution (up to some unknown parameters), but does not need to know the complete distribution. Unknown parameters are estimated with maximum quasi-likelihood estimation.

Just as in the single product setting, a CEP policy is not suitable. Some amount of price experimentation is necessary to ensure that eventually the unknown parameters are learned. In Chapter 3, the sample variance of chosen selling prices is regarded as a measure of price dispersion, and learning of the parameters is ensured by requiring a minimum growth rate on this sample variance. In a multiple product setting, we measure the price dispersion by the smallest eigenvalue of the so-called design matrix. Our proposed pricing policy is to always choose the price that is optimal with respect to the current parameter estimates, but with the requirement that this smallest eigenvalue satisfies a certain lower bound. By carefully tuning this lower bound, we obtain the optimal balance between learning and instant optimization. Our policy is one of the first dynamic pricing and learning policies that can handle a multiple-products setting.

Chapter 4 is based on den Boer (2011).

Chapter 5: pricing a single product with finite inventory

Chapter 5 considers a firm selling a finite number of products during a finite time period. This is a classical model in the literature on dynamic pricing and revenue management, and it has been studied extensively. Existing studies that consider dynamic pricing and learning for this

model are somewhat limited in their scope, either because they assume a rather simplistic demand model, or because they only analyze the performance of pricing policies when initial inventory and the length of the selling period grow very large.

We study dynamic pricing and learning in this setting, for any initial inventory level or length of the selling period. Our main result is that CEP has a good performance, and additional price experimentation is not necessary. This differs from the infinite inventory setting considered in Chapter 3 and 4, where CEP has a very poor performance. The reason for this difference is that in the finite inventory setting a certain "endogenous learning" property is satisfied, that leads to abundant price dispersion even when no active price experimentation is applied. As a result, parameter estimates converge quickly to their true value, and the chosen prices converge quickly to the optimal prices.

Chapter 5 is based on den Boer and Zwart (2011b).

Chapter 6: pricing in a changing market environment

The literature on dynamic pricing and learning generally assumes that the relation between price and expected demand is stable, and does not change over time. In practice, this strong condition is often violated: the behavior of markets may change regularly for different reasons, and pricing policies should be able to respond to these fluctuations. In Chapter 6, we study dynamic pricing and learning of a single product with infinite inventory, in a setting where the market process is subject to variations. Using weighted least-squares type estimators, we derive bounds on the performance of a variant of CEP, and provide guidelines how to optimally choose the estimator. This enables the firm to hedge against the risk of fluctuations in the market. Numerical examples illustrate the methodology in several practical situations, such as pricing in view of market saturation effects, or pricing in a competitive environment. This study is the first on dynamic pricing and learning in a changing market that provides a pricing policy with an explicit characterization of its performance.

Chapter 6 is based on den Boer (2012a).

Chapter 7: convergence rates for maximum quasi-likelihood estimation

The unknown parameters of the demand models in Chapter 3, 4, and 5 are estimated using maximum quasi-likelihood estimation. This is an extension of maximum-likelihood estimation to models where only knowledge on the first two moments of the distribution is available. In Chapter 7, we derive bounds on the convergence rates of the expected squared estimation error for this type of estimator. These convergence rates are used to characterize the performance of pricing policies in Chapter 3, 4, and 5. In addition, they are applicable in many other sequential decision problems under uncertainty.

Chapter 7 is based on den Boer and Zwart (2011a).

Chapter 2

Literature

This chapter surveys the rich literature on dynamic pricing with learning. Section 2.1 briefly reviews some of the pioneering historical work done on pricing and demand estimation, and discusses how this work initially was difficult to apply in practice. In Section 2.2 we sketch important references and developments for dynamic pricing in general, and in Section 2.3 we focus on the literature that deals with dynamic pricing and learning. Connections between dynamic pricing and related research areas are addressed in Section 2.4.

2.1 Historical origins of pricing and demand estimation

Dynamic pricing with learning can be seen as the combined application of two research fields: (1) statistical learning, specifically applied to the problem of estimating demand functions, and (2) price optimization. Both these fields are already quite old, dating back more than a century. In this section we briefly describe the historical fundaments out of which dynamic pricing and learning has emerged, by pointing to some key references on static pricing and estimating demand functions that have been important in the progress of the field.

2.1.1 Demand functions in pricing problems

Cournot (1838) is generally acknowledged as the first to use a mathematical function to describe the price-demand relation of products, and subsequently solve the mathematical problem of determining the optimal selling price. As vividly described by Fisher (1898), such an application of mathematical methods to study an economical problem was quite new and controversial at the time, and the work was neglected for several decades. Cournot showed that if $F(p)$ denotes the demand as function of price p, where $F(p)$ is continuous, decreasing in p, and $pF(p)$ converges to zero as p grows large, then the price that maximizes the revenue $pF(p)$ can be found by equating the derivative of $pF(p)$ to zero. If $F(p)$ is concave, there is a unique solution, which is the optimal price (this is contained in Chapter IV of Cournot (1838)). In this way, Cournot was the first to solve a "static pricing" problem by mathematical methods.

2.1.2 Demand estimation

To make theoretical knowledge on optimal pricing theory applicable in practical problems, one needs to have an estimate of the demand function. The first known empirical work on demand

curves is the so-called King-Davenant Law (Davenant, 1699) which relates the supply and price of corn (see Creedy, 1986, for an exposition on the origins of this work). More advanced research on estimating demand curves, by means of statistical techniques such as correlation and linear regression, took off in the beginning of the 20st century. Benini (1907), Gini (1910) and Lehfeldt (1914) estimate demand curves for various goods as coffee, tea, salt, and wheat, using various curve-fitting methods. Further progress on methodology was made, among others, by Moore (1914, 1917), Wright (1928) and Tinbergen (1930); the monumental work of Schultz (1938) gives a thorough overview of the state-of-the-art of demand estimation in his time, accompanied by many examples. Further references and more information on the historical progress of demand estimation can be found in Stigler (1954, section II), Stigler (1962, particularly section iii), Brown and Deaton (1972), Christ (1985), and Farebrother (2006).

2.1.3 Practical applicability

Estimating demand curves of various products was in first instance not aimed at profit optimization of commercial firms, but rather used to support macro-economic theories on price, supply, and demand. Application of the developed methods in practical problems was initially far away. An illustrative quote is from Hicks (1935), who doubted the possibilities of applying the theory of monopoly pricing on practical problems, exactly because of the difficulty of estimating the demand curve:

> It is evidently the opinion of some of the writers under discussion that the modern theory of monopoly is not only capable of throwing considerable light on the general principles underlying an individualistic economic structure, but that it is also capable of extensive use in the analysis of particular practical economic problems, that is to say, in applied economics. Personally, I cannot but feel sceptical about this. [...] There does not seem to be any reason why a monopolist should not make a mistake in estimating the slope of the demand curve confronting him, and should maintain a certain output, thinking it was the position which maximized his profit, although he could actually have increased his profit by expanding or contracting. (Hicks, 1935, page 18,19).

Hawkins (1957) reviews some of the attempts made by commercial firms to estimate the demand for their products. Most of these attempts were not successful, and suffered from difficulties of obtaining sufficiently many data for reliable estimates, and of changes in the quality of the product and the prices of competitors. Even a very detailed study of General Motors on automobile demand ends, somewhat ironically, with:

> The most important conclusion from these analyses of the elasticity of demand for automobiles with respect to price is that no exact answer to the question has been obtained. (Horner et al., 1939, p. 137).

In view of these quotations, it is rather remarkable that dynamic pricing and learning has nowadays found its way in practice; many applications have been reported in various branches such as airline companies, the hospitality sector, car rental, retail stores, internet advertisement, and many more. A main cause for this is the fact that historical sales data nowadays is often digitally available; this significantly reduces the efforts needed to estimate the demand function. In addition, whenever products are sold via the Internet or using digital price tags, the costs associated with adjusting the prices in response to updated information or changed circumstances are practically zero. In contrast, a price-change in the pre-digital era would often induce costs, for example because a new catalog had to be printed or price tags had to be replaced.

2.2 Dynamic pricing

In this section we discuss the literature on dynamic pricing. There is a huge amount literature on this subject, and we do not intend to give a complete overview of the field. Instead, we briefly describe some of the major research streams and key references, in order to provide a context in which one can position the literature on dynamic pricing with learning discussed in Section 2.3.

The literature on dynamic pricing by a monopolist firm can roughly be classified as follows:

- Models where the demand function is dynamically changing over time.
- Models where the demand function is static, but where pricing dynamics are caused by the inventory level.

In the first class of models, reviewed in Section 2.2.1, the demand function changes according to changing circumstances: for example, the demand function may depend on the time-derivative of price, on the current inventory level, on the amount of cumulative sales, on the firm's pricing history, et cetera. In the second class of models, reviewed in Section 2.2.2, it is not the demand function itself that causes the pricing dynamics: a product offered in two different time periods against the same selling price, is expected to generate the same amount of average demand. Instead, the price dynamics are caused by inventory effects; for example, if the product on sale is almost sold out and no re-ordering is possible on the short term, it may be beneficial to increase the price. Naturally, it is also possible to study models that fall both in classes, if both the demand function is dynamically changing and the price dynamics are influenced by inventory effects; some of these literature is also reviewed in Section 2.2.2.

2.2.1 Dynamic pricing with dynamic demand

Demand depends on price-derivatives

Evans (1924) is one of the first to depart from the static pricing setting introduced by Cournot (1838). In a study on optimal monopoly pricing, he assumes that the (deterministic) demand is not only a function of price, but also of the time-derivative of price. This models the fact that buyers do not only consider the current selling price in their decision to buy a product, but also the future price path. The purpose of the firm is to calculate a price function, on a continuous time interval, that maximizes the profit. Using techniques from calculus of variations, the optimal price function is calculated. Various extensions to this model are made by Evans (1925), Roos (1925, 1927a,b, 1934), Tintner (1937), and Smithies (1939). Thompson et al. (1971) study an extended version of the model of Evans (1924), where optimal production level, investment level, and output price have to be determined. Closely connected to this work is Simaan and Takayama (1976), who consider a model where supply is the control variable; the time-derivative of price at each moment is a known function of the current supply and current price. Methods from control theory are used to derive properties of the optimal supply path.

Demand depends on price history

A related set of literature considers the effect of reference prices on the demand function. Reference prices are perceptions of customers about the price that the firm has been charging in the past; see Mazumdar et al. (2005) for a review on the subject. A difference between the reference price and the actual selling price influences the demand, and as a result, each posted selling price

does not only affect the current demand but also the future demand. Dynamic pricing models and properties of optimal pricing strategies in such a setting are studied by Greenleaf (1995), Kopalle et al. (1996), Fibich et al. (2003), Heidhues and Köszegi (2005), Ahn et al. (2007), Popescu and Wu (2007).

Demand depends on amount of sales

Another stream of literature on dynamic pricing emerged from diffusion and adoption models for new products. A key reference is Bass (1969), and reviews of diffusion models are given by Mahajan et al. (1990), Baptista (1999), and Meade and Islam (2006). In these models, the demand for products does not only depend on the selling price, but also on the amount of cumulative sales. This allows modeling several phenomena related to market saturation, advertisement, word-of-mouth effects, and product diffusion. Robinson and Lakhani (1975) study dynamic pricing in such a model, and numerically compare the performance of several pricing policies. Their work stimulated much further research on optimal dynamic pricing policies, see e.g. Clarke et al. (1982), Kalish (1983), Clarke and Dolan (1984), and the references therein. The models studied in these papers are deterministic, and somewhat related to the literature following Evans (1924): both types of pricing problems are solved by principles from optimal control theory, and the optimal pricing strategy is often characterized by a differential equation.

Chen and Jain (1992), Raman and Chatterjee (1995), and Kamrad et al. (2005) extend these models by incorporating randomness in the demand. In Chen and Jain (1992), the demand is determined by a finite-state Markov chain for which each state corresponds to a deterministic demand function that depends on price and cumulative sales. The optimal price path is characterized in terms of a stochastic differential equation, and compared to the optimal policy in a fully deterministic setting. Raman and Chatterjee (1995) model uncertainty by adding a Wiener process to the (known) deterministic component of the demand function. They characterize the pricing policy that maximizes discounted cumulative profit, and compare it with the optimal price path in the fully deterministic case. Under some specific assumptions, closed form solutions are derived. Similar models that incorporate demand uncertainty are analyzed by Kamrad et al. (2005). For various settings they provide closed-form solutions of the optimal pricing policies.

2.2.2 Dynamic pricing with inventory effects

There are two important research streams on dynamic pricing models where the dynamics of the optimal pricing policy are caused by the inventory level: (i) "revenue management" type of problems, where a finite amount of perishable inventory is sold during a finite time period, and (ii) joint pricing and inventory procurement problems.

Selling a fixed, finite inventory during a finite time period

In this stream of literature, a firm is assumed to have a certain number of products at its disposal, which are sold during a finite time period. There is no replenishment; inventory that is unsold at the end of the selling horizon is lost, and can not be transferred to another selling season. In these problems, the dynamic nature of optimal prices is not caused by changes in the demand, but rather by fact that the marginal value of remaining inventory is changing over time. As a result, the optimal selling price in these settings is not a fixed quantity, but depends on the remaining amount of inventory and the remaining duration of the selling season.

Kincaid and Darling (1963) may be the first to characterize and analyze the optimal pricing policy in such a setting. A more recent key reference is Gallego and van Ryzin (1994). They consider a continuous-time setting where demand is modeled as a Poisson process, with arrival rate that depends on the posted selling price. The pricing problem is formulated as a stochastic optimal control problem, and the optimal solution is characterized using the Hamilton-Jacobi-Bellman equation. A closed-form solution may in general not exist, but for a specific demand function a closed-form optimal solution is derived. The authors furthermore propose two heuristic pricing policies, provide bounds on their performance, and discuss various extensions to the model.

Numerous extensions and variations of the model by Gallego and van Ryzin (1994) have been studied: settings with restrictions on the number of allowable prices or price changes (Feng and Gallego, 1995, Bitran and Mondschein, 1997, Feng and Xiao, 2000a,b), extensions to multiple products that share the same finite set of resources (Gallego and van Ryzin, 1997, Kleywegt, 2001) or multiple stores (Bitran et al., 1998). For a thorough overview of this literature, we refer to the books Talluri and van Ryzin (2004), Phillips (2005), Özer and Phillips (2012), and the reviews by Bitran and Caldentey (2003), Elmaghraby and Keskinocak (2003).

Feng and Gallego (2000) and Zhao and Zheng (2000) characterize the optimal pricing policy in case of time dependent demand intensities; in these models, the price dynamics are caused both by inventory effects and by dynamic demand behavior.

The same holds for models that study strategically behaving customers: customers who, when arriving at the (online) store, do not immediately decide whether to buy the product, but instead wait for a while to anticipate possible decreases in the selling price. In contrast, so-called myopic customers instantly decide whether to buy the product at the moment they arrive at the store. In such settings, the demand at a certain moment depends on the past, present, and future selling prices. Dynamic pricing in view of strategic customers has received a considerable amount of research attention in recent years; a representative sample is Aviv and Pazgal (2008), Elmaghraby et al. (2008), Liu and van Ryzin (2008), Levin et al. (2009), Cachon and Swinney (2009) and Su (2010). These studies sometimes have a game-theoretic flavor, since both the firm and the strategic customers have a decision problem to solve, with contradicting interests.

Jointly determining selling prices and inventory procurement

A main assumption of the literature discussed above is that the initial capacity level is fixed. In many situations in practice this is a natural condition: the number of seats in an aircraft, rooms in a hotel, tables in a restaurant, or seats in a concert hall are all fixed for a considerable time period, and modifications in the capacity occur at a completely different time scale than dynamic price changes. In many other settings, however, the initial capacity is a decision variable to the firm; in particular, when the firm can decide how many items of inventory should be produced or procured. Pricing and inventory management can then be considered as a simultaneous optimization problem.

This research field bridges the gap between the pricing and inventory management literature. Many different settings and models are subject to study, with different types of production, holding and ordering costs, different replenishment policies (periodic or continuous), finite or infinite production capacity, different models for the demand function. Extensive reviews of the literature on simultaneous optimization of price and inventory decisions can be found in Eliashberg and Steinberg (1993), Elmaghraby and Keskinocak (2003, Section 4.1), Yano and Gilbert (2005), Chan et al. (2004), and Chen and Simchi-Levi (2012).

2.3 Dynamic pricing and learning

In the static monopoly pricing problem considered by Cournot (1838), the demand function is deterministic and completely known to the firm. These assumptions are somewhat unrealistic in practice, and eventually it was realized by researchers that demand uncertainty should be incorporated into the problem. One of the first to pursue this direction is Mills (1959), who assumes that the demand is the sum of a random term with zero mean and a deterministic function of price. He studies how a monopolist firm that sells finitely many products in a single time period should optimally set its production level and selling price. Further extensions of this model and properties of pricing problems with random demand are studied by Karlin and Carr (1962), Nevins (1966), Zabel (1970), Baron (1970, 1971), Sandmo (1971) and Leland (1972). An important research question in these studies is how the firm's optimal decisions are influenced by the demand uncertainty, and by the firms' attitude towards risk (risk-neutral, risk-averse, or risk-preferred).

In the models mentioned above, the expected demand as a function of the selling price is still assumed to be completely known by the firm, which makes these models somewhat unrealistic and not usable in practice. The common goal of the literature on dynamic pricing and learning is to develop pricing policies that take the intrinsic uncertainty on the price-demand relation into account.

In the next two sections we discuss the literature on dynamic pricing and learning. Section 2.3.1 considers the literature on the problem of a price-setting firm with infinite inventory and unknown demand function. This basically is the monopoly pricing problem described in Section 2.1.1, with uncertainty on the demand function. The full-information case of this problem is static; the price dynamics are completely caused by the fact that the firm learns about the price-demand relation through accumulating sales data. Section 2.3.2 discusses literature on pricing policies for firms selling a fixed, finite amount of inventory, with unknown demand function. For this problem, the full-information case is already dynamic by itself, as discussed in Section 2.2.2, and the learning aspect of the problem provides an additional source of the price dynamics.

2.3.1 No inventory restrictions

Early work

The first analytical work on dynamic monopoly pricing with unknown demand curve seems to have been done by members of the Uppsala Econometric Seminar, in 1953-54. Billström et al. (1954) contains a mimeographed report of work presented at the 16th Meeting of the Econometric Society in Uppsala, August 1954. The original report has not been published, but an English reprint has appeared in Billström and Thore (1964) and Thore (1964). These two works consider the problem of a monopolist facing a linear demand curve that depends on two unknown parameters. Thore (1964) proposes to use a dynamic pricing rules that satisfies $\text{sign}(p_t - p_{t-1}) = \text{sign}((p_{t-1} - p_{t-2})(r_{t-1} - r_{t-2}))$, where p_t, r_t denote the price and revenue in period t. Put in words, the idea is as follows: if a previous price increase led to an increase in revenue, the price will again be increased; otherwise it will be decreased. Similarly, if a previous price decrease led to an increase in revenue, the price will again be decreased; otherwise, it will be increased. In addition, Thore (1964) proposes to let the magnitude of the price adjustment, $p_t - p_{t-1}$, depend on the magnitude of $r_{t-1} - r_{t-2}$. He specifies two pricing rules in detail,

$$p_t - p_{t-1} = \text{constant} \cdot \sqrt{|r_{t-1} - r_{t-2}|} \cdot \text{sign}((p_{t-1} - p_{t-2})(r_{t-1} - r_{t-2})), \tag{2.1}$$

and

$$p_t - p_{t-1} = \text{constant} \cdot \frac{r_{t-1} - r_{t-2}}{p_{t-1} - p_{t-2}}, \tag{2.2}$$

and analyzes convergence properties of the resulting dynamical systems. Billström and Thore (1964) perform simulation experiments for pricing rule (2.1), both in a deterministic demand setting and in a setting where a normally distributed disturbance term is added to the demand. They also extend the model to incorporate inventory replenishment, and provide a rule of thumb for the optimal choice of the constant in (2.1).

These studies emerging from the Uppsala Econometrics Seminar have not received much research attention in subsequent years. Clower (1959) studies a monopolist firm facing a linear, deterministic demand function whose parameters may change over time. He discusses several price-adjustment mechanisms that may be applied by the firm to adapt its prices to changing situations. Baumol and Quandt (1964) propose rules of thumb for the monopolist pricing problem, and asses their performance by a set of numerical experiments. In their Appendix A they propose exactly pricing rule (2.2), although they are apparently unaware of the work of Thore (1964). They investigate some convergence and stability properties of the resulting dynamical system, both in a discrete-time and continuous-time framework. Baetge et al. (1977) extend the simulation results of Billström and Thore (1964) to non-linear demand curves, and further study the optimal choice of the constant in (2.1). A final study in this line of research is from Witt (1986). He studies a model where a monopolist has to decide on price, output level in the current period and capacity (maximum output) in the next period. Expected demand is linear with unknown coefficients, and may change over time. Three decision rules are compared to each other via a computer simulation. In addition, their performance is compared with a laboratory experiment, where test subjects had to determine their optimal pricing strategy.

Bayesian approaches

Several authors study the dynamic pricing and learning problem within a Bayesian framework. On of the first is Aoki (1973), who applies methods from stochastic adaptive control theory. He considers a setting where the demand function depends on unknown parameters, which are learned by the decision maker in a Bayesian fashion. The purpose is to minimize (a function of) the excess demand. He shows how the optimal Bayesian policy can, in theory, be computed via dynamic programming, but that in many situations no closed-form analytical expression of the solution exists. He proposes two approximation policies. In the first, certainty equivalent pricing (CEP), at each decision moment the price is chosen that would be optimal if the current parameter estimates were correct. In the second, called an approximation under static price expectation, the firm acts at each decision moments as if the chosen price will be maintained throughout the remainder of the selling period. Aoki (1974) he shows that the prices generated by these policies converge a.s. to the optimal price.

Similar work is by Chong and Cheng (1975), under more restrictive assumptions of a linear demand function with two unknown parameters and normally distributed disturbance terms. They show how to calculate the optimal pricing policy using a dynamic programming approach. If the intercept is known, the optimal price is of the certainty equivalent type, but this is not the case if both the intercept and slope are unknown. Three algorithms are proposed to approximate the optimal price, and the performance of two of these - CEP, and a policy based on adaptive dual control theory - are compared to each other by means of simulations.

Closely related Bayesian studies are Nguyen (1984, 1997) and Lobo and Boyd (2003). Nguyen

(1984) considers a quantity-setting monopolist firm facing random demand in multiple periods, where the demand function depends on an unknown parameter which is learned by the firm in a Bayesian fashion. Structural properties of the optimal policy are derived, and its performance is compared to a myopic one-period policy. Nguyen (1997, Section 5) discusses these questions in the context of a price-setting monopolist. Lobo and Boyd (2003) consider the same setting as Chong and Cheng (1975), and compare by means of a computer simulation the performance of four pricing policies with each other.

Cope (2007) assumes that the firm only picks prices from a finite set of predetermined selling prices. This allows for a discretization of the reservation-price distribution of customers, and the construction of a general Dirichlet prior. Cope mentions that in theory an optimal price strategy can be calculated by dynamic programming, but that in practice this is computationally intractable. He develops approximations for the value function in the dynamic program, and numerically compares the performance of the resulting pricing heuristics with CEP. In addition, he shows that his pricing heuristics converge to the optimal price if an average-reward criterion is used, and that their performance do not suffer much from a misspecified prior distribution.

Manning (1979) and Venezia (1984) are two related studies that focus on optimal design of market research. Manning (1979) considers a monopolist firm facing a finite number of customers. By doing market research, the firm can ask n potential customers about their demand at some price p. Such market research is not for free, and the main question of the paper is to determine the optimal amount of market research. This setting is closely related to pricing rules that split the selling season in two parts (e.g. the first pricing rule proposed by Witt (1986)): in the first phase, price experimentation takes place in order to learn the unknown parameters, and in the second phase of the selling season, the myopic price is used. Venezia (1984) considers a linear demand model with unknown parameters, one of which behaves like a random walk. The firm learns about these parameters using Bayes' rule. In addition, the firm can learn the true current value of this random walk by performing market research (which costs money). Using dynamic programming, the optimal market-research policy is calculated.

A common theme in the references mentioned above is that it is often intractable to compute the optimal Bayesian policy, and that therefore approximations are necessary. Rothschild (1974) points to a more fundamental problem of the Bayesian framework. He assumes that there are only two prices the firm can choose, with demand for each price Bernoulli distributed with unknown mean. The dynamic pricing problem is thus viewed as a two-armed bandit problem. The optimal Bayesian policy can be computed via the corresponding dynamic programming formulation. The key result of Rothschild (1974) is that, under the optimal Bayesian strategy, with positive probability a suboptimal price is chosen infinitely often and the optimal price is chosen only finitely many times. McLennan (1984) derives a similar conclusion in a related setting: the set of admissible prices is continuous, and the relation between price and expected demand is one of two known linear demand curves. It turns out that, under an optimal Bayesian policy, the sequence of prices may converge with positive probability to a price different from the optimal price. This work is extended by Harrison et al. (2011a), who show that in several instances a myopic Bayesian policy may lead to incomplete learning. They propose two modifications of the myopic Bayesian policy that avoid incomplete learning, and prove bounds on their performance.

The economics and econometrics literature also contains several studies on joint pricing and Bayesian learning. Prescott (1972), Grossman et al. (1977), Mirman et al. (1993) consider simple two-period models, and study the necessity and effects of price experimentation. Trefler (1993) focuses on the direction of experimentation, and applies his results on several pricing problems. Rustichini and Wolinsky (1995) and Keller and Rady (1999) consider a setting where the market environment changes in a Markovian fashion between two known demand functions, and study properties of optimal experimentation. Balvers and Cosimano (1990) consider a dynamic pric-

ing model where the coefficients of a linear demand model change over time. Easley and Kiefer (1988), Kiefer and Nyarko (1989), Aghion et al. (1991) are concerned with Bayesian learning in general stochastic control problems with uncertainty. They study the possible limits of Bayesian belief vectors, and show that in some cases these limits may differ from the true value. This implies that active experimentation is necessary to obtain strongly consistent control policies.

Non-Bayesian approaches

Despite the disadvantages of the Bayesian framework outlined above (computational intractability of the optimal solution, the negative results by Rothschild (1974) and McLennan (1984)), it has taken several decades before pricing policies in a non-Bayesian setting where studied. An early exception is Aoki (1974), who proposes a pricing scheme based on stochastic approximation in a non-Bayesian framework. He proves that the prices converge almost surely to the optimal price, and compares the policy with Bayesian pricing schemes introduced in Aoki (1973).

More recent work in a non-Bayesian context is Carvalho and Puterman (2005a,b). They propose a so-called one-step ahead pricing policy: based on a Taylor expansion of the expected revenue for the next period, the price is chosen that approximately maximizes the sum of the revenues in the next two periods. This is in contrast to certainty equivalent pricing, where only the expected revenue of one period is maximized. In Carvalho and Puterman (2005a), this idea is applied to a binomial demand distribution with expectation a logit function of the price. By means of a simulation, the performance of the policy is compared with CEP, and with a variant of CEP where each period with a certain time-dependent probability a random price is chosen. In Carvalho and Puterman (2005b), a log-normal demand distribution is assumed, and three more pricing policies are considered in the simulation.

A disadvantage of the many pricing heuristics that have been proposed in the literature, both in a Bayesian and a non-Bayesian setting, is that a qualitative statement of their performance is often missing. In many studies the performance of pricing policies is only evaluated numerically, without any analytical results. This changes with the work of Kleinberg and Leighton (2003), who suggest to quantify the performance of a pricing policy by Regret(T): the expected loss in T time periods incurred by not choosing optimal prices. They consider a setting where buyers arrive sequentially to the firm, and buy only if their willingness-to-pay (WtP) exceeds the posted price. Under some additional assumptions, they show that if the WtP of the individual buyers is an i.i.d. sample of a common distribution, then there is no pricing policy that achieves Regret$(T) = o(\sqrt{T})$; in addition, there is a pricing policy that achieves Regret$(T) = O(\sqrt{T \log(T)})$[1]. In an adversarial or worst-case setting, where the WtP of individual buyers is not assumed to be i.i.d., they show that no pricing policy can achieve Regret$(T) = o(T^{2/3})$, and that there is a pricing policy with Regret$(T) = O(T^{2/3} \log(T)^{1/3})$.

The proof that no policy can achieve regret $o(\sqrt{T})$ is quite involved and requires many assumptions on the demand function. Broder and Rusmevichientong (2012) show a $\sqrt{T}$ lower bound on the regret in a different setting, with Bernoulli distributed demand depending on two unknown parameters. The proof makes use of information-theoretic inequalities and techniques found in Besbes and Zeevi (2011). They also provide a pricing policy that exactly achieves regret $\sqrt{T}$ growth rate. If there is only a single unknown parameter, and in addition the demand curve satisfy a certain "well-separated assumption", they show that this can be improved to $\log(T)$. The key idea of this well-separated condition is that it excludes uninformative prices: prices at which the expected demand given a certain parameter estimate is equal to the true expected demand.

[1]Here $f(T) = O(g(T))$ means $\sup_{T\ in\mathbb{N}} f(T)/g(T) < \infty$, and $f(T) = o(g(t))$ means $\limsup_{T\to\infty} f(T)/g(T) = 0$, when f and g are functions on $\mathbb{N}$.

The existence of such prices appear to play an important role in the best achievable growth rate of the regret; see also the discussion on the subject (in a Bayesian setting) in Harrison et al. (2011a).

Assuming a linear demand function with normally distributed disturbance terms, Harrison et al. (2011b) show a similar $\sqrt{T}$ lower bound on the regret, using alternative proof techniques. In addition, extending Broder and Rusmevichientong (2012) and Chapter 3 of this thesis, they formulate sufficient conditions for any pricing policy to achieve regret $O(\sqrt{T})$. They also study dynamic pricing and learning for multiple products, in a setting comparable to Chapter 4.

Tehrani et al. (2012) assume that the demand model lies in a finite set of known demand functions. This enables them to formulate the dynamic pricing problem as a multi-armed bandit with dependent arms. They propose a pricing policy based on the likelihood-ratio test, and show that its regret is bounded assuming that there are no uninformative prices.

Eren and Maglaras (2010) study dynamic pricing in a robust optimization setting. They show that if an infinite number of goods can be sold during a finite time interval, it is optimal to use a price-skimming strategy. They also study settings where learning of the demand function occurs, but under the rather strong assumption that observed demand realizations are without noise.

2.3.2 Finite inventory

We here discuss the literature on dynamic pricing and learning in presence of a finite inventory that cannot be replenished. Most of the studies assume a finite selling season, corresponding to one of the most studied models in the dynamic pricing and revenue management literature. Some studies however assume an infinite time horizon, and consider the objective of maximizing total discounted reward.

Early work

Lazear (1986) considers a simplified model where a firm sells one item during at most two periods. In the first period a number of customers visit the store; if none of them buys the item, the firm adapts its prior belief on the value of the product, updates the selling price, and tries to sell the item in the second period. The author shows that the expected profit increases by having two selling periods instead of one. He extends his model in several directions, notably by allowing strategic behavior of customers who may postpone their purchase moment if they anticipate a price decrease.

Bayesian approaches

Aviv and Pazgal (2005b) start a research stream on Bayesian learning in dynamic pricing with finite inventory. They consider a continuous-time setting where customers arrive according to a Poisson process with unknown rate. Once arrived, the probability of purchasing an item is determined by a reservation price distribution. For reasons of tractability, they assume that the purchase probability is exponentially decreasing in the selling price, and that this function is known to the seller. The seller has a prior belief on the arrival rate, given by a gamma distribution, which is updated via Bayes' rule; the posterior is then also gamma distributed. After explicitly stating the optimal pricing policy in the full information case, the authors characterize the optimal pricing scheme in the incomplete-information case by means of a differential equation. This equation does in general not admit an explicit analytical solution, and therefore three

pricing heuristics are proposed: a certainty equivalent heuristic, a fixed price policy, and a naive pricing policy that ignores uncertainty on the market. Numerical experiments suggest that CEP performs quite well. An almost identical setting is studied by Lin (2006), who proposes a pricing policy and evaluates its performance via simulations.

Sen and Zhang (2009) extend the model of Aviv and Pazgal (2005b) by assuming that the purchase probabilities of arriving customers are not known to the firm. They assume that the demand distribution is an element of a finite known set, and consider a discrete-time setting with Bayesian learning and a gamma prior on the arrival rate. The optimal pricing policy can be explicitly calculated, and in an extensive computational study, its performance is compared to both a perfect-information setting and a setting where no learning occurs.

Araman and Caldentey (2009) and Farias and van Roy (2010) study a closely related problem: they consider a firm who sells a finite amount of non-perishable inventory during an infinite time horizon. The purpose is to maximize cumulative expected discounted revenues. Similar to Aviv and Pazgal (2005a), they both assume that customers arrive according to a Poisson process with unknown rate, and that arriving customers buy a product according to their reservation price, the distribution of which is known to the firm. The unknown arrival rate is learned via Bayesian updates of the prior distribution. Araman and Caldentey (2009) consider a two-point prior distribution, whereas Farias and van Roy (2010) assume that the prior is a finite mixture of gamma distributions; in both settings, the posterior distributions are in the same parametric family as the prior, which makes the problem tractable. Araman and Caldentey (2009) propose a pricing heuristic based on an asymptotic approximation of the value function of the corresponding intensity control problem. They compare its performance numerically with CEP, static pricing, and a two-price policy. Farias and van Roy (2010) propose another heuristic, called decay balancing, and show several numerical experiments that suggest that it often performs better than both the heuristic proposed by Araman and Caldentey (2009) and CEP. In addition they prove a performance bound on decay balancing, showing that the resulting expected discounted revenue is always at least one third of the optimal value. Furthermore, they consider an extension to a setting with multiple stores that may use different selling prices.

A slightly different but related setting is studied by Aviv and Pazgal (2005a). They consider a seller of finite inventory during a finite selling season, where the demand function changes according to an underlying Markov chain. The finite state space of this Markov chain, and the demand functions associated with each state, are known to the firm. The current state of the system is unknown to the seller, but it is learned via Bayesian learning based on observed demand realizations. By including this Bayesian belief vector into the state space of the dynamic pricing problem, the authors obtain a Markov decision problem with full information, which can in theory be solved. Computationally it is intractable, however, and therefore several approximate solutions are discussed.

Non-Bayesian approaches

Bertsimas and Perakis (2006) consider a non-Bayesian setting. They assume a linear demand model with normally distributed noise terms and unknown slope, intercept and variance. These parameters are estimated using least-squares linear regression. They formulate a dynamic program that, in theory, can provide the optimal pricing policy. Due to the large size of the state space, however, it is computationally intractable. Therefore several approximations are studied, and their performance is compared to each other in a computational study. The authors also consider a dynamic pricing and learning problem in a oligopolistic environment, where in addition the price elasticity of demand is slowly varying over time. They discuss methods for estimat-

ing the demand of the firm and of its competitors, mention several methods of determining the selling prices, and provide a small computational study.

Besbes and Zeevi (2009) propose a pricing algorithm for both a parametric and non-parametric setting, and consider the objective of optimizing the minimax regret. They consider an asymptotic regime, where both initial inventory and demand go to infinity at equal speed. In both a parametric and non-parametric setting they provide an upper bound on the minimax regret attained by their proposed pricing policy, and a lower bound on the minimax regret that is attained by any policy. Wang et al. (2011) improve some of these bounds, and Besbes and Zeevi (2012) extend Besbes and Zeevi (2009) to a setting where multiple products share the same finite resources.

Besbes and Maglaras (2012) consider dynamic pricing in a setting where certain financial milestone constraints in terms of sales and revenues targets are imposed. They formulate a pricing policy and study its performance in an asymptotic regime, where inventory, sales horizon, and revenue and sales target grow to infinity at equal speed.

Robust approaches

A number of studies take a robust approach, where the demand function is not learned over time, but assumed to lie in some known uncertainty set. Thiele (2006), Lim and Shanthikumar (2007), Bergemann and Schlag (2008a,b) study this in a single-product setting, and Lim et al. (2008), Thiele (2009) in a multi-product setting. A disadvantage of these robust approaches is that no learning takes place, despite the accumulation of sales data. Lobel and Perakis (2011) attempt to bridge the gap between robust and data-driven approaches to dynamic pricing, by considering a setting where the uncertainty set is deduced from data samples.

Machine-learning approaches

A considerable stream of literature on dynamic pricing and learning has emerged from the computer science community. In general, the focus of these papers is not to provide a mathematical analysis of the performance of pricing policies, but rather to design a realistic model for electronic markets and subsequently apply machine learning techniques. An advantage of this approach is that one can model many phenomena that influence the demand, such as competition, fluctuating demand, and strategic buyer behavior. A disadvantage is that these models are often too complex to analyze analytically, and insights on the behavior of various pricing strategies can only be obtained by performing numerical experiments.

Machine-learning techniques that have been applied to dynamic pricing problems are evolutionary algorithms (Ramezani et al., 2011), particle swarm optimization (Mullen et al., 2006), reinforcement learning and Q-learning (Kutschinski et al., 2003, Chinthalapati et al., 2006), simulated annealing (Xia and Dube, 2007), Markov chain Monte Carlo methods (Chung et al., 2012), the aggregating algorithm (Levina et al., 2009) by Vovk (1990), and goal-directed and derivative-following strategies in simulation (DiMicco et al., 2003).

2.3.3 Joint pricing and inventory problems

A few studies consider the problem of simultaneously determining an optimal pricing and inventory replenishment policy under demand uncertainty.

Most of them consider learning in a Bayesian framework. Subrahmanyan and Shoemaker (1996) assume that the unknown demand function lies in a finite known set of demand functions, which is learned over time in a Bayesian fashion. The optimal policy is determined by a dynamic program. Several numerical experiments are provided to offer insight in the properties of the pricing policy. Bitran and Wadhwa (1996) and Bisi and Dada (2007) study a similar type of problem, where an unknown parameter is learned in a Bayesian manner, and the optimal decisions are determined by a dynamic program. Bitran and Wadhwa (1996) perform extensive computational experiments, and Bisi and Dada (2007) derive several properties of the optimal policy. Lariviere and Porteus (1995) consider the situation of a manufacturer that sells to a retailer. The manufacturer decides on a wholesale price offered to the retailer, and the retailer has to choose an optimal inventory replenishment policy. Both learn about a parametrized demand function in a Bayesian fashion. Properties of the optimal policy, both for the manufacturer and the retailer, are studied. Gaul and Azizi (2010) assume that a product is sold in different stores. The problem is to determine optimal prices in a finite number of periods, as well as to decide if and how inventory should be reallocated between stores. Parameters of the demand function are learned by Bayesian updating, and numerical experiments are provided to illustrate the method.

Burnetas and Smith (2000) consider a joint pricing and inventory problem in a non-parametric setting. They propose an adaptive stochastic-approximation policy, and show that the expected profit per period converges to the optimal profit under complete information. A robust approach to the dynamic pricing and inventory control problem with multiple products is studied by Adida and Perakis (2006). The focus of that paper is the formulation of the robust optimization problem, and to study its complexity properties. Related is the work of Petruzzi and Dada (2002). These authors assume that there is no demand noise, which means that the unknown parameters that determine the demand function are completely known once a demand realization is observed that does not lead to stock-out.

2.4 Methodologically related areas

Dynamic pricing under uncertainty is closely related to multi-armed bandit problems. This is a class of problems that capture many essential features of optimization problems under uncertainty, including the well-known exploration-exploitation trade-off: the decision maker should properly balance the two objectives of maximizing instant reward (exploitation of current knowledge) and learning the unknown properties of the system (exploration). This trade-off between learning and instant optimization is also frequently observed in dynamic pricing problems. The literature on multi-armed bandit problems is large; some key references are Thompson (1933), Robbins (1952), Lai and Robbins (1985), Gittins (1989), Auer et al. (2002); see further Vermorel and Mohri (2005), Cesa-Bianchi and Lugosi (2006), Powell (2010). If in a dynamic pricing problem, the number of admissible selling prices is finite, the problem can be modeled as a multi-armed bandit problem. This approach is e.g. taken by Rothschild (1974), Xia and Dube (2007), and Cope (2007).

Another important area related to dynamic pricing and learning is the study of convergence rates of statistical estimates. Lai and Wei (1982) study how the speed of convergence of least-squares linear regression estimates depend on the amount of dispersion in the explanatory variables. Their results are applied in several dynamic pricing problems with linear demand functions, such as Le Guen (2008) and Cooper et al. (2012). Similarly, results on the convergence rate of maximum-likelihood estimators, as in Borovkov (1998), are crucial in the analysis of pricing policies by Broder and Rusmevichientong (2012) and Besbes and Zeevi (2011).

Chapter 3

Dynamic pricing and learning for a single product with infinite inventory

3.1 Introduction

In this chapter, we study dynamic pricing and learning for a monopolist firm that sells a single type of product with infinite inventory, in a stable market environment. In some sense, this is the most elementary dynamic-pricing-and-learning problem possible. The optimal price in the full-information case of this setting, where the firm knows the relation between demand and price, is namely just a single number. This means that with full information, *static* pricing is optimal, and the need for *dynamic* pricing is only caused by the uncertainty about the demand distribution.

The term "infinite inventory" here should be interpreted as that the firm always has sufficient inventory to meet all demand; or that, if stock-outs occur, demand can be back-logged and satisfied in later periods.

The relation between selling price and demand is assumed to belong to a known, parametrized family of demand functions. We do not require that the firm has complete knowledge on these demand distributions; only knowledge on the relation between the first two moments of demand and the selling price is assumed. This results in a generic model that includes practically all demand functions used in practice. In addition, the model is more robust to mis-specification than models where a complete demand distribution is assumed. To estimate the unknown parameters of the model, we deploy maximum quasi-likelihood estimation; this is a natural extension of maximum-likelihood estimation to settings where only the first two moments of the distribution are known. Based on these estimates, the firm has to determine selling prices for a (discrete but possibly infinite) sequence of decision moments, with the objective of maximizing the expected revenue.

An intuitively appealing pricing policy is to set the price at each decision moment equal to the price that would be optimal if the current parameter estimates were correct. Such a policy is usually called passive learning, myopic pricing, or certainty equivalent pricing, for obvious reasons: each decision is made as if the current parameter estimates are equal to their true values. We show that this policy, although intuitively appealing, is not suitable: the seller may never learn the value of the optimal price, and the parameter estimates may converge to a value different from the true parameter values. The intuition behind this negative result is that certainty equivalent pricing only focuses on instant revenue maximization, and not on optimizing the quality of the parameter estimates.

To solve this issue, we propose a new dynamic pricing policy, called Controlled Variance Pricing (CVP). They key idea is to always choose the price closest to the certainty equivalent price, but such that a certain amount of price dispersion or price variation is guaranteed. The price dispersion, measured by the sample variance the selling prices, influences the quality of the parameter estimates, and by carefully tuning the growth rate of the price dispersion, we achieve an optimal balance between learning these parameter values and optimizing revenue.

We show analytically that CVP will eventually provide the correct value of the unknown parameters, and thus the value of the optimal price. We also provide bounds on the speed of convergence. Furthermore, we obtain an asymptotic upper bound on the regret, which measures the expected amount of money lost due to not using the optimal price. In particular, we show that the regret after T time periods is $O(T^{1/2+\delta})$, where $\delta > 0$ can be chosen arbitrarily small. This bound is close to $O(\sqrt{T})$, which in several settings has been shown to be the best achievable asymptotic upper bound of the regret (see e.g. Kleinberg and Leighton, 2003, Besbes and Zeevi, 2011, Broder and Rusmevichientong, 2012). Apart from this theoretical result, we also numerically compare the performance of CVP with another existing pricing policy from the literature. These numerical experiments suggest that CVP performs well for different demand functions and time scales.

Our pricing algorithm provides several advantages over other policies from the literature. First, our analysis of Controlled Variance Pricing is valid for a large class of demand models. This is in contrast to Lobo and Boyd (2003), Carvalho and Puterman (2005a,b), Bertsimas and Perakis (2006) and Harrison et al. (2011b), where the analysis is restricted to specific models (e.g. linear) or distributions (e.g. log-normal). In addition, CVP is the first parametric approach to dynamic pricing with unknown demand where only knowledge on the first two moments of the demand distribution is required. Furthermore, the expected demand can depend on two unknown parameters, this is more natural than a single unknown parameter as assumed in the Bayesian approaches Lin (2006), Araman and Caldentey (2009), Farias and van Roy (2010), Harrison et al. (2011a).

Another feature of CVP is that it balances learning and optimization at each decision moment, enabling convergence of the prices to the optimal price. This differs from policies that strictly separate the time horizon in exploration and exploitation phases, as in Broder and Rusmevichientong (2012) or Besbes and Zeevi (2009); here only the average price converges to the optimal price. Moreover, in this latter type of policies the number of exploration prices that need to be chosen beforehand increases when the number of unknown parameters increases. CVP only requires one variable to choose, independent of the number of unknown parameters This makes the method suitable for extensions to models with multiple products, as is elaborated in Chapter 4 of this thesis. Another difference between CVP and these policies is that CVP uses all available historical data to form parameter estimates, whereas the analysis of the algorithms by Broder and Rusmevichientong (2012) and Besbes and Zeevi (2009) only uses data from the exploration phases. In light of the results of den Boer (2012b), it is unclear what happens if all available data would be used to form estimates.

The rest of this chapter is organized as follows. The demand model is described in Section 3.2.1, followed by a short discussion on the model assumptions (Section 3.2.2) and the method to estimate the unknown parameters (Section 3.2.3). We show in Section 3.3 that certainty equivalent pricing is not consistent, which motivates the introduction of Controlled Variance Pricing (CVP). We show that under this policy the parameter estimates converge to the true vale, and show that the regret admits the upper bound $O(T^{1/2+\delta})$, where T is the number of time periods and $\delta > 0$ is arbitrarily small. Section 3.4 discusses the quality of the regret bound, the relation between regret and price dispersion, differences with a related control problem, and applicability of the key ideas of CVP to other sequential decision problems. In Section 3.5, CVP is numerically compared to another pricing policy from the literature, on different time scales and different demand

functions. All mathematical proofs are contained in Section 3.6.

Notation. With $\log(t)$ we denote the natural logarithm. If $x_1, x_2, \ldots, x_t$ is a sequence, then $\bar{x}_t = \frac{1}{t}\sum_{i=1}^{t} x_i$ denotes the sample mean and $\text{Var}(x)_t = \frac{1}{t}\sum_{i=1}^{t}(x_i - \bar{x}_t)^2$ the sample variance. For a vector $x \in \mathbb{R}^n$, x^T denotes the transpose and $||x||$ denotes the Euclidean norm of x. For non-random sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$, $x_n = O(y_n)$ means that there exists a $K > 0$ such that $|x_n| \leq K|y_n|$ for all $n \in \mathbb{N}$.

3.2 Model, assumptions, and estimation method

3.2.1 Model

We consider a monopolist firm that sells a single product. Time is discretized, and time periods are denoted by $t \in \mathbb{N}$. At the beginning of each time period the firm determines a selling price $p_t \in [p_l, p_h]$. The prices $0 < p_l < p_h$ are the minimum and maximum price that are acceptable to the firm. After setting the price, the firm observes a realization d_t of the demand $D_t(p_t)$, which is a random variable, and collects revenue $p_t \cdot d_t$. We assume that the inventory is sufficient to meet all demand, i.e. stock-outs do not occur.

The random variable $D_t(p_t)$ denotes the demand in period t, against selling price p_t. Given the selling prices, the demand in different time periods is independent, and for each $t \in \mathbb{N}$ and $p_t = p \in [p_l, p_h]$, $D_t(p_t)$ is distributed as $D(p)$, for which we assume the following parametric model:

$$\begin{aligned} E[D(p)] &= h(a_0^{(0)} + a_1^{(0)} p), \\ \text{Var}[D(p)] &= \sigma^2 v(E[D(p)]). \end{aligned} \tag{3.1}$$

Here $h : \mathbb{R}_+ \to \mathbb{R}_+$ and $v : \mathbb{R}_+ \to \mathbb{R}_{++}$ are both thrice continuously differentiable known functions, with $\dot{h}(x) = \frac{\partial h(x)}{\partial x} > 0$ for all $x \geq 0$. Furthermore, σ and $a^{(0)} = (a_0^{(0)}, a_1^{(0)})$ are unknown parameters with $\sigma > 0, a_0^{(0)} > 0, a_1^{(0)} < 0$, and $a_0^{(0)} + a_1^{(0)} p_h \geq 0$.

Write $e_t = D(p_t) - E[D(p_t) \mid p_1, \ldots, p_{t-1}, d_1, \ldots, d_{t-1}]$. We make the technical assumption on the demand that for some $r > 3$,

$$\sup_{t\in\mathbb{N}} E[|e_t|^r \mid p_1, \ldots, p_{t-1}, d_1, \ldots, d_{t-1}] < \infty \text{ a.s.} \tag{3.2}$$

The expected revenue collected in a single time period where price p is used, is denoted by $r(p) = p \cdot h(a_0^{(0)} + a_1^{(0)} p)$; to emphasize the dependence on the parameter values, we write $r(p, a_0, a_1) = p \cdot h(a_0 + a_1 p)$ as a function of p and (a_0, a_1).

We assume that there is an open neighborhood $U \subset \mathbb{R}^2$ of $(a_0^{(0)}, a_1^{(0)})$ such that for all $(a_0, a_1) \in U$, $r(p, a_0, a_1)$ has a unique maximizer

$$p(a_0, a_1) = \operatorname*{arg\,max}_{p_l < p < p_h} p \cdot h(a_0 + a_1 p),$$

and such that $r''(p(a_0, a_1), a_0, a_1) < 0$. This ensures that the optimal price $p_{\text{opt}} = p(a_0^{(0)}, a_1^{(0)})$ is unique and well-defined, and lies strictly between p_l and p_h.

The marginal costs of the sold product equal zero, therefore maximizing profit is equivalent to maximizing revenue. Note that a situation with positive marginal costs $c > 0$ can easily be

captured by replacing p by $p - c$.

A pricing policy ψ is a method that for each t generates a price $p_t \in [p_l, p_h]$, based on the previously chosen prices $p_1, \ldots, p_{t-1}$ and demand realizations $d_1, d_2, \ldots, d_{t-1}$. This p_t may be a random variable.

The performance of a pricing policy is measured in terms of regret, which is the expected revenue loss caused by not using the optimal price p_{opt}. For a pricing policy ψ that generates prices $p_1, p_2, \ldots, p_T$, the regret after T time periods is defined as

$$\text{Regret}(T, \psi) = E\left[\sum_{t=1}^{T} r(p_{\text{opt}}, a^{(0)}) - r(p_t, a^{(0)})\right].$$

The objective of the seller is to find a pricing policy ψ that maximizes the total expected revenue over a finite number of T time periods. This is equivalent to minimizing Regret(T, ψ). Note however that the regret can not directly be used by the seller to find an optimal policy, since its value depends on the unknown parameters $a^{(0)}$.

3.2.2 Discussion of model assumptions

We do not assume complete knowledge about the demand distribution, only about the first two moments. This makes the demand model a little more robust to misspecifications.

In equation (3.1) we assume that the variance of demand is a function of the expectation. This holds for many common demand models, like Bernoulli (with $v(x) = x(1-x), \sigma = 1$), Poisson ($v(x) = x, \sigma = 1$) and normal distributions ($v(x) = 1$ and arbitrary $\sigma > 0$). All these examples also satisfy the moment condition (3.2).

The assumptions on the functions h, v and parameters σ, $a_0^{(0)}$, $a_1^{(0)}$ imply that the expected demand is strictly decreasing in the price, and that the variance is strictly positive; i.e. demand is non-deterministic. These are both natural assumptions on the demand distribution.

The assumption on the existence and uniqueness of $\arg\max_{p_l < p < p_h} p \cdot h(a_0 + a_1 p)$ for all a_0, a_1 in an open neighborhood U of $a^{(0)}$, is satisfied by many functions h that are used in practice to model the relation between price and expected demand; examples are $h(x) = x$, $h(x) = \exp(x)$, $h(x) = (1 + \exp(-x))^{-1}$. A sufficient condition to satisfy this assumption is that the revenue function $r(p, a^{(0)})$ is strictly concave in p, and attains it maximum strictly between p_l and p_h.

3.2.3 Estimation of unknown parameters

The unknown parameters $a^{(0)}$ can be estimated with maximum quasi-likelihood estimation. This is a natural extension of maximum-likelihood estimation to settings where only the first two moments of the distribution are known; see Wedderburn (1974), McCullagh (1983), Godambe and Heyde (1987) and the books by McCullagh and Nelder (1983), Heyde (1997) and Gill (2001).

Given prices $p_1, \ldots, p_t$ and demand realizations $d_1, \ldots, d_t$, the maximum quasi-likelihood estimator (MQLE) of $(a_0^{(0)}, a_1^{(0)})$, denoted by $\hat{a}_t = (\hat{a}_{0t}, \hat{a}_{1t})$, is the solution to the two-dimensional

equation

$$l_t(\hat{a}_t) = \sum_{i=1}^{t} \frac{\dot{h}(\hat{a}_{0t} + \hat{a}_{1t}p_i)}{\sigma^2 v(h(\hat{a}_{0t} + \hat{a}_{1t}p_i))} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (d_i - h(\hat{a}_{0t} + \hat{a}_{1t}p_i)) = 0. \tag{3.3}$$

If the probability density function (or probability mass function in case of discrete demand distribution) of $D(p)$ can be written in the form $\exp(\sigma^{-1}(d\theta - g(\theta)))$, where θ is some function of $h(a_0 + a_1 p)$, then (3.3) corresponds to the maximum-likelihood equations (Wedderburn, 1974). Many demand distributions that are used in practice, such as Poisson, Bernoulli, and normal distributions, fall in this class (see McCullagh and Nelder, 1983, Gill, 2001). In case of normally distributed demand with h the identity function, (3.3) is also equivalent to the normal equations of ordinary least squares, viz.

$$l_t(\hat{a}_t) = \sum_{i=1}^{t} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (d_i - \hat{a}_{0t} - \hat{a}_{1t}p_i) = 0. \tag{3.4}$$

The solution to the quasi-likelihood equations may in general not always be unique. A standard way to select the "right" solution is to pick the solution with lowest mean-square error, cf. Heyde (1997, Section 13.3). In our numerical results, Section 3.5, we did not encounter problems with multiple solutions of (3.3).

3.3 Performance of pricing policies

3.3.1 Inconsistency of certainty equivalent pricing

An intuitively natural pricing policy is to estimate after each time period the unknown parameters, and to set the next price equal to the price that is optimal with respect to these estimates. More precisely, choose two different initial prices $p_1, p_2 \in [p_l, p_h]$; after $t \geq 2$ time periods, calculate the MQLE estimators $\hat{a}_t$ with (3.3), and set the next price p_{t+1} equal to

$$p_{t+1} = \operatorname*{arg\,max}_{p \in [p_l, p_h]} r(p, \hat{a}_{0t}, \hat{a}_{1t}). \tag{3.5}$$

This pricing policy is known under the name certainty equivalent pricing, myopic pricing, or passive learning.

Under different settings, certainty equivalent policies are known to produce suboptimal outcomes: see e.g. the simulation results of such policies in Lobo and Boyd (2003) and Carvalho and Puterman (2005a,b). Anderson and Taylor (1976) studied a linear system $y_t = a_0^{(0)} + a_1^{(0)} x_t + \epsilon_t$ with unknown parameters $a_0^{(0)}, a_1^{(0)}$ and input variables x_t; the objective is to steer y_t to a desired value y_*. The certainty equivalent policy is to set

$$x_{t+1} = \min\left\{x_{\max}, \max\left\{x_{\min}, (y_* - \hat{a}_{0t})/\hat{a}_{1t}\right\}\right\},$$

where $\hat{a}_{0t}$, $\hat{a}_{1t}$ are the least square estimates of $a_0^{(0)}$, $a_1^{(0)}$, based on $(x_1, y_1), \ldots, (x_t, y_t)$, and $x_{\min}$, $x_{\max}$ are the minimum and maximum admissible values for x_t. Lai and Robbins (1982) showed that there are parameter values such that using this certainty equivalent policy, the controls x_t converge with positive probability to a value different from the optimal control $x = (y_* - a_0^{(0)})/a_1^{(0)}$; this implies that the certainty equivalent policy is not strongly consistent. (Interestingly, in a Bayesian setting the certainty equivalence policy is strongly consistent, as shown by

Chen and Hu, 1998). The proof idea of Lai and Robbins (1982) can easily be extended to our case, when h is the identity and v is constant. In that case the expected demand is a linear function of the price, and the MQLE equations (3.3) are equivalent to the normal equations for ordinary linear regression. A difference with Lai and Robbins (1982) is that they posed specific conditions on p_1, p_2, p_l, p_h; in our result these assumptions are left out.

Proposition 3.1. *Suppose that demand is normally distributed with constant variance and expected demand a linear function of the price (i.e. $h(x) = x$, $v(x) = 1$), and suppose that certainty equivalent pricing is used. Then with positive probability, p_t does not converge to p_{opt}.*

The idea of the proof, contained in Section 3.6, is to show by induction that with positive probability, $p_t = p_h > p_{\text{opt}}$ for all $t \geq 3$. On this event, the price sequence $(p_t)_{t\in\mathbb{N}}$ is exactly known, which facilitates analysis of the behavior of the sample path of $(\hat{a}_t)_{t\in\mathbb{N}}$.

Proposition 3.1 shows that certainty equivalent pricing is not strongly consistent for a linear demand function with constant variance. Its scope however is somewhat limited in the sense that it does only partially describe the asymptotic behavior of the policy. It is proven that with a positive, but possibly very small probability, the prices converge to $p_h \neq p_{\text{opt}}$. If this would happen in practice, the price manager would simply increase p_h. Moreover, simulations suggest that p_t may also converge to a value strictly between p_l and p_h, and that the limit price is with probability one different from p_{opt}. To provide a mathematical proof of this is, however, still an open problem.

3.3.2 Controlled Variance Pricing

An intuition for what goes wrong with the certainty equivalent policy is that the prices p_t converge "too quickly" to a certain value. As a result, not enough new information is obtained to further improve the parameter estimates, and thus they will not converge to the correct values. The key idea is to control the speed at which the prices converge. This is done by constructing a lower bound on the sample variance of the chosen prices. In particular we require that at each time period t, $\text{Var}(p)_t \geq ct^{\alpha-1}$, for some $c > 0$ and $\alpha \in (0, 1)$.

The pricing policy we propose, called Controlled Variance Pricing, chooses at each time period the certainty equivalent price (3.5), unless this means that the lower bound on the sample variance of the prices $\text{Var}(p)_{t+1} \geq c(t+1)^{\alpha-1}$ is not satisfied. In that case, the next price should be chosen not too close to the average price chosen so far; in particular, p_{t+1} is then not allowed to lie in the interval

$$TI(t) = \left(\bar{p}_t - \sqrt{c\left[(t+1)^\alpha - t^\alpha\right] \frac{t+1}{t}} \, , \bar{p}_t + \sqrt{c\left[(t+1)^\alpha - t^\alpha\right] \frac{t+1}{t}} \right), \tag{3.6}$$

which is referred to as the *taboo interval* at time t. Choosing p_{t+1} outside the taboo interval creates extra price dispersion, by guaranteeing $\text{Var}(p)_{t+1} \geq c(t+1)^{\alpha-1}$ (see Proposition 3.2).

Controlled Variance Pricing

Initialization: Choose initial prices $p_1, p_2 \in [p_l, p_h], p_1 \neq p_2$.
Choose $\alpha \in (0, 1)$ and $c \in \left(0, 2^{-\alpha}(p_1 - p_2)^2 \min\{1, (3\alpha)^{-1}\}\right)$.
For all $t \geq 2$:
Estimation: Calculate the MQLE estimates $\hat{a}_t$ according to (3.3).
Pricing: If
(a) there is no solution $\hat{a}_t$, or

(b) $\hat{a}_{0t} \leq 0$ or $\hat{a}_{1t} \geq 0$, or
(c) $\hat{a}_{0t} + \hat{a}_{1t}p < 0$ for some $p \in [p_l, p_h]$,
set $p_{t+1} \in \{p_1, p_2\}$ such that $|p_{t+1} - \bar{p}_t| = \max(|p_1 - \bar{p}_t|, |p_2 - \bar{p}_t|)$.
Now assume $\hat{a}_t$ exists and $\hat{a}_{0t} > 0$, $\hat{a}_{1t} < 0$, $\hat{a}_{0t} + \hat{a}_{1t}p \geq 0$ for all $p \in [p_l, p_h]$.
Set

$$p_{t+1} = \underset{p \in [p_l, p_h]}{\arg\max}\, r(p, \hat{a}_t), \tag{3.7}$$

if this results in $\text{Var}(p)_{t+1} \geq c(t+1)^{\alpha-1}$.
Else, set

$$p_{t+1} = \underset{p \in [p_l, p_h] \setminus TI(t)}{\arg\max}\, r(p, \hat{a}_t), \tag{3.8}$$

where $TI(t)$ is the taboo interval (3.6) at time t.

In cases (a) - (c), we choose one of the initial prices p_1, p_2, that is most far away from $\bar{p}_t$. This ensures that the bound on the variance $\text{Var}(p)_t \geq ct^{\alpha-1}$ remains valid. The upper bound on the constant c ensures that $[p_l, p_h] \setminus TI(t)$ is nonempty for all $t \geq 2$, and that $\text{Var}(p)_t \geq ct^{\alpha-1}$ is satisfied for $t = 2$.

A desirable property of a pricing policy is that the price p_t converges to the optimal price p_{opt}, and thus the sample variance $\text{Var}(p)_t$ converges to zero. The speed at which the sample variance goes to zero turns out to be strongly related to the quality of the parameter estimates: in particular, the parameter estimates $\hat{a}_t$ converge quickly to the correct values $a^{(0)}$ if the sample variance $\text{Var}(p)_t$ converges slowly to zero; then the price however converges slowly, which may be costly. We here observe a trade-off between exploration (quick convergence of parameter estimates to the correct values) and exploitation (quick convergence of prices to the optimal price). The balance between exploration and exploitation is captured in the parameter α of CVP. The following proposition establishes a relation between α and the sample variance of the prices:

Proposition 3.2. *With CVP, $Var(p)_t \geq ct^{\alpha-1}$ for all $t \geq 2$.*

The results on consistency and convergence rates of parameter estimates that we use to establish performance bounds for CVP, are stated in terms of the eigenvalues of the design matrix. The following lemma relates these eigenvalues to the sample variance of the prices. Its proof is straightforward and contained in Section 3.6.

Lemma 3.1. *Let $\lambda_{max}(t), \lambda_{min}(t)$ be the largest and smallest eigenvalue of the design matrix*

$$P_t = \begin{pmatrix} t & \sum_{i=1}^t p_i \\ \sum_{i=1}^t p_i & \sum_{i=1}^t p_i^2 \end{pmatrix}, \quad (t \geq 2),$$

where $p_1, \ldots, p_t \in [p_l, p_h]$ and $p_1 \neq p_2$. Then $\lambda_{max}(t) \leq (1 + p_h^2)t$ and $tVar(p)_t \leq (1 + p_h^2)\lambda_{min}(t)$.

In the following Proposition 3.3 and Theorem 3.1, we assume that CVP with $\alpha \in (1/2, 1)$ is used. We show that a solution $\hat{a}_t$ to the estimation equations (3.3) eventually exists, and the parameter estimates $\hat{a}_t$ converge to the correct value $a^{(0)}$. In addition we provide an upper bound on the mean square convergence rate, in terms of the parameter α.

Proposition 3.3. *Let $\alpha > 1/2$. A solution $\hat{a}_t$ to* (3.3) *eventually exists, and $\hat{a}_t \to a^{(0)}$ a.s. In addition, if we define*

$$T_\rho = \sup\{t \in \mathbb{N} \mid \textit{ there is no solution } \hat{a}_t \textit{ of } (3.3) \textit{ such that } \left|\left|\hat{a}_t - a^{(0)}\right|\right| \leq \rho\}, \tag{3.9}$$

then there exists a $\rho_0 > 0$ *such that for all* $0 < \rho < \rho_0$, $E[T_\rho^{1/2}] < \infty$ *and*

$$E\left[\left|\left|\hat{a}_t - a^{(0)}\right|\right|^2 \mathbf{1}_{t>T_\rho}\right] = O\left(\frac{\log(t)}{t^\alpha}\right), \tag{3.10}$$

where $\mathbf{1}_{t>T_\rho}$ *denotes the indicator function of the event* $t > T_\rho$.

The proposition follows from Chapter 7, where strong consistency and convergence rates for quasi-likelihood estimates are discussed. Theorem 7.1 implies the assertion $E[T_\rho^{1/2}] < \infty$. (Note that there, the required condition $1/2 < r\alpha - 1$ is valid for all $1/2 < \alpha \leq 1$, because of our moment condition (3.2), with $r > 3$). The convergence rates (3.10) follow from Theorem 7.2 and Remark 7.2 in Chapter 7, together with Lemma 3.1.

Proposition 3.3 enables us to calculate the following upper bound on the regret:

Theorem 3.1.

$$\textit{Regret}(T, CVP) = O\left(T^\alpha + T^{1-\alpha}\log(T)\right),$$

provided $\alpha > 1/2$.

We use a Taylor series expansion of the revenue function $r(p)$ to show $|r(p) - r(p_{\text{opt}})| = O((p - p_{\text{opt}})^2)$. The implicit function theorem is invoked to obtain $|p(a) - p_{\text{opt}}| = O\left(||a - a^{(0)}||\right)$. The theorem can then be derived from Proposition 3.3 and the rate at which the size of the taboo interval converges to zero. The details of the proof are given in Section 3.6.

3.4 Discussion

3.4.1 Quality of regret bound

The term $T^{1-\alpha}\log(T)$ in Theorem 3.1 comes from bounds (3.10) on the quality of the parameter estimates, the term T^α comes from the length of the taboo interval (3.6). The parameter α captures the trade-off between learning and optimization. If α is large then much emphasis is put on learning: the parameters converge quickly to their correct values, but due to the large size of the taboo interval, the prices converge slowly. If α is small then the emphasis is on optimizing instant revenue: the taboo interval is then very small, thus the next-period price is close to the certainty equivalent optimal price; however, for small α, only a relatively slow convergence of the parameter estimates is guaranteed by Proposition 3.3. The optimal choice of α in Theorem 3.1 clearly is $1/2$, but since α should be larger than $1/2$, we get the following

Corollary 3.1.

$$\textit{Regret}(T, CVP) = O\left(T^{1/2+\delta}\right),$$

with $\alpha = 1/2 + \delta$, *for arbitrarily small* $\delta > 0$.

This result would be a little more elegant if the term T^δ, $\delta > 0$, could be removed. The relevant theorems of Chapter 7 however require $\alpha > 1/2$ and it appears that in general this requirement cannot easily be removed.

The bound from Corollary 3.1 is close to $O(\sqrt{T})$. In several settings it has been shown that this is the best achievable asymptotic upper bound on the regret (see e.g. Kleinberg and Leighton, 2003, Besbes and Zeevi, 2011, Broder and Rusmevichientong, 2012). It is not completely clear if the "gap" T^δ between Corollary 3.1 and the best achievable bound $\sqrt{T}$ can be removed. By using much technical machinery, one can make this gap slightly smaller: the multi-product pricing

policy from Chapter 4, applied to the single-product setting, achieves Regret$(T) = O(\sqrt{T \log(T)})$. However, here there is still a "gap" of $\sqrt{\log(T)}$. A further discussion on this issue is provided in section 4.4.4.

3.4.2 Generality of result

Concerning the generality of the result, we note that Proposition 3.3 holds for any pricing policy that guarantees $\text{Var}(p)_t \geq ct^{\alpha-1}$. The relation between regret and $\text{Var}(p)_t$ through the parameter α, as in Theorem 3.1, depends however on the specifics of the used pricing policy.

A lower bound on Regret(t) in terms of $\text{Var}(p)_t$ can easily be constructed. For example, if the revenue function $r(p)$ is strictly concave in p, then one can show that $\text{Regret}(t, \psi) \geq k \sum_{i=1}^{t} (p_i - p_{\text{opt}})^2$, for some positive constant k and any policy ψ. Since

$$\arg\min_{p \in \mathcal{P}} \sum_{i=1}^{t} (p_i - p)^2 = \bar{p}_t,$$

this implies $t\text{Var}(p)_t \leq k^{-1}\text{Regret}(t, \psi)$ a.s. To derive an upper bound on the regret in terms of the growth rate of $\text{Var}(p)_t$, for arbitrary policies ψ, seems much less straightforward. It is an interesting direction for future research to completely characterize the relation between regret and empirical variance.

3.4.3 Differences with the multiperiod control problem

For the multiperiod control problem mentioned in Section 3.3.1, Lai and Robbins (1982) showed that there is a policy with Regret$(T) = O(\log(T))$. This problem is very much akin to the dynamic pricing problem with linear demand function: in the first problem the optimal control $x(\hat{a}_{0t}, \hat{a}_{1t})$ as function of the parameter estimates equals $(y^* - \hat{a}_{0t})/\hat{a}_{1t}$, in the latter problem the optimal price $p(\hat{a}_{0t}, \hat{a}_{1t})$ equals $-\hat{a}_{0t}/(2\hat{a}_{1t})$ (for the moment neglecting bounds on x, p and assuming $\hat{a}_{0t}, \hat{a}_{1t}$ have the correct sign). When $y^* = 0$, these optimal controls only differ by a factor 2. An intuitive explanation why we do not achieve Regret$(T) = O(\log(T))$ in the dynamic pricing problem, despite the similarities with the multiperiod control problem, is the presence of what Harrison et al. (2011a) call "indeterminate equilibria". An indeterminate equilibrium occurs if there are estimates $(\hat{a}_0, \hat{a}_1)$ such that the average observed output at $x(\hat{a}_0, \hat{a}_1)$ "confirms" the correctness of these estimates, i.e. if $(\hat{a}_0, \hat{a}_1)$ satisfies $a_0^{(0)} + a_1^{(0)} x(\hat{a}_0, \hat{a}_1) = \hat{a}_0 + \hat{a}_1 x(\hat{a}_0, \hat{a}_1)$. It is not difficult to show that there are infinitely many indeterminate equilibria, both in the multiperiod control problem and the dynamic pricing problem. In the multiperiod control problem, each indeterminate equilibrium $(\hat{a}_0, \hat{a}_1) \neq (a_0^{(0)}, a_1^{(0)})$ still gives an optimal control $x(\hat{a}_0, \hat{a}_1) = x(a_0^{(0)}, a_1^{(0)})$, while in the dynamic pricing problem, each indeterminate equilibrium $(\hat{a}_0, \hat{a}_1) \neq (a_0^{(0)}, a_1^{(0)})$ yields a suboptimal price $p(\hat{a}_0, \hat{a}_1) \neq p(a_0^{(0)}, a_1^{(0)})$. This means that in the multiperiod control problem, convergence of the parameter estimates to any arbitrary indeterminate equilibrium implies convergence of the controls to the optimal control; while in the dynamic pricing problem, *only* convergence of the parameter estimates to the "true" indeterminate equilibrium $a^{(0)}$ implies convergence of the controls to the optimal control. This makes the dynamic pricing problem structurally more complex than the multiperiod control problem.

3.4.4 Applicability to other sequential decision problems

In Sections 3.3.1 and 3.3.2 we discuss inconsistency of certainty equivalent pricing, and study the performance of Controlled Variance pricing, in the specific context of dynamic pricing under uncertainty. We believe however that the ideas developed in this paper can be applied to many other types of sequential decision problems with uncertainty. We provide a brief sketch of problems for which the controlled variance pricing idea may be a fruitful approach. At each time instance $t \in \mathbb{N}$ the decision maker chooses a control $x_t \in \mathbb{R}^d$, ($d \in \mathbb{N}$), and observes a realization y_t of a random variable $Y(x_t, \theta)$, whose probability distribution depends on x_t and on an unknown parameter θ in a parameter space $\Theta \subset \mathbb{R}^d$; subsequently a cost $c(x_t, y_t, \theta)$ is encountered. The decision maker estimates θ using historical data $(x_i, y_i)_{i \leq t}$, with an appropriate statistical estimation technique (e.g. maximum-likelihood estimation). A certainty equivalent control rule then sets the next control to

$$x_{t+1} = \arg\min_x E[c(x, Y(x, \hat{\theta}_t), \hat{\theta}_t)], \tag{3.11}$$

where $\hat{\theta}_t$ denotes the estimate of θ at time t. If the quality of the parameter estimates $||\hat{\theta}_t - \theta||$ depends on some measure of dispersion of the controls, then a controlled variance rule sets the next control to (3.11), subject to a lower bound on the measure of dispersion. The optimal lower bound depends on the problem characteristics, and captures in some sense the trade-off between estimation and instant optimization, i.e. exploration and exploitation.

3.5 Numerical evaluation

We numerically compare the performance of Controlled Variance Pricing to the policy MLE-cycle, which was introduced by Broder and Rusmevichientong (2012). We test normally, Poisson, and Bernoulli distributed demand, since these are commonly used demand models in practice. For each distribution we test two functions h that models the relation between price and expected demand. The function v need not be specified, since it is already determined by the demand distribution: $v(x) = 1$ for normal demand, $v(x) = x$ for Poisson demand, and $v(x) = x(1-x)$ for Bernoulli demand. All six sets of demand distribution and the function h are listed in Table 3.1. For each set, we randomly generate 10,000 different instances of parameters a_0, a_1. For normally distributed demand we also generate a value for σ; for Poisson and Bernoulli demand, $\sigma = 1$. The parameters are drawn from a uniform distribution. The support of these uniform distributions is chosen such that the optimal price lies between 3 and 8. For normal demand an additional requirement is that $h(a_0 + a_1 p_{\text{opt}}) - 3\sigma > 0$ and $\frac{\sigma}{h(a_0 + a_1 p_{\text{opt}})} > \frac{1}{20}$. This implies that at the optimal price, the probability that demand is negative is small (less than 0.135 %), and the coefficient of variation at the optimal price is not extremely small (at least $\frac{1}{20}$). For Bernoulli demand an additional requirement is $h(a_0 + a_1 p) \in (0, 1)$ for all $p_l \leq p \leq p_h$. Table 3.2 list summary statistics for the chosen parameter values.

For both policies that we compare, the lowest and highest admissible price are set to $p_l = 1$, $p_h = 10$. The policy CVP uses $\alpha = 0.5001$, and initial prices $p_1 = 4$, $p_2 = 7$. For the constant c in the taboo interval, we try three different values: 1, 3, and 5. The exploration prices of MLE-cycle are set to $p_1 = 4$, $p_2 = 7$. We vary the number of exploration phases per cycle. In particular we try 1, 2, and 3 consecutive exploration phases (n consecutive exploration phases means that during the $2n$ exploration periods in each cycle, the price alternates between p_1 and p_2).

For each set of instances we calculate the average relative regret over 10,000 instances. Thus, for

Table 3.1: Problem sets, with parameter range

	Distr.	h(x)	a_0	a_1	σ
1.	Normal	x	[0.1, 20]	$[\frac{-a_0}{11}, \frac{-a_0}{16}]$	$[\frac{1}{20}, \frac{1}{3}] \cdot (a_0 + a_1 p_{\text{opt}})$
2.	Normal	$x^{3/4}$	[0.1, 20]	$[\frac{-a_0}{11}, \frac{-a_0}{14}]$	$[\frac{1}{20}, \frac{1}{3}] \cdot (a_0 + a_1 p_{\text{opt}})^{3/4}$
3.	Poisson	exp(x)	$[\frac{11}{3}, 20]$	$[\frac{-1}{3}, \frac{-1}{8}]$	1
4.	Poisson	x	$[\frac{11}{3}, 20]$	$[\frac{-a_0}{11}, \frac{-a_0}{16}]$	1
5.	Bernoulli	$(1 + \exp(-x))^{-1}$	$[\log(-3a_1 - 1) - 3a_1, \log(-8a_1 - 1) - 8a_1]$	$[-1, \frac{-4}{9}]$	1
6.	Bernoulli	$x^{3/4}$	[0.8, 1.1]	$[\frac{-a_0}{11}, \frac{-a_0}{14}]$	1

each instance, corresponding to a choice of parameter values, we measure the relative regret

$$\frac{\sum_{t=1}^{T} r_{\text{opt}} - r(p_t)}{T r_{\text{opt}}} \times 100\%,$$

and then we average over all instances:

$$\frac{1}{10,000} \sum_{i=1}^{10,000} \frac{\sum_{t=1}^{T} r_{\text{opt}} - r(p_t)}{T r_{\text{opt}}} \times 100\%.$$

This quantity is measured for $T \in \{10, 50, 100, 500, 1000\}$. The results are listed in Table 3.3. In these tables, the header CVP(c) denotes the policy CVP with constant c, the header MLE-c(n) denotes the policy MLE-cycle with n consecutive exploration phases.

Table 3.2: Sample statistics of parameters

Problem set 1

	a_0	a_1	σ	p_{opt}
max	19.9973	-0.0066	3.2775	7.9993
mean	10.0518	-0.7712	0.9652	6.5984
min	0.1050	-1.8004	0.0042	5.5002
std	5.7519	0.4517	0.7246	0.7187

Problem set 2

	a_0	a_1	σ	p_{opt}
max	19.9989	-0.0075	2.8178	7.9998
mean	10.0050	-0.8125	0.8181	7.0703
min	0.1009	-1.8044	0.0037	6.2860
std	5.7400	0.4704	0.6135	0.4964

Problem set 3

	a_0	a_1	σ	p_{opt}
max	19.9995	-0.1250	1.0000	7.9991
mean	11.8249	-0.2286	1.0000	4.7182
min	3.6669	-0.3333	1.0000	3.0004
std	4.7345	0.0600	0	1.3508

Problem set 4

	a_0	a_1	σ	p_{opt}
max	19.9983	-0.2353	1.0000	7.9991
mean	11.8751	-0.9094	1.0000	6.6062
min	3.6687	-1.8095	1.0000	5.5006
std	4.7217	0.3762	0	0.7230

Problem set 5

	a_0	a_1	σ	p_{opt}
max	9.8596	-0.4445	1.0000	7.9998
mean	4.8056	-0.7255	1.0000	5.3353
min	0.3068	-1.0000	1.0000	3.0006
std	1.9504	0.1606	0	1.4570

Problem set 6

	a_0	a_1	σ	p_{opt}
max	1.1000	-0.0574	1.0000	8.0000
mean	0.9497	-0.0770	1.0000	7.0780
min	0.8000	-0.0997	1.0000	6.2858
std	0.0866	0.0088	0	0.4952

The results from Table 3.3 suggest that CVP performs comparable to MLE-cycle, or even better. This hold for all tested time scales $T = 10, 50, 100, 500$ and 1000, and all six sets of problem instances. If we consider the results for $T = 1000$, we see that CVP outperforms MLE-cycle on all problem sets except 1. On a shorter time scale, $T = 100$, this holds for all problem sets. One of the reasons for this difference may be that MLE-cycle only uses the data from the exploration phases to form parameter estimates, whereas CVP uses all the available historical data. (Note, however, that den Boer (2012b) shows that adding data does not necessarily improve the quality of parameter estimates).

Table 3.3: Average relative regret

Problem set 1: Normal demand, $h(x) = x$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	5.0 %	5.0 %	5.0 %	7.6 %	8.9 %	8.6 %
50	3.2 %	3.1 %	3.2 %	5.0 %	6.7 %	7.2 %
100	2.9 %	2.9 %	2.9 %	3.9 %	5.3 %	6.5 %
500	2.7 %	2.7 %	2.7 %	2.0 %	3.0 %	4.1 %
1000	2.7 %	2.6 %	2.7 %	1.5 %	2.2 %	3.2 %

Problem set 2: Normal demand, $h(x) = x^{3/4}$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	6.8 %	7.2 %	7.5 %	9.4 %	11.0 %	10.4 %
50	4.0 %	3.7 %	3.8 %	7.0 %	8.6 %	8.9 %
100	3.2 %	2.8 %	2.8 %	5.9 %	7.0 %	8.1 %
500	1.9 %	1.4 %	1.4 %	3.4 %	4.0 %	5.2 %
1000	1.4 %	1.0 %	1.0 %	2.6 %	3.0 %	4.1 %

Problem set 3: Poisson demand, $h(x) = \exp(x)$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	2.3 %	2.7 %	3.3 %	5.8 %	7.7 %	9.1 %
50	0.9 %	1.3 %	1.9 %	3.1 %	6.3 %	7.3 %
100	0.6 %	1.0 %	1.4 %	2.3 %	5.0 %	6.6 %
500	0.3 %	0.4 %	0.7 %	1.2 %	2.9 %	4.2 %
1000	0.2 %	0.3 %	0.5 %	0.8 %	2.2 %	3.3 %

Problem set 4: Poisson demand, $h(x) = x$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	8.1 %	8.6 %	9.1 %	9.4 %	9.5 %	8.7 %
50	5.5 %	5.5 %	5.6 %	8.5 %	8.1 %	8.0 %
100	4.8 %	4.5 %	4.3 %	7.6 %	6.9 %	7.3 %
500	3.4 %	2.7 %	2.4 %	4.9 %	4.2 %	4.8 %
1000	2.8 %	2.1 %	1.9 %	3.9 %	3.2 %	3.8 %

Problem set 5: Bernoulli demand, $h(x) = (1 + \exp(-x))^{-1}$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	18.4 %	18.5 %	18.3 %	21.0 %	19.2 %	20.7 %
50	9.5 %	10.0 %	10.5 %	15.8 %	17.1 %	18.0 %
100	6.8 %	7.2 %	7.6 %	13.5 %	14.4 %	16.5 %
500	3.6 %	3.5 %	3.5 %	8.6 %	8.9 %	11.0 %
1000	2.8 %	2.5 %	2.5 %	6.8 %	6.9 %	8.7 %

Problem set 6: Bernoulli demand, $h(x) = x^{3/4}$

t	CVP(1)	CVP(3)	CVP(5)	MLE-c (1)	MLE-c (3)	MLE-c (5)
10	11.3 %	11.5 %	11.6 %	11.4 %	12.3 %	10.4 %
50	9.2 %	9.8 %	10.1 %	11.1 %	10.8 %	10.4 %
100	8.0 %	8.3 %	8.4 %	11.0 %	10.3 %	10.0 %
500	5.8 %	5.4 %	5.0 %	9.9 %	8.1 %	7.9 %
1000	5.0 %	4.4 %	3.9 %	9.0 %	6.8 %	6.5 %

For some instances, in particular problem set 3, the relative regret decreases very fast: CVP(1) has regret below 1% already from $T = 50$. The sets 5 and 6, with Bernoulli distributed demand, show a more slowly decreasing relative regret. We also see that the optimal value of the constant c in CVP depends on T. For example in problem set 6, $c = 1$ performs best for $T = 10, 50, 100$, whereas $c = 5$ is the best choice for $T = 500, 1000$.

We wish to emphasize that these results are not meant as an exhaustive comparison between the numerical performance of CVP and MLE-cycle. In that case we should also have fine-tuned the value of the exploration prices p_1, p_2. The simulation results nevertheless are an indication that CVP may perform well in practical applications.

3.6 Proofs

Proof of Proposition 3.1

The proof is similar to Section 2 of Lai and Robbins (1982), with the difference that we do not make assumptions on the values of p_1, p_2, p_l, p_h.

Without loss of generality assume $p_1 < p_2$, define $a = ((p_h - p_1)^2 + (p_h - p_2)^2)p_h^{-1}$, and recall that $e_t = D(p_t) - E[D(p_t) \mid p_1, \ldots, p_{t-1}, d_1, \ldots, d_{t-1}]$. Write $\sigma^2 = E[e_t^2]$, for all $t \in \mathbb{N}$. Let $\delta > 0$, and consider the event

$$A = \left\{ \begin{array}{l} (p_2 - 2p_h)e_1 + (2p_h - p_1)e_2 \geq -a_1^{(0)}(p_2 - p_1)2p_h \\ (p_1 - p_h)e_1 + (p_2 - p_h)e_2 \geq (-2p_h a_1^{(0)} - a_0^{(0)} + \delta)a + (2p_h - p_1 - p_2)\delta \\ |\bar{e}_t| \leq \delta \text{ for all } t \geq 3 \end{array} \right\}.$$

We first show that for sufficiently large δ, the event A occurs with strictly positive probability. The first two inequalities of A are satisfied when

$$\left(\begin{array}{cc} p_2 - 2p_h & 2p_h - p_1 \\ p_1 - p_h & p_2 - p_h \end{array} \right) \left(\begin{array}{c} e_1 \\ e_2 \end{array} \right) \geq \left(\begin{array}{c} -a_1^{(0)}(p_2 - p_1)2p_h \\ a(-2p_h a_1^{(0)} - a_0^{(0)} + \delta) + (2p_h - p_1 - p_2)\delta \end{array} \right). \tag{3.12}$$

The determinant of the coefficient matrix equals $(p_2 - 2p_h)(p_2 - p_h) + (p_1 - p_h)(p_1 - 2p_h)$, which is strictly positive. A solution to this linear system therefore exists, and (3.12) happens with positive probability. Let $B \subset \mathbb{R}^2$ be a bounded subset of the solutions (e_1, e_2) of (3.12), s.t. $P((e_1, e_2) \in B) > 0$. Choose $\delta > \sqrt{8}\sigma + \sup_{(e_1,e_2)\in B} \frac{1}{3}|e_1 + e_2|$. It follows from the Kolmogorov inequality (see e.g. Chow and Teicher, 2003, Theorem 6, page 133) that for any $\epsilon > \sqrt{8}\sigma$,

$$\begin{aligned} & P\left(\sup_{3 \leq t} \left| \frac{1}{t-2} \sum_{i=3}^{t} e_i \right| > \epsilon \right) \leq \sum_{j=1}^{\infty} P\left(\sup_{2^j < t \leq 2^{j+1}} \left| \frac{1}{t-2} \sum_{i=3}^{t} e_i \right| > \epsilon \right) \\ \leq & \sum_{j=1}^{\infty} P\left(\sup_{2^j < t \leq 2^{j+1}} |\sum_{i=3}^{t} e_i| > (2^j - 1)\epsilon \right) \leq \sum_{j=1}^{\infty} P\left(\sup_{1 \leq t \leq 2^{j+1}} |\sum_{i=3}^{t} e_i| > (2^j - 1)\epsilon \right) \\ \leq & \sum_{j=1}^{\infty} \frac{1}{(2^j - 1)^2 \epsilon^2} \sigma^2 2^{j+1} \leq \sum_{j=1}^{\infty} 8\sigma^2 \epsilon^{-2} 2^{-j} \\ = & 8\sigma^2 \epsilon^{-2} < 1, \end{aligned}$$

since $2^{j+1}/(2^j-1)^2 \le 8\cdot 2^{-j}$, $j \ge 1$. This implies

$$\begin{aligned}
P(|\bar{e}_t| \le \delta \text{ for all } t \ge 3) &= P(\sup_{t\ge 3}|\bar{e}_t| \le \delta) = P\left(\sup_{t\ge 3}\left|\frac{e_1+e_2}{t} + \frac{1}{t}\sum_{i=3}^{t} e_i\right| \le \delta\right) \\
&\ge P\left((e_1,e_2)\in B \text{ and } \sup_{t\ge 3}\left|\frac{1}{t}\sum_{i=3}^{t} e_i\right| \le \delta - \sup_{(e_1,e_2)\in B}\frac{1}{3}|e_1+e_2|\right) \\
&= P\left((e_1,e_2)\in B\right)\cdot P\left(\sup_{t\ge 3}\left|\frac{1}{t}\sum_{i=3}^{t} e_i\right| \le \left(\delta - \sup_{(e_1,e_2)\in B}\frac{1}{3}|e_1+e_2|\right)\right) \\
&\ge P\left((e_1,e_2)\in B\right)\cdot P\left(\sup_{t\ge 3}\left|\frac{1}{t-2}\sum_{i=3}^{t} e_i\right| \le \left(\delta - \sup_{(e_1,e_2)\in B}\frac{1}{3}|e_1+e_2|\right)\right) \\
&> 0.
\end{aligned}$$

This proves that for δ sufficiently large, the event A occurs with probability $P(A) > 0$.

If for some t, $\hat{a}_{1t} \ge 0$ or $\hat{a}_{0t} \le 0$ then clearly the parameter estimates have the wrong sign; it would be foolish for a price manager to use the certainty equivalent price

$$p_{t+1} = \arg\max_{p_l \le p \le p_h} p\cdot(\hat{a}_{0t} + \hat{a}_{1t}p)$$

in that case. We therefore assume that $p_{t+1} = p_h$ whenever $\hat{a}_{1t} \ge 0$. (Alternatively one might impose some extra conditions in the set A, and still use the certainty equivalent price when the estimates have the wrong sign).

We show by induction that on the event A, $p_{t+1} = p_h$, for all $t \ge 2$.

t=2. The line through the points $(p_i, a_0^{(0)} + a_1^{(0)}p_i + e_i)$, $i = 1, 2$ has slope $\hat{a}_{12} = a_1^{(0)} + (e_2-e_1)(p_2-p_1)^{-1}$ and intercept $\hat{a}_{02} = (e_1p_2 - e_2p_1)(p_2-p_1)^{-1}$. When $\hat{a}_{12} \ge 0$ then $p_3 = p_h$. If $\hat{a}_{12} < 0$ then $p_3 = \frac{\hat{a}_{02}}{-2\hat{a}_{12}} \ge p_h$ is implied by

$$(e_1p_2 - e_2p_1)(p_2-p_1)^{-1} \ge -2(a_1^{(0)} + (e_2-e_1)(p_2-p_1)^{-1})p_h,$$

which, by multiplying (p_2-p_1) and rearranging terms, is equivalent to the condition

$$(p_2 - 2p_h)e_1 + (2p_h - p_1)e_2 \ge -a_1^{(0)}(p_2-p_1)2p_h.$$

t $\ge$ 3. Suppose that for all $i = 3, \ldots, t$, $p_i = p_h$. Then $\bar{p}_i = p_h - \frac{2p_h - p_1 - p_2}{i}$ $(3 \le i \le t)$. Defining $C_t = \sum_{i=1}^{t}(p_i - \bar{p}_t)e_i$, and $V_t = \sum_{i=1}^{t}(p_i - \bar{p}_t)^2$, the least-squares estimates are

$$\begin{pmatrix} \hat{a}_{0t} \\ \hat{a}_{1t} \end{pmatrix} = \begin{pmatrix} a_0^{(0)} \\ a_1^{(0)} \end{pmatrix} + \begin{pmatrix} \bar{e}_t - \bar{p}_t C_t/V_t \\ C_t/V_t \end{pmatrix}.$$

For all $t \ge 2$, V_t and C_t can be rewritten as

$$V_t = \sum_{i=2}^{t}\frac{i-1}{i}(p_i - \bar{p}_{i-1})^2, \qquad C_t = \sum_{i=2}^{t}\frac{i-1}{i}(p_i - \bar{p}_{i-1})(e_i - \bar{e}_{i-1}).$$

and by some algebra and an induction argument, it follows that

$$V_t = V_2 + (2p_h - p_1 - p_2)^2(\frac{1}{2} - t^{-1}), \qquad C_t = C_2 + (2p_h - p_1 - p_2)(\bar{e}_t - \bar{e}_2).$$

where $V_2 = \frac{1}{2}(p_2 - p_1)^2$ and $C_2 = \frac{1}{2}(p_2 - p_1)(e_2 - e_1)$.

If $\hat{a}_{1t} \geq 0$ then $p_{t+1} = p_h$.

Now suppose $\hat{a}_{1t} < 0$. Then

$$\begin{aligned}
& p_{t+1} = p_h \\
\Leftrightarrow & \frac{\hat{a}_{0t}}{-2\hat{a}_{1t}} \geq p_h \\
\Leftrightarrow & \hat{a}_{0t} \geq -2\hat{a}_{1t} p_h \\
\Leftrightarrow & a_0^{(0)} + \bar{e}_t - \bar{p}_t C_t / V_t \geq -2p_h a_1^{(0)} - 2p_h C_t / V_t \\
\Leftrightarrow & \bar{e}_t + (2p_h - \bar{p}_t) C_t / V_t \geq -2p_h a_1^{(0)} - a_0^{(0)} \\
\Leftrightarrow & (p_h + \frac{2p_h - p_1 - p_2}{t}) C_t / V_t \geq -2p_h a_1^{(0)} - a_0^{(0)} - \bar{e}_t.
\end{aligned}$$

Observe that on the event A, $-\bar{e}_t < \delta$, $p_h + \frac{2p_h - p_1 - p_2}{t} \geq p_h$, $V_t \leq V_2 + (2p_h - p_1 - p_2)^2 \cdot \frac{1}{2}$, $C_t = C_2 + (2p_h - p_1 - p_2)(\bar{e}_t - \bar{e}_2) \geq C_2 + (2p_h - p_1 - p_2)(-\delta - \bar{e}_2)$ and thus it suffices to show

$$C_2 + (2p_h - p_1 - p_2)(-\delta - \bar{e}_2) \geq (-2p_h a_1^{(0)} - a_0^{(0)} + \delta)(V_2 + (2p_h - p_1 - p_2)^2 \cdot \frac{1}{2}) p_h^{-1},$$

i.e.

$$\begin{aligned}
& \frac{1}{2}(p_2 - p_1)(e_2 - e_1) - (2p_h - p_1 - p_2)\bar{e}_2 \\
\geq & (-2p_h a_1^{(0)} - a_0^{(0)} + \delta)(\frac{1}{2}(p_2 - p_1)^2 + (2p_h - p_1 - p_2)^2 \cdot \frac{1}{2}) p_h^{-1} + (2p_h - p_1 - p_2)\delta.
\end{aligned}$$

Rewriting the lefthandside we get the condition

$$\begin{aligned}
& e_1(p_1 - p_h) + e_2(p_2 - p_h) \\
\geq & (-2p_h a_1^{(0)} - a_0^{(0)} + \delta)(\frac{1}{2}(p_2 - p_1)^2 + (2p_h - p_1 - p_2)^2 \cdot \frac{1}{2}) p_h^{-1} + (2p_h - p_1 - p_2)\delta \\
= & (-2p_h a_1^{(0)} - a_0^{(0)} + \delta) a + (2p_h - p_1 - p_2)\delta.
\end{aligned}$$

Proof of Proposition 3.2

We proof the assertion by induction. For $t = 2$, observe that the upper bound $c \leq 2^{-\alpha}(p_1 - p_2)^2$ on the constant c implies $\text{Var}(p)_2 = \frac{(p_1 - p_2)^2}{2} \geq c 2^{\alpha - 1}$. Now let $t \geq 2$ and suppose that $\text{Var}(p)_t \geq c t^{\alpha - 1}$. If (3.3) has no solution, $\hat{a}_{0t} \leq 0, \hat{a}_{1t} \geq 0$, or $\hat{a}_{0t} + \hat{a}_{1t} p < 0$ for some $p \in [p_l, p_h]$, then $|p_{t+1} - \bar{p}_t| = \max(|p_1 - \bar{p}_t|, |p_2 - \bar{p}_t|) \geq \frac{|p_1 - p_2|}{2}$. Observe that for all $t \geq 2$ and $\alpha \in (0, 1)$, $(t+1)^\alpha - t^\alpha \leq \alpha t^{\alpha - 1}$ Together with the bound $c \leq 2^{-\alpha}(3\alpha)^{-1}(p_1 - p_2)^2$ this implies

$$(t+1)\text{Var}(p)_{t+1} = t\text{Var}(p)_t + \frac{t}{t+1}(p_{t+1} - \bar{p}_t)^2 \geq c t^\alpha + c\left[(t+1)^\alpha - t^\alpha\right] = c(t+1)^\alpha.$$

If (3.7) is chosen, then automatically $\text{Var}(p)_{t+1} \geq c(t+1)^{\alpha - 1}$.

If (3.8) is chosen, then by construction of the taboo interval (3.6) we have

$$(t+1)\text{Var}(p)_{t+1} = t\text{Var}(p)_t + \frac{t}{t+1}(p_{t+1} - \bar{p}_t)^2 \geq c t^\alpha + c\left[(t+1)^\alpha - t^\alpha\right] = c(t+1)^\alpha.$$

Proof of Lemma 3.1

From $\lambda_{\max}(t) + \lambda_{\min}(t) = \text{tr}(P_t) = t(1 + \overline{p^2}_t) > 0$ and $\lambda_{\max}(t)\lambda_{\min}(t) = \det(P_t) = t^2\text{Var}(p)_t > 0$, it follows that $\lambda_{\min}(t) > 0$. Together with $\overline{p^2}_t \leq p_h$ we thus have $\lambda_{\max}(t) \leq t(1 + \overline{p^2}_t) \leq t(1 + p_h^2)$. Furthermore, $\lambda_{\min}(t) = \lambda_{\max}(t)^{-1} \det(P_t) = \lambda_{\max}(t)^{-1} t^2 \text{Var}(p)_t \geq (1 + p_h^2)^{-1} t \text{Var}(p)_t$.

Proof of Theorem 3.1

Since $r(p, a)$ is twice continuously differentiable in p, it follows from a Taylor series expansion that, given a, for all $p \in [p_l, p_h]$ there is a $\tilde{p} \in [p_l, p_h]$ on the line segment between p and p_{opt}, such that

$$r(p, a) = r(p_{\text{opt}}, a) + r'(p_{\text{opt}}, a)(p - p_{\text{opt}}) + \frac{1}{2} r''(\tilde{p}, a)(p - p_{\text{opt}})^2,$$

where $r'(p, a)$ and $r''(p, a)$ denote the first and second derivatives of r with respect to p. The assumption $p_l < p_{\text{opt}} < p_h$ implies $r'(p_{\text{opt}}) = 0$, and with $K = \sup_{p \in [p_l, p_h]} |r''(p)| < \infty$ it follows that

$$|r(p) - r(p_{\text{opt}})| \leq \frac{K}{2}(p - p_{\text{opt}})^2, \qquad (p \in \mathcal{P}). \tag{3.13}$$

On an open neighborhood of p_{opt} in $\mathcal{P}$, $p_{\text{opt}} = p(a^{(0)})$ is the unique solution to $r'(p, a^{(0)}) = 0$. Since by assumption $r''(p, a_0, a_1)$ exists and is nonzero at the point $(p(a^{(0)}), a^{(0)})$, it follows from the implicit function theorem (see e.g. Duistermaat and Kolk, 2004) that there are open neighborhoods U of $p(a^{(0)})$ in $\mathbb{R}$ and V of $a^{(0)}$ in $\mathbb{R}^2$, such that for each $a \in V$ there is a unique $p \in U$ with $r'(p, a) = 0$. Moreover, the mapping $V \to U, a \mapsto p(a)$, is continuously differentiable in V. Consequently for all $a \in V$ there is a $\tilde{a} \in V$ on the line segment between a and $a^{(0)}$, such that

$$p(a) = p(a^{(0)}) + \left.\frac{\partial p(a)}{\partial a^T}\right|_{\tilde{a}} (a - a^{(0)}),$$

which implies that we can choose V such that for all $a \in V$,

$$|p(a) - p(a^{(0)})| = O\left(||a - a^{(0)}||\right). \tag{3.14}$$

Let $\rho \in (0, \rho_0)$ be such that $\{a : \left|\left|a - a^{(0)}\right|\right| \leq \rho\} \subset V$, and such $a_0 > 0$, $a_1 < 0$ whenever $\left|\left|a - a^{(0)}\right|\right| \leq \rho$. Let T_ρ be as in (3.9). Then for all $t \in \mathbb{N}$,

$$\begin{aligned} E[|p_t - p_{\text{opt}}|^2] &= E[|p_t - p_{\text{opt}}|^2 \cdot \mathbf{1}_{t > T_\rho}] + E[|p_t - p_{\text{opt}}|^2 \cdot \mathbf{1}_{t \leq T_\rho}] \\ &\leq E[|p_t - p_{\text{opt}}|^2 \cdot \mathbf{1}_{t > T_\rho}] + (p_h - p_l)^2 P(t \leq T_\rho) \\ &\leq E[|p_t - p_{\text{opt}}|^2 \cdot \mathbf{1}_{t > T_\rho}] + (p_h - p_l)^2 \frac{E[T_\rho^{1/2}]}{t^{1/2}}, \end{aligned} \tag{3.15}$$

where $\mathbf{1}_A$ denotes the indicator function of the event A.

Since $r'(p(a^{(0)}), a^{(0)}) = 0$ and $r''(p(a^{(0)}), a^{(0)}) < 0$, it follows from continuity arguments that $r'(p(a), a) = 0$ and $r''(p(a), a) < 0$ for all a in an open neighborhood of $a^{(0)}$. This implies that if $\left|\left|\hat{a}_t - a^{(0)}\right|\right|$ is sufficiently small and t sufficiently large, $\underset{p \in [p_l, p_h] \setminus TI(t)}{\arg\max} r(p, \hat{a}_t)$ lies on the boundary of the taboo interval $TI(t)$. It follows that there is a $\rho' \in (0, \rho_0)$ such that for all $t > T_{\rho'}$,

$$|p_t - p(\hat{a}_t)| \leq |TI(t)|, \tag{3.16}$$

where $|TI(t)|$ denotes the length of the taboo interval $TI(t)$. Combining (3.14), (3.15), and (3.16),

$$\begin{aligned}
E[|p_t - p_{\text{opt}}|^2] &\leq E[|p_t - p_{\text{opt}}|^2 \cdot \mathbf{1}_{t>T_{\rho'}}] + (p_h - p_l)^2 \frac{E[T_{\rho'}^{1/2}]}{t^{1/2}}, \\
&\leq 2E[|p_t - p(\hat{a}_t)|^2 \cdot \mathbf{1}_{t>T_{\rho'}}] + 2E[|p(\hat{a}_t) - p(a^{(0)})|^2 \cdot \mathbf{1}_{t>T_{\rho'}}] + (p_h - p_l)^2 \frac{E[T_{\rho'}^{1/2}]}{t^{1/2}}, \\
&= O\left(E[|TI(t)|^2] + E\left[\left|\left| \hat{a}_t - a^{(0)} \right|\right|^2 \cdot \mathbf{1}_{t>T_{\rho'}} \right] + (p_h - p_l)^2 \frac{E[T_{\rho'}^{1/2}]}{t^{1/2}} \right), \\
&= O\left(t^{\alpha-1} + \frac{\log(t)}{t^{\alpha}} \right),
\end{aligned}$$

by Proposition 3.3, from which follows

$$\begin{aligned}
\text{Regret}(T) = \sum_{t=1}^{T} E[r(p_{\text{opt}}) - r(p_t)] &= O\left(\sum_{t=1}^{T} E[(p_t - p_{\text{opt}})^2] \right) \\
&= O\left(\sum_{t=1}^{T} t^{\alpha-1} + t^{-\alpha}\log(t) \right) \\
&= O\left(T^{\alpha} + T^{1-\alpha}\log(T) \right).
\end{aligned}$$

Chapter 4

Dynamic pricing and learning for multiple products with infinite inventory

4.1 Introduction

In the preceding chapter, we study pricing policies that are suitable for a firm selling a single type of product. In practice, firms often sell multiple types of products, and the demand for one product is influenced by the selling prices of the other products. This means that learning the demand function and determining optimal prices have to be considered for all products simultaneously; one can, in general, not simply apply the methods from Chapter 3 independently for each individual product. We therefore study in the current chapter dynamic pricing and learning in a setting with multiple products.

Similar as in the single-product case, we consider a parametric setting where the seller knows the relation between selling prices and the first two moments of the demand distributions, up to some unknown parameters. The value of these unknown parameters can be estimated by maximum quasi-likelihood estimation (MQLE); this is an extension of classical maximum-likelihood estimation to settings where only the first two moments of the distribution are known.

We propose an adaptive pricing policy which is based on the following principle: in each time period, the seller estimates the unknown parameters with MQLE; subsequently, he chooses the prices that generate the highest expected revenue, given that these parameter estimates are correct, and with an additional requirement on a certain measure of price dispersion. This policy balances at each time step exploration and exploitation: the requirement on the price dispersion makes sure that the parameter estimates converge to the true values, the current knowledge of the parameter estimates is exploited by choosing the optimal prices w.r.t. these estimates.

We measure price dispersion by the smallest eigenvalue of the design matrix, which is specified below, and require that it grows with a certain pre-specified rate. This rate guarantees strong consistency of the MQL estimates. There is no simple recursive relation between these smallest eigenvalues in two consecutive time periods. We therefore work with an expression which grows at the same rate, namely the inverse of the trace of the inverse design matrix. Using the Sherman-Morrison formula, we show that a simple quadratic constraint on the chosen prices is sufficient to establish the desired growth rate of the smallest eigenvalue of the design matrix.

The performance of pricing policies is measured in terms of Regret(T), which is the expected amount of revenue loss after T time periods, caused by not using the optimal price. We provide two conditions - one assuring a sufficient amount of price dispersion, the other bounding the cu-

mulative deviation from the certainty equivalence prices - such that any pricing policy satisfying these conditions admits an upper bound on the regret in terms of the amount of price dispersion. We show that our proposed adaptive pricing policy satisfies these conditions, and by optimally choosing the price dispersion rate, we obtain the bound Regret$(T) = O(T^{2/3})$.

In many demand models that are used in practice, the demand functions are so-called *canonical* link functions. For this important class of demand functions, we show that Regret$(T) = O(\sqrt{T \log(T)})$ can be achieved. This bound is close to $O(\sqrt{T})$, which in several (single product) settings has been shown to be the lowest provable asymptotic upper bound on the regret (see e.g. Kleinberg and Leighton, 2003, Besbes and Zeevi, 2011, Broder and Rusmevichientong, 2012). The upper bound Regret$(T) = O(\sqrt{T \log(T)})$ is based on new sufficient conditions that guarantee strong consistency of MQLE. The proof of this result is based on an extension of a theorem by Lai (1974) to martingale difference sequences, which may be of independent interest.

One of the strengths of our approach to dynamic pricing and learning for multiple products, is that our results are valid for a very large class of demand functions and distributions. Other works, such as Le Guen (2008) or Harrison et al. (2011b), restrict to linear demand functions or sub-Gaussian demand distributions. In addition, we construct a pricing policy that facilitates learning of the unknown parameters; in contrast, in a robust approach as Lim et al. (2008), no learning takes place.

The remainder of this chapter is organized as follows. Section 4.2 introduces the model and notation, discusses some of the assumptions we make, and introduces the maximum quasi-likelihood estimator. Section 4.3 describes the proposed adaptive pricing policy. In Section 4.4.1 we provide an upper bound on the regret of a pricing policy, in terms of the amount of price dispersion. Section 4.4.2 shows that in case of canonical link functions these bounds can be improved. Some auxiliary results needed to prove these regret bounds are contained in Section 4.4.3, and the quality of the regret bounds is discussed in Section 4.4.4. Numerical illustrations are provided in Section 4.5, and all mathematical proofs are contained in Section 4.6.

4.2 Model, assumptions, and estimation method

4.2.1 Model and notation

In this section, we consecutively discuss the dynamic pricing setting under consideration, the parametric demand model deployed by the seller, assumptions on the revenue function, the definition of a policy, and the definition of the regret. Subsequently we explain some notation used in this paper.

We consider a firm that sells $n \in \mathbb{N}$ different types of products. Time is discretized, and time periods are denoted by $t \in \mathbb{N}$. A time period can represent a day or a week, but also say five minutes. At the beginning of each time period $t \in \mathbb{N}$, the firm determines for each product $k = 1, \ldots, n$ a selling price $p_k(t) > 0$. After setting the prices the firm observes a realization of the demand d_{kt} for each product $k = 1, \ldots, n$, and collects revenue $\sum_{k=1}^{n} p_k(t) d_{kt}$. We assume that all demand can be met, thus stock-outs do not occur.

Write $\mathbf{p}(t) = (p_0(t), p_1(t), \ldots, p_n(t))^T$, where $p_0(t) = 1$ for all t, and $p_k(t)$ is the selling price of product k in period t, $(1 \leq k \leq n)$. The term $p_0(t) = 1$ is convenient for notational reasons. We assume that the prices lie in a compact, convex, non-empty set $\mathcal{P} \subset \{1\} \times \mathbb{R}^n_{>0}$. The set $\mathcal{P}$ is called the set of admissible prices. A common choice is $\mathcal{P} = \{1\} \times \prod_{k=1}^{n} [p_{lk}, p_{hk}]$ where $0 < p_{lk} < p_{hk}$

denote the lowest and highest price for product k that is acceptable to the firm. Our assumptions on $\mathcal{P}$ are more flexible, allowing joint price constraints e.g. of the form $p_1 \leq p_2$.

The random variable $D_{kt}(\mathbf{p}(t))$ denotes the demand for product k in period t, given selling price vector $\mathbf{p}(t)$. Given the selling prices, the demand in different time periods and for different products are independent of each other, and for each $t \in \mathbb{N}$, $k = 1, \ldots, n$ and $\mathbf{p}(t) \in \mathcal{P}$, the demand d_{kt} is a realization of a random variable $D_k(\mathbf{p}(t))$. The seller assumes the following parametric model:

$$E[D_k(\mathbf{p})] = h_k(\mathbf{p}^T \beta_k^{(0)}), \qquad (\mathbf{p} \in \mathcal{P}), \tag{4.1}$$

$$\text{Var}[D_k(\mathbf{p})] = \sigma_k^2 v_k(E[D_k(\mathbf{p})]), \qquad (\mathbf{p} \in \mathcal{P}). \tag{4.2}$$

Here for all $k = 1, \ldots, n$, the functions $h_k : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and $v_k : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ are both thrice continuously differentiable, with $\dot{h}_k(x) = \frac{\partial h_k(x)}{\partial x} > 0$, $v_k(x) > 0$ for all $x \geq 0$, σ_k^2 are unknown positive scalars, and $\beta_k^{(0)} = (\beta_{k0}^{(0)}, \ldots, \beta_{kn}^{(0)})^T \in \mathbb{R}^{n+1}$ are unknown parameter vectors. The functions h_k are called *link functions*. With $\beta^{(0)}$ we denote the $n \times (n+1)$ matrix whose k-th row equals $(\beta_{k0}^{(0)}, \ldots, \beta_{kn}^{(0)})$.

Let $(\mathcal{F}_t)_{t \in \mathbb{N}}$ be the filtration generated by $\{d_{ki}, p_{ki} : k = 1, \ldots, n, i = 1, \ldots, t\}$, i.e. by all prices and demand realizations up to and including time t, for $t \in \mathbb{N}$, and let $\mathcal{F}_0$ be the trivial σ-algebra. A technical assumption on the demand is

$$\sup_{\mathbf{p} \in \mathcal{P}, k = 1, \ldots, n} E[|D_k(\mathbf{p}) - E[D_k(\mathbf{p}) \mid \mathcal{F}_{t-1}]|^\gamma] < \infty \text{ a.s., for some } \gamma > 3. \tag{4.3}$$

The expected revenue collected in a single time period by product k against price $\mathbf{p}$, is denoted by $r_k(\mathbf{p}) = E[p_k D_k(\mathbf{p})] = p_k h(\mathbf{p}^T \beta_k^{(0)})$. The total expected revenue in a single time period t against selling price $\mathbf{p}$ is $r(\mathbf{p}) = \sum_{k=1}^n r_k(\mathbf{p})$. We also write $r_k(\mathbf{p}, \beta_k)$ and $r(\mathbf{p}, \beta)$ as a function of both the price vector $\mathbf{p}$ and the parameter values $\beta_k \in \mathbb{R}^{n+1}$ resp. $\beta \in \mathbb{R}^{(n+1) \times n}$.

We assume there is an open, bounded neighborhood $V \in \mathbb{R}^{n \times (n+1)}$ around $\beta^{(0)}$, such that for all $\beta \in V$, the function $\mathcal{P} \to \mathbb{R}, \mathbf{p} \mapsto r(\mathbf{p}, \beta)$ has a unique maximizer

$$\mathbf{p}(\beta) = \arg\max_{\mathbf{p} \in \mathcal{P}} r(\mathbf{p}, \beta) \in \text{int}(\mathcal{P}), \tag{4.4}$$

such that the matrix of all second derivatives of r w.r.t. $\mathbf{p}$ (excluding the first component $\mathbf{p}_0 = 1$),

$$H(\mathbf{p}, \beta) = \left(\frac{\partial^2 r(\mathbf{p}, \beta)}{\partial p_i \partial p_j} \right)_{1 \leq i,j \leq n}, \tag{4.5}$$

is negative definite at the point $\mathbf{p}(\beta)$. In (4.4), and throughout this chapter, int$(\mathcal{P})$ is defined as $\{1\} \times \text{int}(\{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid (1, p_1, \ldots, p_n) \in \mathcal{P}\})$. The correct optimal price $\mathbf{p}(\beta^{(0)})$ is also denoted by $\mathbf{p}_{\text{opt}}$.

A pricing policy ψ is a method that for each $t \in \mathbb{N}$ generates a price $\mathbf{p}(t) \in \mathcal{P}$. This price may depend on the previously chosen prices $\mathbf{p}(1), \ldots, \mathbf{p}(t-1)$ and demand realizations $\{d_{ki} : k = 1, \ldots, n, i = 1, \ldots, t-1\}$, i.e. $\mathbf{p}(t)$ is $\mathcal{F}_{t-1}$-measurable.

The performance of a pricing policy is measured by the regret, which is the expected revenue loss caused by not using the optimal price $\mathbf{p}_{\text{opt}}$. For a pricing policy ψ that generates prices

$\mathbf{p}(1), \mathbf{p}(2), \ldots, \mathbf{p}(T)$, the regret after T time periods is defined as

$$\text{Regret}(T, \psi) = E\left[\sum_{t=1}^{T} r(\mathbf{p}_{\text{opt}}, \beta^{(0)}) - r(\mathbf{p}(t), \beta^{(0)})\right].$$

The objective of the seller is to find a pricing policy ψ that gives the highest expected revenue over T time periods. This is equivalent to minimizing Regret(T, ψ). Note that this objective cannot directly be used by the seller to find a policy, since it depends on the unknown parameters $\beta^{(0)}$.

Notation. With $\text{tr}(A)$ and $\det(A)$ we denote the trace and determinant of a matrix A, with $\lambda_{\text{max}}(A)$ and $\lambda_{\text{min}}(A)$ its largest and smallest eigenvalue (when these are real-valued). The transpose of a (column) vector v is denoted by v^T. Given price vectors $\mathbf{p}(1), \ldots, \mathbf{p}(t)$, the design matrix $P(t)$ is defined as

$$P(t) = \sum_{i=1}^{t} \mathbf{p}(i)\mathbf{p}^T(i). \tag{4.6}$$

Since the largest and smallest eigenvalue of $P(t)$ play an important role in the analysis, we use shorthand notation $\lambda_{\text{max}}(t) = \lambda_{\text{max}}(P(t))$ and $\lambda_{\text{min}}(t) = \lambda_{\text{min}}(P(t))$. The natural logarithm of $x > 0$ is denoted by $\log(x)$. If it is clear from the context which pricing policy ψ is used, we sometimes write Regret(T) instead of Regret(T, ψ).

4.2.2 Discussion of model assumptions

We only assume knowledge on the first two moments of the demand, not on the complete distribution. This makes the demand model a little more robust. The assumption that the variance is a function of the first moment is valid for several demand distributions that are commonly used in practice, for example, if the distribution of $D_k(\mathbf{p})$ is normal ($v_k(h) = 1$), Bernoulli ($v_k(h) = h(1-h)$), or Poisson ($v_k(h) = h$). The moment assumption (4.3) is not common in the literature on dynamic pricing, and allows for heavy-tailed demand distributions. The conditions on the uniqueness of the optimal price $\mathbf{p}(\beta)$ and on the Hessian matrix (4.5) are satisfied when the revenue function $r(\mathbf{p}, \beta^{(0)})$ is strictly concave in $\mathbf{p}$. This is for example the case if the demand functions are linear ($h_k(x) = x$ for each $k = 1, \ldots, n$) and the matrix $\left(\beta_{kl}^{(0)} + \beta_{lk}^{(0)}\right)_{k,l=1,\ldots,n}$ is negative definite.

4.2.3 Estimation of unknown parameters

The unknown parameters $\beta^{(0)}$ can be estimated with maximum quasi-likelihood estimation. This is a natural extension of ordinary maximum-likelihood estimation to settings where only the first two moments of the distribution are known. For more details we refer to Wedderburn (1974), McCullagh (1983), Godambe and Heyde (1987), McCullagh and Nelder (1983), Heyde (1997) and Gill (2001).

Given price vectors $\mathbf{p}(1), \ldots, \mathbf{p}(t)$ and demand realizations $\{d_{ki} \mid k = 1, \ldots, n, i = 1, \ldots, t\}$, the maximum quasi-likelihood estimate (MQLE) of $\beta_k^{(0)}$, denoted by $\hat{\beta}_k(t) \in \mathbb{R}^{n+1}$, is defined as a solution to the $(n+1)$-dimensional equation

$$l_{kt}(\beta_k) = \sum_{i=1}^{t} \frac{\dot{h}_k(\mathbf{p}^T(i)\beta_k)}{\sigma_k^2 v_k(h_k(\mathbf{p}^T(i)\beta_k))}\mathbf{p}(i)(d_{ki} - h_k(\mathbf{p}^T(i)\beta_k) = 0. \tag{4.7}$$

4.3 Adaptive pricing policy

A natural and intuitive pricing policy is to set at each time period the selling prices equal to the prices that are optimal, given that the current parameter estimates are correct. This pricing policy is usually called myopic pricing or certainty equivalent pricing. At each step, the firm acts as if it is certain about its parameter estimates. Although this policy is very intuitive and easy to understand, its performance is very poor: in Chapter 3, we show for a single product with normally distributed demand function whose expectation depends linearly on the selling price, that with certainty equivalent pricing, the parameter estimates may converge to the wrong value, and the price may converge to a limit price which is not equal to the optimal price. There we propose an alternative pricing policy, called Controlled Variance Pricing, and show that under this policy the price converges to the optimal price. The key idea of this policy is to use at each time period the optimal price given the current parameter estimates, with an additional constraint on the price dispersion. In this single product case, the price dispersion at time t is measured by the sample variance of the prices chosen up to time t, and is required to satisfy a carefully chosen, time-dependent lower bound. This pricing rule balances at each time step learning of the parameters and instant revenue optimization, i.e. exploration and exploitation.

We now introduce an adaptive pricing policy for multiple products, which is inspired by the same principles as Controlled Variance Pricing. The key idea is to choose the optimal price given the current parameter estimates, with the additional requirement that $\lambda_{\min}(t)$, the smallest eigenvalue of the design matrix (4.6), grows with a certain rate. More precisely we require that $\lambda_{\min}(t) \geq L_1(t)$, where $L_1(t)$ is a positive monotone increasing non-random function on $\mathbb{N}$. As we will show, a proper choice of the function $L_1(t)$ implies sufficient price dispersion for the parameter estimates $\hat{\beta}(t)$ to converge to the true value $\beta^{(0)}$.

Since there is no simple explicit expression relating two consecutive smallest eigenvalues $\lambda_{\min}(t)$ and $\lambda_{\min}(t+1)$, we instead work with the trace of the inverse design matrix, $\mathrm{tr}(P(t)^{-1})$. This can be justified by the fact that for any positive definite $n \times n$ matrix A,

$$\mathrm{tr}(A^{-1})^{-1} \leq \lambda_{\min}(A) \leq n\mathrm{tr}(A^{-1})^{-1}. \tag{4.8}$$

Thus, $\mathrm{tr}(P(t)^{-1}) = O\left(L_1(t)^{-1}\right)$ is equivalent to $\lambda_{\min}(P(t)) = \Omega(L_1(t))$. The expression $\mathrm{tr}(P(t)^{-1})$ admits a recursive form via the Sherman-Morrison formula (Bartlett (1951), see Hager (1989) for a historical treatment of these type of formulas). In particular, one can show

$$\mathrm{tr}(P(t+1)^{-1}) - \mathrm{tr}(P(t)^{-1}) = -\frac{\left|\left|P(t)^{-1}\mathbf{p}(t+1)\right|\right|^2}{1 + \mathbf{p}^T(t+1)P(t)^{-1}\mathbf{p}(t+1)}. \tag{4.9}$$

If $\mathrm{tr}(P(t)^{-1}) \leq \frac{1}{L_1(t)}$ and $\mathbf{p}(t+1)$ is chosen such that the right hand side of (4.9) satisfies a carefully chosen constraint, we can make sure that $\mathrm{tr}(P(t+1)^{-1}) \leq \frac{1}{L_1(t+1)}$.

Let $\mathcal{L}$ be the class of non-decreasing differentiable functions $L : \mathbb{N} \to \mathbb{R}_{>0}$ such that $\dot{L}(t) = o(1)$, and $t \mapsto \frac{1}{L(t)}$ is convex. Examples of functions contained in $\mathcal{L}$ are $t \mapsto c\sqrt{t\log(t)}$ or $t \mapsto ct^a$, $(c > 0, 0 < a < 1)$. It is not difficult to derive that for any $L \in \mathcal{L}$, $L(t) = o(t)$, and there exists a $C_L \in \mathbb{N}$ such that $L_1(C_L t) \leq C_L L_1(t)$ for all $t \in \mathbb{N}$.

The details of the adaptive pricing policy, named Φ_A, are outlined below:

Adaptive pricing policy Φ_A for n products

Initialization: Choose $L_1 \in \mathcal{L}$.
Choose $n+1$ linearly independent initial price vectors $\mathbf{p}(1), \mathbf{p}(2), \ldots, \mathbf{p}(n+1)$ in $\mathcal{P}$.
For all $t \geq n+2$:
Estimation: For each $k = 1, \ldots, n$, calculate the MQLE $\hat{\beta}_k(t)$ using the MQL equations (4.7).
Pricing:
I) If for some k, $\hat{\beta}_k(t)$ does not exist, or $\text{tr}(P(t)^{-1})^{-1} \not\geq L_1(t)$, then set $\mathbf{p}(t+1) = \mathbf{p}(1)$, $\mathbf{p}(t+2) = \mathbf{p}(2)$, ..., $\mathbf{p}(t+j) = \mathbf{p}(j)$, where j is the smallest integer such that $\text{tr}(P(t+j)^{-1})^{-1} \geq L_1(t+j)$.
II) If for all k, $\hat{\beta}_k(t)$ exists, and $\text{tr}(P(t)^{-1})^{-1} \geq L_1(t)$, let $\mathbf{p}_{\text{ceqp}} = \mathbf{p}(\hat{\beta}(t))$, and consider the following cases:
IIa) If

$$\text{tr}\left(\left(P(t) + \mathbf{p}_{\text{ceqp}}\mathbf{p}_{\text{ceqp}}^T\right)^{-1}\right)^{-1} \geq L_1(t+1), \tag{4.10}$$

then choose $\mathbf{p}(t+1) = \mathbf{p}_{\text{ceqp}}$.
IIb) If (4.10) does not hold, then choose $\mathbf{p}(t+1)$ that maximizes

$$\max_{\mathbf{p}\in\mathcal{P}} r(\mathbf{p}, \hat{\beta}(t)) \text{ s.t. } \frac{\left|\left|P(t)^{-1}\mathbf{p}\right|\right|^2}{1 + \mathbf{p}^T P(t)^{-1}\mathbf{p}} \geq \frac{\dot{L}_1(t)}{L_1(t)^2}, \tag{4.11}$$

provided there is a feasible solution.
IIc) If (4.10) does not hold, and (4.11) has no feasible solution, then set $\mathbf{p}(t+1) = \mathbf{p}(1)$, $\mathbf{p}(t+2) = \mathbf{p}(2)$, ..., $\mathbf{p}(t+j) = \mathbf{p}(j)$, where j is the smallest integer such that $\left|\left|P(t+j)^{-1}\mathbf{p}\right|\right|^2 \left[1 + \mathbf{p}^T P(t+j)^{-1}\mathbf{p}\right]^{-1} \geq \dot{L}_1(t+j)L_1(t+j)^{-2}$ is satisfied by some $\mathbf{p} \in \mathcal{P}$.

Ad I) and IIc) in the policy description deal with possible non-existence of the MQLE $\hat{\beta}_k(t)$ and other short-timescale effects: in that case, all previously chosen prices are repeated until the MQLE exists and there is sufficient price dispersion. In the proof of Proposition 4.2 we show that the term "j" in I) and IIc) is always finite.

Ad IIa) describes the situation where the certainty equivalent price $\mathbf{p}(\hat{\beta}(t))$ induces sufficient price dispersion; in that case, the next price is equal to the certainty equivalent price.

Ad IIb) shows which price to choose when the certainty equivalent price induces insufficient price dispersion. In that case, an additional constraint in (4.11) has to be satisfied.

Computational methods to solve the MQL equations (4.7) are discussed in Osborne (1992) and Heyde and Morton (1996). The value of $\mathbf{p}_{\text{ceqp}} = \mathbf{p}(\hat{\beta}(t))$ is the solution in $\mathcal{P}$ of a maximization problem in n variables. The maximization problem (4.11) has one additional constraint, which can be written as $g(\mathbf{p}) \geq 0$, where g is a quadratic polynomial in $\mathbf{p}$. Both these problems can be solved using standard techniques for constrained optimization (see e.g. Bertsekas, 1982, 1999, Fletcher, 2000).

For sufficiently large t, the maximization problem (4.11) always has a feasible solution:

Proposition 4.1 (Feasibility of (4.11))**.** *There is a $T_0 \in \mathbb{N}$, depending only on $\mathcal{P}$ and L_1, such that for all $t \geq T_0$: if*

$$tr\left(P(t)^{-1}\right)^{-1} \geq L_1(t), \tag{4.12}$$

$$tr\left(\left(P(t) + \mathbf{p}(\hat{\beta}(t))\mathbf{p}(\hat{\beta}(t))^T\right)^{-1}\right)^{-1} < L_1(t+1), \tag{4.13}$$

then the set

$$\left\{ \mathbf{p} \in \mathcal{P} \mid \frac{\left|\left|P(t)^{-1}\mathbf{p}\right|\right|^2}{1+\mathbf{p}^T P(t)^{-1}\mathbf{p}} \geq \frac{\dot{L}_1(t)}{L_1(t)^2} \right\}$$

is nonempty.

The following proposition states that for sufficiently large t, the adaptive pricing policy Φ_A induces a lower bound on $\text{tr}(P(t)^{-1})^{-1}$, and thus by (4.8) also on $\lambda_{\min}(t)$.

Proposition 4.2 (Growth rate of $\text{tr}(P(t)^{-1})^{-1}$)**.** *There are T_1, $C_L \in \mathbb{N}$, depending only on T_0, L_1, and $P(n+1)$, such that for all $t \geq T_1$,*

$$tr(P(t)^{-1})^{-1} \geq C_L^{-1} L_1(t). \tag{4.14}$$

4.4 Bounds on the regret

In this section, we provide upper bounds on the regret induced by pricing policies. The bounds depends on two characteristics of a pricing policy: the first is a lower bound L_1 on the smallest eigenvalue $\lambda_{\min}(t)$ of the design matrix $P(t)$; this bound quantifies the amount of emphasis on learning the unknown parameters. The second characteristic is the cumulative difference between the chosen prices and the certainty equivalence prices. Theorem 4.1 in Section 4.4.1 states an upper bound on the regret, in terms of these two characteristics. We apply this result on the pricing policy introduced in Section 4.3, and show that it can achieve $\text{Regret}(T) = O(T^{2/3})$.

In Section 4.4.2, we consider the case of canonical link functions h_k, which means that a certain relation between the functions h_k and v_k is satisfied (the details are in Section 4.4.2). Canonical link functions are encountered in several demand models of practical interest. We extend existing statistical results on the strong consistency of MQLE, and show that $\text{Regret}(T) = O(\sqrt{T \log(T)}$ can be achieved. As intermediate result we obtain an extension of Theorem 3 of Lai (1974) to martingale difference sequences.

Section 4.4.4 discusses the quality of the upper bounds derived in Sections 4.4.1 and 4.4.2.

4.4.1 General link functions

In order to state the main results of this section, we develop some notation that deals with possible non-existence of solutions to the quasi-likelihood equations. In particular, for $\rho > 0$ and $k = 1, \ldots, n$, we define

$$T_{\rho,k} = \sup\{n \in \mathbb{N} : \text{there is no } \beta \in B_{\rho,k} \text{ such that } l_{kt}(\beta) = 0\}, \tag{4.15}$$

where $B_{\rho,k} = \left\{ \beta \in \mathbb{R}^{n+1} \mid \left|\left| \beta - \beta_k^{(0)} \right|\right| \leq \rho \right\}$, and

$$T_\rho = \max\{T_{\rho,1}, \ldots, T_{\rho,n}\}. \tag{4.16}$$

The importance of T_ρ becomes clear from following proposition, which relates L_1 to the rate at which the parameter estimate $\hat{\beta}(t)$ converges to the true value $\beta^{(0)}$, and in addition provides moment bounds on T_ρ.

Proposition 4.3 (Strong consistency and convergence rates)**.** *Let $L_1 \in \mathcal{L}$, and suppose there are $t_0 \in \mathbb{N}$, $c > 0$ and $\alpha \in (\frac{1}{2}, 1)$ such that $\lambda_{min}(t) \geq L_1(t) \geq ct^\alpha$ a.s. for all $t \geq t_0$. Then there is a $\rho_0 > 0$*

such that $T_\rho < \infty$ *a.s. and* $E\left[T_\rho^\eta\right] < \infty$*, for all* $0 < \eta < \gamma\alpha - 1$ *and* $0 < \rho \leq \rho_0$*. In addition, for all* $k = 1, \ldots, n$ *and* $t > T_\rho$*, there exists a solution* $\hat{\beta}_k(t)$ *to* (4.7)*,* $\lim_{t\to\infty} \hat{\beta}_k(t) = \beta_k^{(0)}$ *a.s., and*

$$E\left[\left|\left|\hat{\beta}_k(t) - \beta_k^{(0)}\right|\right|^2 \mathbf{1}_{t>T_\rho}\right] = O\left(L_1(t)^{-1}\log(t) + tL_1(t)^{-2}\right).$$

The assertions about T_ρ follow from applying Theorem 7.1 (Chapter 7) for each $T_{\rho,k}$, $k = 1, \ldots, n$, and noting that $T_\rho \leq \sum_{k=1}^n T_{\rho,k}$ a.s. The other statements follow from Theorem 7.2.

The following theorem provides an upper bound on the regret, in terms of the function L_1. Let $\rho_1 \in (0, \rho_0)$ such that $\{(\beta_1, \ldots, \beta_n) \in \mathbb{R}^{n\times(n+1)} \mid \beta_k \in B_{\rho,k}, k = 1, \ldots, n\} \subset V$, where V is defined in Section 4.2.

Theorem 4.1. *Let* $t_0 \in \mathbb{N}$*,* $L_1 \in \mathcal{L}$ *such that* $L_1(t) \geq ct^\alpha$ *for all* $t \geq t_0$ *and some* $c > 0$*,* $\alpha \in (\frac{1}{2}, 1)$*. Let* $0 < \rho \leq \rho_1$*, and let* T_2 *be a random variable on* $\mathbb{N}$ *such that* $T_2 \geq T_\rho$ *a.s. and* $E[T_2^{1/2}] < \infty$*. If* ψ *is a pricing policy that satisfies*

(i) $\lambda_{min}(t) \geq L_1(t)$ *a.s., for all* $t \geq t_0$*,*

(ii) $\sum_{t=1}^T \left|\left|\mathbf{p}(t) - \mathbf{p}(\hat{\beta}(t-1))\right|\right|^2 \mathbf{1}_{t>T_2} \leq K_2 L_1(T)$ *a.s., for all* $T \geq t_0$ *and some* $K_2 > 0$*,*

then $\text{Regret}(\psi, T) = O\left(L_1(T) + \sum_{t=1}^T \left(\frac{\log(t)}{L_1(t)} + \frac{t}{L_1(t)^2}\right)\right)$.

Note that if $L_1(t) = O(t/\log(t))$, the bound in Theorem 4.1 equals

$$\text{Regret}(\psi, T) = O\left(L_1(T) + \sum_{t=1}^T \frac{t}{L_1(t)^2}\right). \tag{4.17}$$

In the formulation of the theorem, one naturally can set T_2 just equal to T_ρ. With the current formulation we allow for a little more flexibility, e.g., we allow to set $T_2 = \max\{T_\rho, T\}$, for a non-random $T \in \mathbb{N}$. Furthermore, note that the theorem bears resemblance with Theorem 6 of Harrison et al. (2011b). A difference is that they restrict to a linear demand function, Gaussian distributed noise terms, and a special class of policies called "orthogonal policies".

In Theorem 4.1, the choice $L_1(t) = ct^{2/3}$, for some $c > 0$, yields $\text{Regret}(\psi, T) = O(T^{2/3})$. This choice is optimal in the sense that for this choice of L_1,

$$L_1(T) + \sum_{t=1}^T \left(\frac{\log(t)}{L_1(t)} + \frac{t}{L_1(t)^2}\right) = o\left(\tilde{L}_1(T) + \sum_{t=1}^T \left(\frac{\log(t)}{\tilde{L}_1(t)} + \frac{t}{\tilde{L}_1(t)^2}\right)\right),$$

for all $\tilde{L}_1 \in \mathcal{L}$ such that $L_1 = o(\tilde{L}_1)$ or $\tilde{L}_1 = o(L_1)$.

The following theorem shows that the adaptive pricing policy Φ_A from Section 4.3 satisfies the conditions of Theorem 4.1.

Theorem 4.2. *Consider the adaptive pricing policy* Φ_A*, with* $L_1(t) \geq ct^\alpha$ *for all* $t \geq t_0$ *and some* $c > 0$*,* $\alpha \in (\frac{1}{2}, 1)$*. There exist* $K_2 \in \mathbb{N}$*,* $\rho \in (0, \rho_1]$ *and a random variable* T_2*, that satisfy the conditions of Theorem 4.1.*

4.4.2 Canonical link functions

The terms $\sum_{t=1}^{T}\left(\frac{\log(t)}{L_1(t)}+\frac{t}{L_1(t)^2}\right)$ in the regret bound of Theorem 4.1, come from the convergence rates in Proposition 4.3. These rates are valid for general functions h_k and v_k. However, in several special cases, which are of particular practical interest, Theorem 7.3 shows that Proposition 4.3 can be improved to

$$E\left[\left|\left|\hat{\beta}_k(t)-\beta_k^{(0)}\right|\right|^2\mathbf{1}_{t>T_\rho}\right]=O\left(L_1(t)^{-1}\log(t)\right). \tag{4.18}$$

This is the case when the functions h_k are *canonical*, which means $\dot{h}_k(x)=v_k(h_k(x))$, for all $x\in\mathbb{R}$, $k=1,\ldots,n$. Some examples where this occurs, are normally distributed demand with $h_k(x)=x$, Poisson distributed demand with $h_k(x)=\exp(x)$, and Bernoulli distributed demand with $h_k(x)=\frac{\exp(x)}{1+\exp(x)}$.

It is easy to see that, by slightly altering the proof of Theorem 4.1, these improved bounds (4.18) for canonical link functions imply

$$\text{Regret}(\psi,T)=O\left(L_1(T)+\sum_{t=1}^{T}L_1(t)^{-1}\log(t)\right), \tag{4.19}$$

assuming exactly the same conditions as in Theorem 4.1. The choice $L_1(t)=ct^{\frac{1}{2}+\delta}$, for some $c>0$ and small $\delta>0$, then implies $\text{Regret}(\psi,T)=O(T^{\frac{1}{2}+\delta})$, which is a substantial improvement to the rate $T^{2/3}$ derived in Section 4.4.1.

However, one can show that the optimal choice that minimizes the right hand side of (4.19), is $L_1(t)=c\sqrt{t\log(t)}$, $(c>0)$. This choice is optimal in the following sense: if $L_1(t)=c\sqrt{t\log(t)}$ and $\tilde{L}_1\in\mathcal{L}$ is such that $L_1=o(\tilde{L}_1)$ or $\tilde{L}_1=o(L_1)$, then

$$L_1(T)+\sum_{t=1}^{T}L_1(t)^{-1}\log(t)=o\left(\tilde{L}_1(T)+\sum_{t=1}^{T}\tilde{L}_1(t)^{-1}\log(t)\right).$$

The choice $L_1(t)=c\sqrt{t\log(t)}$ does not satisfy the requirement in Proposition 4.3 that L_1 should grow at least as t^α, for some $\alpha\in(\frac{1}{2},1)$. This raises the question whether this requirement can be weakened. We show that this is indeed the case; in particular, we show that Proposition 4.3 is still valid if $L_1(t)\geq c\sqrt{t\log(t)}$, for a sufficiently large $c>0$. It then immediately follows that in Theorem 4.1, the choice $L_1(t)=c\sqrt{t\log(t)}$ with sufficiently large c, leads to $\text{Regret}(T)=O(\sqrt{T\log(T)})$, when the link functions are canonical.

Proposition 4.4 (Strong consistency and convergence rates)**.** *Suppose there are $t_0\in\mathbb{N}$, $c>0$ such that $L_1(t)\geq c\sqrt{t\log(t)})$ a.s. for all $t\geq t_0$. Then, for all sufficiently small $\rho>0$ there exists a $c_\rho^*>0$, such that for all $0<\eta<\frac{\gamma-1}{2}$: $T_\rho<\infty$ a.s. and $E\left[T_\rho^\eta\right]<\infty$, provided $c>c_\rho^*$. In addition, for all $k=1,\ldots,n$ and $t>T_\rho$, there exists a solution $\hat{\beta}_k(t)$ to* (4.7)*, $\lim_{t\to\infty}\hat{\beta}_k(t)=\beta_k^{(0)}$ a.s., and*

$$E\left[\left|\left|\hat{\beta}_k(t)-\beta_k^{(0)}\right|\right|^2\mathbf{1}_{t>T_\rho}\right]=O\left(L_1(t)^{-1}\log(t)\right).$$

The proof is based on Theorems 7.1 and 7.3, and on Proposition 4.5 contained in the next section.

4.4.3 Auxiliary results

This section contains a number of auxiliary results that are needed to prove the regret bounds in Section 4.4.1 and 4.4.2.

Proposition 4.5. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. Write $S_n = \sum_{i=1}^n X_i$ and suppose $\sup_{i\in\mathbb{N}} E[X_i^2 \mid \mathcal{F}_{i-1}] \leq \sigma^2 < \infty$ a.s. for some $\sigma > 0$. Let $\eta > 0$, $r > 2(\eta+1)$, $c > 2\sigma\sqrt{\eta}$, and define the random variable $T = \sup\{n \in \mathbb{N} \mid |S_n| \geq c\sqrt{n\log(n)}\}$, where T takes values in $\mathbb{N}\cup\{\infty\}$. If $\sup_{i\in\mathbb{N}} E\left[|X_i|^r\right] \leq C < \infty$ for some $C > 0$, then*

$$T < \infty \text{ a.s.}, \quad \text{and } E\left[T^\eta\right] < \infty.$$

A key ingredient to Proposition 4.5 is the following theorem. This was proven in Lai (1974, Theorem 3) for i.i.d. random variables; we extend it to martingale difference sequences.

Theorem 4.3. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. Write $S_n = \sum_{i=1}^n X_i$, and suppose $\sup_{i\in\mathbb{N}} E[X_i^2 \mid \mathcal{F}_{i-1}] \leq \sigma^2 < \infty$ a.s., for some $\sigma > 0$. Let $a > -1$, $p > 2(a+2)$ and $\delta > \sigma\sqrt{1+a}$. If $\sup_{i\in\mathbb{N}} E|X_i|^p \leq C < \infty$ for some $C > 0$, then*

$$\sum_{n=1}^{\infty} n^a P\left(|S_n| > \delta\sqrt{2n\log(n)}\right) < \infty, \tag{4.20}$$

$$\sum_{n=1}^{\infty} n^a P\Big(\sup_{1\leq i\leq n} |S_i| > \delta\sqrt{2n\log(n)}\Big) < \infty. \tag{4.21}$$

The proof makes use of the following result, which is based on Stout (1970).

Lemma 4.1. *Let $(X_i)_{i\in\mathbb{N}}$, S_n and σ^2 be as in Theorem 4.3. If $\max_{1\leq i\leq n} |X_i|/(\sigma\sqrt{n}) \leq c$ a.s., for some $c > 0$, then for all $0 \leq \epsilon \leq c^{-1}$,*

$$P\left(S_n > \epsilon\sigma\sqrt{n}\right) \leq \exp(-(\epsilon^2/2)(1-\epsilon c/2)).$$

The proof of Theorem 4.3 is not valid when $\delta < \sigma\sqrt{1+a}$. The method to proof Proposition 4.4 can therefore not easily be extended for all $c_\rho^* > 0$.

4.4.4 Quality of regret bounds

The results from Section 4.4.1 and 4.4.2 show that the adaptive pricing policy Φ_A can achieve a regret bound $\text{Regret}(T) = O(T^{2/3})$ in the case of general link functions, and $\text{Regret}(T) = O(\sqrt{T\log(T)})$ in the case of canonical link functions. In this section, we address the question how good these bounds are.

For canonical link functions, the rate $\sqrt{T\log(T)}$ is close to $\sqrt{T}$. In several (single product) settings, it has been shown that $\sqrt{T}$ is the best achievable upper bound on the regret, (see e.g. Kleinberg and Leighton, 2003, Besbes and Zeevi, 2011, Broder and Rusmevichientong, 2012). Thus, apart from the $\sqrt{\log(T)}$ term, the adaptive pricing policy achieves the optimal asymptotic growth rate of the regret. This raises the question if the $\sqrt{\log(T)}$ term is only a result from the used proof-techniques, or that the regret really grows at rate $\sqrt{T\log(T)}$.

The term $\sqrt{\log(T)}$ can be traced back to two sources: Proposition 4.5 and Proposition 7.2. Proposition 4.5 is a building block to prove that for sufficiently large t, a solution to the likelihood equations exists in a neighborhood of $\beta^{(0)}$. It considers random variables of the form $T = \sup\{n \in$

$\mathbb{N} \mid |S_n| \geq c\sqrt{n\log(n)}\}$, where S_n is a martingale and $c > 0$, and discusses finiteness of moments of T. Clearly, the $\sqrt{\log(n)}$ term cannot be removed here, since martingales S_n for which $\sup\{n \in \mathbb{N} \mid |S_n| \geq c\sqrt{n}\} = \infty$ a.s. are easily constructed. The second source of the $\sqrt{\log(T)}$ term is Proposition 7.2, where bounds are derived on the expected squared norm of the difference between a least-squares estimate and the true parameter. Similar to Lai and Wei (1982), a $\log(t)$ term appears in the equations. An example provided by Nassiri-Toussi and Ren (1994) shows that at least in some instances, the $\log(t)$ term is present in the asymptotic behavior of the estimates. This implies that Proposition 7.2 is sharp, in the sense that the $\log(t)$-term cannot be removed. However, in this chapter we deal with a particular pricing policy, and it is unclear if this $\log(t)$-term plays a role in the convergence rates of the estimators induced by our pricing policy.

For general link functions, our adaptive pricing policy Φ_A can achieve Regret$(T^{2/3})$. The gap with $\sqrt{T}$ is caused by the bounds on the estimation error in Proposition 4.3. These bounds are based on Theorem 7.2, where for maximum quasi-likelihood estimation with general link functions and adaptive design, mean square error bounds are derived of the form

$$E\left[\left|\left|\hat{\beta}(t) - \beta^{(0)}\right|\right|^2\right] = O\left(\frac{\log(t)}{L_1(t)} + \frac{t}{L_1(t)L_2(t)}\right). \tag{4.22}$$

Here $L_2(t)$ is a non-random lower bound on the one-but-smallest eigenvalue of $P(t)$. The term $\frac{t}{L_1(t)L_2(t)}$ is not present if the link functions are canonical, cf. Proposition 4.4, and essentially causes the difference between the regret bounds for general and for canonical link functions. It is not clear if the bounds (4.22) are tight. But, a noteworthy recent study by Yin et al. (2008) considering maximum quasi-likelihood estimation with adaptive design, general link functions, and multivariate response data, provides convergence rates that, in case of bounded design, imply

$$\left|\left|\hat{\beta}(t) - \beta^{(0)}\right|\right|^2 = o\left(\frac{t}{\lambda_{\min}(t)^2}\log(t)(\log(\log(t)))^{1+2\delta}\right) \text{ a.s., for any } \delta > 0,$$

where $\lambda_{\min}(t)$ denotes the smallest eigenvalue of $P(t)$. These bounds are even slightly worse than the (4.22). The question on tight bounds on the convergence rates of maximum quasi-likelihood estimates with general link functions and adaptive design, remains an open question.

4.5 Numerical illustration

In this section we provide two numerical illustration of the proposed adaptive pricing policy Φ_A. The first considers two products with Poisson distributed demand, and non-canonical link functions. The second is a larger instance, with ten products, normally distributed demand, and canonical link functions.

4.5.1 Two products, Poisson distributed demand

Consider two products with Poisson distributed demand, with expectation

$$E[D_1(p_1, p_2)] = 11.5 - 1.25p_1 + 0.34p_2,$$

$$E[D_2(p_1, p_2)] = 10.22 + 0.25p_1 - 1.55p_2.$$

The lowest and highest admissible price are set to $\mathbf{p}_l = (1, 3, 3)^T$ and $\mathbf{p}_h = (1, 7, 7)^T$, and the three linearly independent initial prices are $\mathbf{p}_1 = (1, 3.0, 6.7)^T$, $\mathbf{p}_2 = (1, 3.3, 3.1)^T$, $\mathbf{p}_3 = (1, 6.7, 6.8)^T$.

The optimal price is $\mathbf{p}_{\text{opt}} = (1, 5.63, 4.37)^T$, with expected revenue 54.7. We apply the adaptive pricing policy Φ_A with $L_1(t) = 0.2 \cdot t^{2/3}$.

The plots in Figure 4.1 show a sample path of the price dispersion $\text{tr}(P(t)^{-1})^{-1}$ divided by $t^{2/3}$, the squared norm $||\hat{\beta}(t) - \beta^{(0)}||^2$ of the difference between the parameter estimates and the true parameter, Regret(t), and Regret$(t)/t^{2/3}$. These pictures illustrate our analytical results that $\text{tr}(P(t)^{-1})^{-1} \geq 0.2t^{2/3}$ for all sufficiently large t, $\lim_{t\to\infty} ||\hat{\beta}(t) - \beta^{(0)}||^2 = 0$, and Regret$(t) = O(t^{2/3})$.

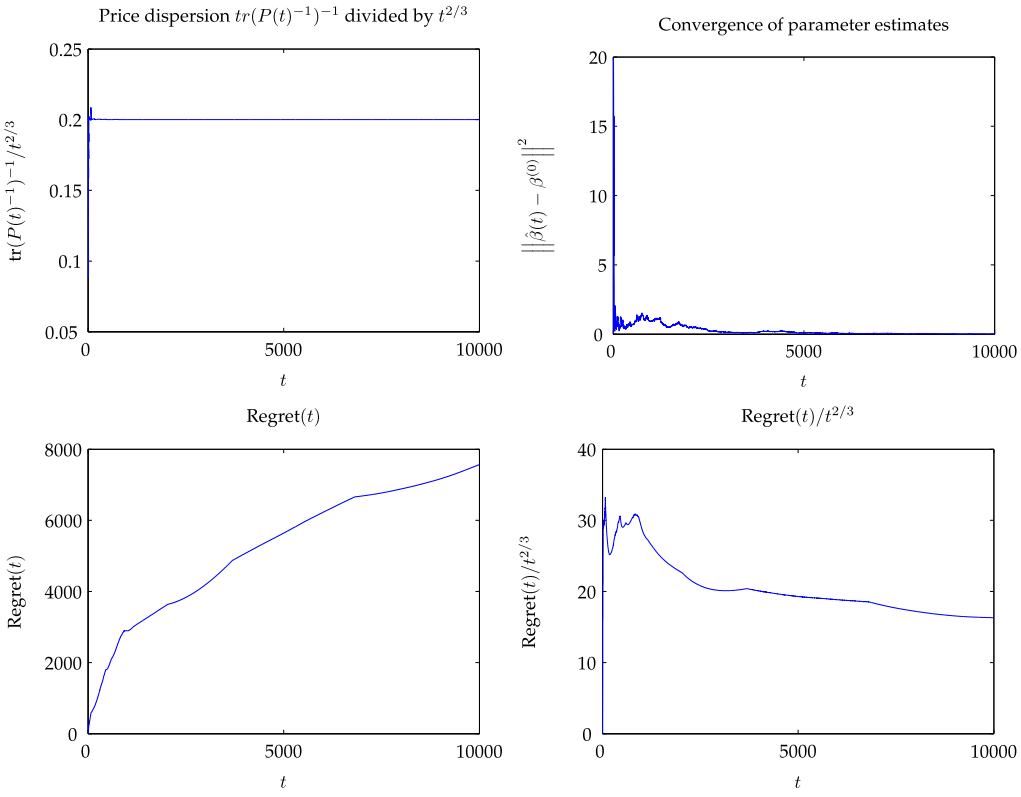

Figure 4.1: Numerical results for Section 4.5.1

4.5.2 Ten products, normally distributed demand

We here consider a large instance, with ten products. The demand for each product k is normally distributed with expectation and variance given by

$$E[D_k(\mathbf{p})] = \beta_{k0}^{(0)} + \beta_{k1}^{(0)} p_1 + \ldots + \beta_{kn}^{(0)} p_n, \quad (k = 1, \ldots, n),$$

$$\text{Var}[D_k(\mathbf{p})] = \sigma_k^2, \quad (k = 1, \ldots, n),$$

where $\beta^{(0)}$ is equal to

$$(\beta_{kl}^{(0)})_{k=1..n,l=0..n} = \begin{pmatrix} 16.32 & -3.10 & 0.10 & 0.09 & 0.19 & 0.11 & 0.16 & 0.10 & 0.12 & 0.06 & 0.16 \\ 19.57 & 0.11 & -3.40 & 0.04 & 0.10 & 0.02 & 0.12 & 0.06 & 0.01 & 0.01 & 0.03 \\ 17.10 & 0.03 & 0.09 & -2.49 & 0.18 & 0.07 & 0.15 & 0.05 & 0.13 & 0.15 & 0.17 \\ 17.70 & 0.10 & 0.02 & 0.10 & -2.37 & 0.17 & 0.03 & 0.08 & 0.08 & 0.13 & 0.15 \\ 18.04 & 0.04 & 0.03 & 0.10 & 0.11 & -2.22 & 0.06 & 0.17 & 0.10 & 0.04 & 0.16 \\ 19.13 & 0.16 & 0.12 & 0.08 & 0.09 & 0.01 & -2.55 & 0.15 & 0.08 & 0.08 & 0.11 \\ 18.12 & 0.17 & 0.05 & 0.16 & 0.09 & 0.05 & 0.07 & -2.02 & 0.07 & 0.13 & 0.04 \\ 15.88 & 0.10 & 0.02 & 0.12 & 0.16 & 0.01 & 0.01 & 0.00 & -3.26 & 0.13 & 0.18 \\ 17.96 & 0.17 & 0.04 & 0.03 & 0.11 & 0.20 & 0.20 & 0.16 & 0.19 & -2.59 & 0.12 \\ 17.45 & 0.02 & 0.07 & 0.14 & 0.19 & 0.19 & 0.09 & 0.05 & 0.02 & 0.18 & -2.37 \end{pmatrix}$$

and

$$(\sigma_1^2, \ldots, \sigma_{10}^2)^T = \begin{pmatrix} 0.55 \\ 0.64 \\ 0.61 \\ 0.64 \\ 0.74 \\ 0.77 \\ 0.92 \\ 0.99 \\ 0.52 \\ 0.62 \end{pmatrix}.$$

The eleven linearly independent initial prices $\mathbf{p}(1), \ldots, \mathbf{p}(11)$ are set to

$$\mathbf{p}(1) = \begin{pmatrix} 1 \\ 18.59 \\ 1.81 \\ 13.09 \\ 6.11 \\ 19.32 \\ 4.23 \\ 10.65 \\ 13.27 \\ 15.64 \\ 1.76 \end{pmatrix}, \mathbf{p}(2) = \begin{pmatrix} 1 \\ 4.48 \\ 1.33 \\ 5.34 \\ 9.26 \\ 10.75 \\ 14.18 \\ 1.23 \\ 14.06 \\ 18.87 \\ 8.36 \end{pmatrix}, \quad \mathbf{p}(3) = \begin{pmatrix} 1 \\ 19.04 \\ 18.34 \\ 19.61 \\ 18.98 \\ 11.24 \\ 10.47 \\ 6.34 \\ 14.4 \\ 18.44 \\ 18.63 \end{pmatrix}, \mathbf{p}(4) = \begin{pmatrix} 1 \\ 15.04 \\ 3.17 \\ 14.61 \\ 5.79 \\ 16.51 \\ 17.67 \\ 1.49 \\ 9.14 \\ 17.78 \\ 14.32 \end{pmatrix}, \quad \mathbf{p}(5) = \begin{pmatrix} 1 \\ 14.3 \\ 4.99 \\ 11.79 \\ 2.33 \\ 3.02 \\ 8.18 \\ 4.65 \\ 7.45 \\ 1.31 \\ 2.81 \end{pmatrix}, \mathbf{p}(6) = \begin{pmatrix} 1 \\ 9.7 \\ 7.76 \\ 4.82 \\ 5.46 \\ 11.88 \\ 16.83 \\ 17.51 \\ 2.94 \\ 10.28 \\ 5.81 \end{pmatrix},$$

$$\mathbf{p}(7) = \begin{pmatrix} 1 \\ 13.06 \\ 1.47 \\ 2.86 \\ 12.06 \\ 16.61 \\ 5.18 \\ 10.57 \\ 4.46 \\ 5.67 \\ 6.66 \end{pmatrix}, \mathbf{p}(8) = \begin{pmatrix} 1 \\ 19.74 \\ 6.61 \\ 2.92 \\ 16.96 \\ 17.55 \\ 16.34 \\ 19.51 \\ 14.3 \\ 19.51 \\ 10.18 \end{pmatrix}, \quad \mathbf{p}(9) = \begin{pmatrix} 1 \\ 9.45 \\ 18.81 \\ 2.26 \\ 2.28 \\ 4.1 \\ 12.21 \\ 1.62 \\ 11.14 \\ 19.42 \\ 10.5 \end{pmatrix}, \mathbf{p}(10) = \begin{pmatrix} 1 \\ 2.1 \\ 17.23 \\ 10.77 \\ 7.21 \\ 10.89 \\ 13.56 \\ 7.34 \\ 11.81 \\ 9.82 \\ 13.74 \end{pmatrix}, \quad \mathbf{p}(11) = \begin{pmatrix} 1 \\ 3.8 \\ 7.1 \\ 3.15 \\ 6.73 \\ 2.26 \\ 9.05 \\ 6.5 \\ 5.31 \\ 12.12 \\ 7.51 \end{pmatrix}.$$

The lowest and highest admissible price are $\mathbf{p}_l = (1, 1, 1, \ldots, 1)^T$ and $\mathbf{p}_h = (1, 20, 20, \ldots, 20)^T$. The optimal price is $\mathbf{p}_{\text{opt}} = (1.00, 5.09, 3.73, 5.23, 3.68, 3.63, 6.90, 3.89, 3.58, 3.51, 4.56)^T$ with expected revenue 381.9. We apply the adaptive pricing policy Φ_A with $L_1(t) = 0.05\sqrt{t \log(t)}$.

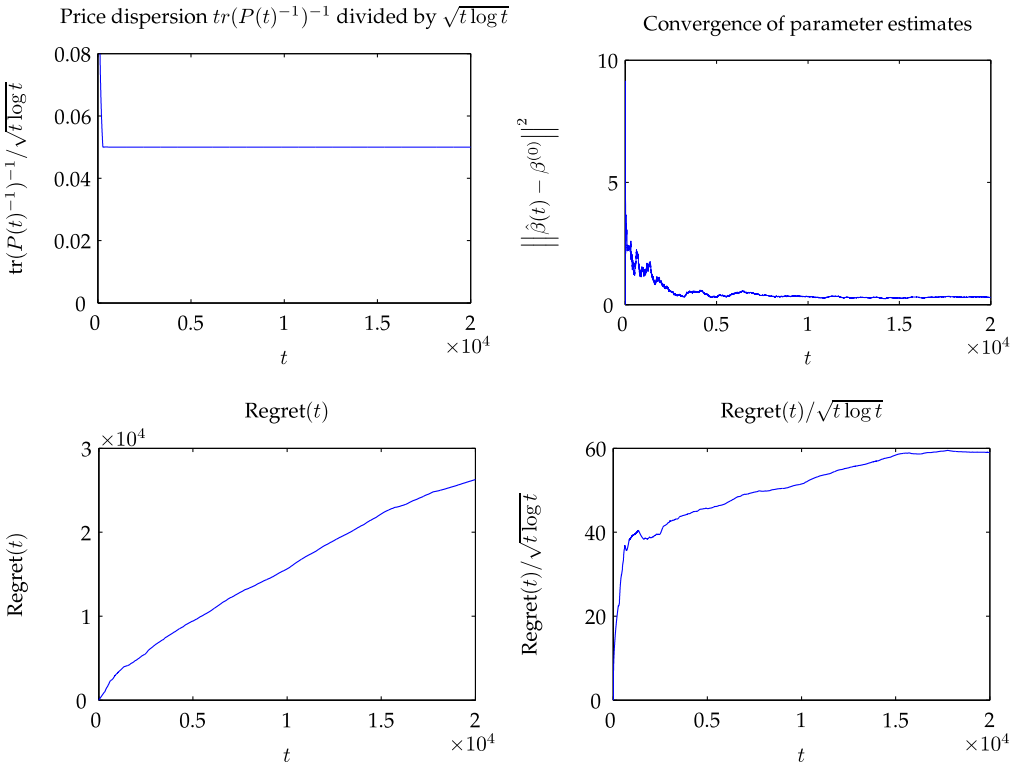

Figure 4.2: Numerical results for Section 4.5.2

The plots in Figure 4.2 show a sample path of the price dispersion $\text{tr}(P(t)^{-1})^{-1}$ divided by $\sqrt{t\log(t)}$, the squared norm $||\hat{\beta}(t) - \beta^{(0)}||^2$ of the difference between the parameter estimates and the true parameter, Regret(t), and Regret$(t)/\sqrt{t\log(t)}$. These pictures illustrate our results that $\text{tr}(P(t)^{-1})^{-1} \geq 0.05\sqrt{t\log(t)}$ for all sufficiently large t,
$\lim_{t\to\infty} ||\hat{\beta}(t) - \beta^{(0)}||^2 = 0$, and Regret$(t) = O(\sqrt{t\log(t)})$.

4.6 Proofs

4.6.1 Proofs of Section 4.3

Proof of Proposition 4.1

Let $t > n+1$ and assume (4.12) and (4.13). Let $\lambda_1 \geq \ldots \geq \lambda_{n+1} > 0$ be the eigenvalues of $P(t)$, and let $v_1, \ldots, v_{n+1}$ be associated eigenvectors. Since $P(t)$ is symmetric, we can assume that $v_1, \ldots, v_{n+1}$ form an orthonormal basis of $\mathbb{R}^{n+1}$.

Choose some $\phi = (\phi_0, \phi_1, \ldots, \phi_n) \in \text{int}(\mathcal{P})$ and $r \in (0,1)$, such that $\{(p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} \mid p_0 = 1, \sup_{k=1,\ldots,n} |p_k - \phi_k| \leq r\} \subset \mathcal{P}$, and let $\phi = \sum_{i=1}^{n+1} \alpha_i v_i$ expressed in the basis induced by the eigenvectors. Define $\mathbf{q} = \phi + \epsilon(v_{n+1,1}\phi - v_{n+1})$, where ϵ is chosen such that

$$|\epsilon| = \min_{k=1,\ldots,n} r(1+\phi_k)^{-1},$$

and

$$\epsilon \geq 0 \text{ if } \alpha_{n+1} \leq 0, \epsilon < 0 \text{ if } \alpha_{n+1} > 0.$$

Note that ϵ^2 is independent of t (but sign(ϵ) is not). We choose $T_0 \in \mathbb{N}$ such that

$$\dot{L}_1(t) \leq \epsilon^2 (n+1)^{-2} \left(1 + L_1(n+1)^{-1} \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2\right)^{-1},$$

for all $t \geq T_0$. The existence of such a T_0 follows from $\dot{L}_1(t) = o(1)$.

Now $\mathbf{q}_0 = 1$, and for all $k = 1, \ldots, n$,

$$|q_k - \phi_k| = |\epsilon||(v_{n+1,1}\phi_k - v_{n+1,k})| \leq |\epsilon|(\phi_k + 1) \leq r,$$

since $|v_{n+1,i}| \leq 1$ for all i. By construction of ϕ and r, this implies $\mathbf{q} \in \mathcal{P}$.

Observe

$$\begin{aligned}\mathbf{q}^T P(t)^{-1}\mathbf{q} &\leq \lambda_{\max}(P(t)^{-1})\,||\mathbf{q}||^2 = \lambda_{\min}(P(t))^{-1}\,||\mathbf{q}||^2 \\ &\leq L_1(t)^{-1}\,||\mathbf{q}||^2 \leq L_1(n+1)^{-1} \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2.\end{aligned}$$

Furthermore,

$$\begin{aligned}
\left|\left|P(t)^{-1}\mathbf{q}\right|\right|^2 &= \left|\left|P(t)^{-1}\left(\sum_{i=1}^{n}(1+\epsilon v_{n+1,1})\alpha_i v_i + ((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)v_{n+1}\right)\right|\right|^2 \\
&= \left|\left|\sum_{i=1}^{n}(1+\epsilon v_{n+1,1})\alpha_i\lambda_i^{-1}v_i + ((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)\lambda_{n+1}^{-1}v_{n+1}\right|\right|^2 \\
&= \sum_{i=1}^{n}(1+\epsilon v_{n+1,1})^2\alpha_i^2\lambda_i^{-2} + ((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)^2\lambda_{n+1}^{-2} \\
&\geq ((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)^2\lambda_{n+1}^{-2} \\
&\geq ((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)^2(n+1)^{-2}L_1(t+1)^{-2},
\end{aligned}$$

since

$$\begin{aligned}
\lambda_{n+1} &\leq (n+1)\text{tr}(P(t)^{-1})^{-1} \leq (n+1)\text{tr}\left(\left(P(t)+\mathbf{p}(\hat{\beta}(t))\mathbf{p}(\hat{\beta}(t))^T\right)^{-1}\right)^{-1} \\
&< (n+1)L_1(t+1).
\end{aligned}$$

Note that $|\epsilon| < 1$ and thus $1+\epsilon v_{n+1,1} \geq 0$. By choice of the sign of ϵ it follows that

$$((1+\epsilon v_{n+1,1})\alpha_{n+1}-\epsilon)^2 \geq (1+\epsilon v_{n+1,1})^2\alpha_{n+1}^2 + \epsilon^2 \geq \epsilon^2,$$

and thus

$$\left|\left|P(t)^{-1}\mathbf{q}\right|\right|^2 \geq \epsilon^2(n+1)^{-2}L_1(t+1)^{-2}. \tag{4.23}$$

The definition of T_0 implies that for $t \geq T_0$,

$$\frac{\left|\left|P(t)^{-1}\mathbf{q}\right|\right|^2}{1+\mathbf{q}^T P(t)^{-1}\mathbf{q}} \geq \frac{\epsilon^2(n+1)^{-2}L_1(t+1)^{-2}}{1+L_1(n+1)^{-1}\max_{\mathbf{p}\in\mathcal{P}}||\mathbf{p}||^2} \geq \frac{\dot{L}_1(t)}{L_1(t+1)^2}.$$

Proof of Proposition 4.2

In IIa) and IIb) of the pricing policy, a decision is made only for the next price $\mathbf{p}(t+1)$. In I) and IIc), decisions are made for a number of forthcoming periods: prices $\mathbf{p}(1), \mathbf{p}(2), \ldots$ are repeated in periods $t+1, t+2, \ldots$, until $\text{tr}(P(t+j)^{-1}) \geq L_1(t+j)$, in I), or until (4.11) has a feasible solution, in IIc). By Proposition 4.1, IIc) does not occur for $t \geq T_0$.

In addition, since prices are merely repeated in I), and $\text{tr}(P(ct)^{-1})^{-1} = c\text{tr}(P(t)^{-1})^{-1}$ for all $c \in \mathbb{N}$, and $L(t) = o(t)$, it follows that the j in I) has to be finite. This implies the existence of a $T_1 \geq T_0$ such that $\text{tr}(P(T_1)^{-1})^{-1} \geq L_1(T_1)$.

From the assumption $\lim_{t\to\infty}\dot{L}_1(t) = 0$ one can derive that there exists a $C_L \in \mathbb{N}$ such that $L_1(C_L t) \leq C_L L_1(t)$ for all $t \in \mathbb{N}$.

We now show for all $t \geq T_1$ that if $\text{tr}\left(P(t)^{-1}\right)^{-1} \geq L_1(t)$, then the following holds:
1) If $\hat{\beta}_k(t)$ does not exist for some k, then

$$\text{tr}\left(P(t+j)^{-1}\right)^{-1} \geq L_1(t+j) \quad \text{for some } 1 \leq j \leq (C_L-1)t,$$

and

$$\text{tr}\left(P(t+i)^{-1}\right)^{-1} \geq C_L^{-1} L_1(t+i) \quad \text{for all } 1 \leq i \leq j;$$

2) If $\hat{\beta}_k(t)$ exists for all k, then

$$\text{tr}\left(P(t+1)^{-1}\right)^{-1} \geq L_1(t+1). \tag{4.24}$$

1) First suppose that $\text{tr}\left(P(t)^{-1}\right)^{-1} \geq L_1(t)$, and $\hat{\beta}_k(t)$ does not exist for some k. Then by I) in the adaptive pricing policy Φ_A, $\mathbf{p}(t+i) = \mathbf{p}(i)$, for $i = 1, \ldots, j$, for some $j \in \mathbb{N}$, where j is the smallest number such that $\text{tr}\left(P(t+j)^{-1}\right)^{-1} \geq L_1(t+j)$. From

$$\text{tr}\left(P(C_L t)^{-1}\right)^{-1} = C_L \text{tr}\left(P(t)^{-1}\right)^{-1} \geq C_L L_1(t) \geq L_1(C_L t)$$

follows that $j \leq (C_L - 1)t$. Moreover, for all $t+i$, $i = 1, \ldots, j$ it holds that

$$\text{tr}\left(P(t+i)^{-1}\right)^{-1} \geq \text{tr}\left(P(t)^{-1}\right)^{-1} \geq L_1(t) \geq C_L^{-1} L_1(C_L t) \geq C_L^{-1} L_1(t+i).$$

Thus, if $\text{tr}\left(P(t)^{-1}\right)^{-1} \geq L_1(t)$, then there is a $1 \leq j \leq (C_L - 1)t$ such that for all $1 \leq i < j$, $\text{tr}\left(P(t+j)^{-1}\right)^{-1} \geq L_1(t+j)$ and $\text{tr}\left(P(t+i)^{-1}\right)^{-1} \geq C_L^{-1} L_1(t+i)$.

2) Now suppose that $\text{tr}\left(P(t)^{-1}\right)^{-1} \geq L_1(t)$, and $\hat{\beta}_k(t)$ does exist for all k. The case (4.10) is trivial; suppose that (4.10) does not hold. Then $\mathbf{p}(t+1)$ is determined by IIb). The Sherman-Morrison formula (Bartlett, 1951),

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u},$$

applied to $A = P(t)$, $u = v = \mathbf{p}(t+1)$, implies

$$\begin{aligned}
&\text{tr}\left(P(t) + \mathbf{p}(t+1)\mathbf{p}^T(t+1)\right)^{-1}) \\
=&\text{tr}(P(t)^{-1}) - \frac{\text{tr}((P(t)^{-1}\mathbf{p}(t+1))(P(t)^{-1}\mathbf{p}(t+1))^T)}{1 + \mathbf{p}(t+1)P(t)^{-1}\mathbf{p}(t+1)} \\
=&\text{tr}(P(t)^{-1}) - \frac{\left|\left|P(t)^{-1}\mathbf{p}(t+1)\right|\right|^2}{1 + \mathbf{p}(t+1)P(t)^{-1}\mathbf{p}(t+1)} \\
\leq&\frac{1}{L_1(t)} + \frac{\partial}{\partial t}\frac{1}{L_1(t)}.
\end{aligned}$$

By a Taylor expansion it follows that $\frac{1}{L_1(t+1)} = \frac{1}{L_1(t)} + \frac{\partial}{\partial t}\frac{1}{L_1(t)}\Big|_{t=\tilde{t}}$, for some $\tilde{t}$ between t and $t+1$. Since $t \mapsto \frac{1}{L_1(t)}$ is convex, $\frac{1}{L_1(t)} + \frac{\partial}{\partial t}\frac{1}{L_1(t)} \leq \frac{1}{L_1(t+1)}$ and thus

$$\text{tr}\left(P(t) + \mathbf{p}(t+1)\mathbf{p}^T(t+1)\right)^{-1})^{-1} \geq L_1(t+1).$$

4.6.2 Proofs of Section 4.4.1

Proof of Theorem 4.1

Since $\mathbf{p}(\beta^{(0)}) \in \text{int}(\mathcal{P})$, it follows that $\frac{\partial r(\mathbf{p},\beta^{(0)})}{\partial \mathbf{p}_k} = 0$ in the point $\mathbf{p}_{\text{opt}}$, for all $k = 1, \ldots, n$. A Taylor series expansion of $\mathbf{p} \mapsto r(\mathbf{p}, \beta^{(0)})$ then implies

$$r(\mathbf{p}_{\text{opt}}, \beta^{(0)}) - r(\mathbf{p}, \beta^{(0)}) \leq \left(\sup_{\mathbf{p} \in \mathcal{P}} |\lambda_{\max}(\nabla^2 r(\mathbf{p}, \beta^{(0)})| \right) \cdot \left|\left| \mathbf{p}_{\text{opt}} - \mathbf{p} \right|\right|^2, \tag{4.25}$$

for all $\mathbf{p} \in \mathcal{P}$.

By assumption, there exists an open neighborhood V of $\beta^{(0)}$ in $\mathbb{R}^{n \times (n+1)}$, such that for all $\beta \in V$, there is a unique optimal price $\mathbf{p}(\beta)$, for which the matrix (4.5) of second partial derivatives w.r.t. $\mathbf{p}_1, \ldots, \mathbf{p}_n$ exists and is negative definite in the point $(\mathbf{p}(\beta), \beta)$. It follows from the implicit function theorem (see e.g. Duistermaat and Kolk, 2004) that V can be chosen such that the function $\beta \mapsto \mathbf{p}(\beta)$ is continuously differentiable with bounded derivatives. Then, by a Taylor expansion,

$$\left|\left| \mathbf{p}(\beta) - \mathbf{p}(\beta^{(0)}) \right|\right| \leq K_1 \left|\left| \beta - \beta^{(0)} \right|\right|,$$

for all $\beta \in V$ and some non-random constant $K_1 > 0$. By choice of ρ_1, $\hat{\beta}(t) \in V$ for all $t > T_\rho$, and since $T_2 \geq T_\rho$ a.s. we obtain

$$\left|\left| \mathbf{p}(\beta^{(0)}) - \mathbf{p}(\hat{\beta}(t)) \right|\right|^2 \mathbf{1}_{t > T_2} \leq K_1 \left|\left| \beta^{(0)} - \hat{\beta}(t) \right|\right|^2 \mathbf{1}_{t > T_2} \text{ a.s.}$$

We have

$$\begin{aligned}
& E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(t) - \mathbf{p}_{\text{opt}} \right|\right|^2 \right] \\
\leq & E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(t) - \mathbf{p}_{\text{opt}} \right|\right|^2 \mathbf{1}_{t > T_2} \right] + E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(t) - \mathbf{p}_{\text{opt}} \right|\right|^2 \mathbf{1}_{t \leq T_2} \right] \\
\leq & 2E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(t) - \mathbf{p}(\hat{\beta}(t-1)) \right|\right|^2 \mathbf{1}_{t > T_2} \right] + 2E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(\hat{\beta}(t-1)) - \mathbf{p}_{\text{opt}} \right|\right|^2 \mathbf{1}_{t > T_2} \right] \\
& + E\left[\sum_{t=t_0}^{T} \left|\left| \mathbf{p}(t) - \mathbf{p}_{\text{opt}} \right|\right|^2 \mathbf{1}_{t \leq T_2} \right] \\
\leq & 2K_2 L_1(T) + 2K_1 E\left[\sum_{t=t_0}^{T} \left|\left| \hat{\beta}(t-1) - \beta^{(0)} \right|\right|^2 \mathbf{1}_{t > T_2} \right] + \sum_{t=t_0}^{T} \max_{\mathbf{q}, \mathbf{q}' \in \mathcal{P}} ||\mathbf{q} - \mathbf{q}'||^2 \, P\left(t \leq T_2\right) \\
= & O\left(L_1(T) + \sum_{t=1}^{T} \left(L_1(t)^{-1} \log(t) + t L_1(t)^{-2} + E\left[T_2^{1/2} \right] t^{-1/2} \right) \right).
\end{aligned}$$

Since $\sum_{t=1}^{T} E\left[T_2^{1/2} \right] t^{-1/2} = O(T^{1/2}) = o(L_1(T))$, it follows by (4.25) that

$$\text{Regret}(T) = O\left(L_1(T) + \sum_{t=1}^{T} \left(L_1(t)^{-1} \log(t) + t L_1(t)^{-2} \right) \right).$$

Proof of Theorem 4.2

Since (i) of Theorem 4.1 follows from Proposition 4.2 and (4.8), it suffices to show $||\mathbf{p}(t)-\mathbf{p}(\hat{\beta}(t-1))||^2\mathbf{1}_{t>T_2} \leq K_2\dot{L}_1(t)\mathbf{1}_{t>T_2}$ a.s., for appropriately chosen K_2 and T_2 and all $t \in \mathbb{N}$. Because $\sum_{t=1}^{T}\dot{L}_1(t) = O(L_1(T))$, this implies (ii).

Choose any $\rho \in (0,\rho_1)$, let C_L, T_1 be as in Proposition 4.2, and set $T_2 = \max\{C_LT_\rho, T_1, T_3, T_4\}$, where T_3, T_4 are non-random constants specified below. Clearly, $E[T_2^\eta] < \infty$ if and only if $E\left[T_\rho^\eta\right] < \infty$, for all $\eta > 0$, and thus $E[T_2^\eta] < \infty$ for all $0 < \eta < \gamma\alpha - 1$. In particular, $\alpha > \frac{1}{2}$ and $\gamma > 3$ implies $E\left[T_2^{1/2}\right] < \infty$.

T_2 is chosen such that I) and IIc) of the pricing policy, do not occur for $t \geq T_2$. For IIc) this follows from $T_2 \geq T_1 \geq T_0$, together with Proposition 4.1. For I), note that since $\text{tr}(P(T_1)^{-1})^{-1} \geq L_1(T_1)$ is shown in the proof of Proposition 4.2, and since $T_2 \geq \max\{C_LT_\rho, T_1\}$, it suffices to show $\text{tr}(P(C_LT_\rho)^{-1})^{-1} \geq L_1(C_LT_\rho)$. This follows since $\text{tr}(P(T_\rho+j)^{-1})^{-1} \geq L_1(T_\rho+j)$ must hold for some $1 \leq j \leq (C_L-1)T_\rho$, cf. the proof of Proposition 4.2.

Let $\beta \in V$ be arbitrary. The uniqueness of the maximum $\mathbf{p}(\beta)$, together with compactness of $\mathcal{P}$, imply that there is a neighborhood $U_\beta \subset \mathcal{P}$ of $\mathbf{p}(\beta)$, such that $r(\mathbf{p}_1,\beta) > r(\mathbf{p}_2,\beta)$ for all $\mathbf{p}_1 \in U_\beta$, $\mathbf{p}_2 \in \mathcal{P}\backslash U_\beta$. For all $\beta \in V$, choose U_β such that

$$\begin{aligned} l &= \inf_{\beta\in V}\inf_{\mathbf{p}\in U_\beta} \lambda_{\min}(\nabla^2 r(\mathbf{p})) > 0, \\ L &= \sup_{\beta\in V}\sup_{\mathbf{p}\in U_\beta} \lambda_{\max}(\nabla^2 r(\mathbf{p})) < \infty; \end{aligned}$$

in view of (4.5), this is always possible.

Now, fix $t > T_2$; then $\hat{\beta}(t) \in V$. For any $\mathbf{p}' \in U_{\hat{\beta}(t)}$ that is a feasible solution of (4.11), we have $r(\mathbf{p}(t+1),\hat{\beta}(t)) \geq r(\mathbf{p}'(\hat{\beta}(t)),\hat{\beta}(t))$, both in case IIa) and IIb), and thus $\mathbf{p}(t+1) \in U_{\hat{\beta}(t)}$. A Taylor expansion yields

$$\begin{aligned} r(\mathbf{p}(\hat{\beta}(t)),\hat{\beta}(t)) - r(\mathbf{p}(t+1),\hat{\beta}(t)) &= \frac{1}{2}(\mathbf{p}(t+1)-\mathbf{p}(\hat{\beta}(t)))^T\nabla^2 r(\tilde{\mathbf{p}}_1,\hat{\beta}(t))(\mathbf{p}(t+1)-\mathbf{p}(\hat{\beta}(t))) \\ &\geq \frac{l}{2}\left|\left|\mathbf{p}(t+1)-\mathbf{p}(\hat{\beta}(t))\right|\right|^2 \end{aligned}$$

and

$$\begin{aligned} r(\mathbf{p}(\hat{\beta}(t)),\hat{\beta}(t)) - r(\mathbf{p}',\hat{\beta}(t)) &= \frac{1}{2}(\mathbf{p}'-\mathbf{p}(\hat{\beta}(t)))^T\nabla^2 r(\tilde{\mathbf{p}}_2,\hat{\beta}(t))(\mathbf{p}'-\mathbf{p}(\hat{\beta}(t))) \\ &\leq \frac{L}{2}\left|\left|\mathbf{p}'-\mathbf{p}(\hat{\beta}(t))\right|\right|^2, \end{aligned}$$

for some $\tilde{\mathbf{p}}_1$, $\tilde{\mathbf{p}}_2 \in U_{\hat{\beta}(t)}$, and consequently,

$$\begin{aligned} \left|\left|\mathbf{p}(t+1)-\mathbf{p}(\hat{\beta}(t))\right|\right|^2 &\leq 2l^{-1}\left[r(\mathbf{p}(\hat{\beta}(t)),\hat{\beta}(t)) - r(\mathbf{p}(t+1),\hat{\beta}(t))\right] \\ &\leq 2l^{-1}\left[r(\mathbf{p}(\hat{\beta}(t)),\hat{\beta}(t)) - r(\mathbf{p}',\hat{\beta}(t))\right] \\ &\leq l^{-1}L\left|\left|\mathbf{p}'-\mathbf{p}(\hat{\beta}(t))\right|\right|^2. \end{aligned}$$

Assertion (ii) of Theorem 4.1 thus follows if for all $t > T_2$, there exists a $\mathbf{p}' \in U_{\hat{\beta}(t)}$ which is a feasible solution of (4.11), such that $||\mathbf{p}'-\mathbf{p}(\hat{\beta}(t))||^2 \leq K_3\dot{L}_1(t)$ for some $K_3 > 0$ independent of

$\hat{\beta}_t$, t. If $\mathbf{p}(t+1) = \mathbf{p}(\hat{\beta}(t))$, then this holds trivially by choosing $\mathbf{p}' = \mathbf{p}(\hat{\beta}(t))$; assume therefore that $\mathbf{p}(t+1)$ is determined by (4.11).

Let $C_0 > 2$, and

$$T_3 = \sup\left\{ t \in \mathbb{N} \mid \begin{array}{l} \text{there exists a } \beta \in V \text{ and } \mathbf{p} \in \mathbb{R}^{n+1} \backslash U_\beta, \\ \text{such that } ||\mathbf{p}(\beta) - \mathbf{p}||^2 \leq C_0^2 \dot{L}_1(t)(1 + \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2) \end{array} \right\},$$

$$T_4 = \sup\{t \in \mathbb{N} \mid C_0^2 \dot{L}_1(t) > 1 \text{ or } L_1(t+1) > \frac{C_0}{2} L_1(t)\}.$$

It follows from $\dot{L}_1(t) = o(1)$ that T_3 and T_4 are finite.

Let $\lambda_1 \geq \ldots \geq \lambda_{n+1}$ be the eigenvalues of $P(t)$, and $v_1, \ldots, v_{n+1}$ the corresponding normalized eigenvectors. Note that all eigenvalues are real and positive. Since $P(t)$ is symmetric we can choose the eigenvectors such that they form an orthonormal basis. Let $\mathbf{p}(\hat{\beta}(t)) = \sum_{i=1}^{n+1} \alpha_i v_i$ be $\mathbf{p}(\hat{\beta}(t))$ expressed in the orthonormal basis of eigenvectors.

Choose C such that $|C| = C_0$, and

$$\text{sign}(C) = \begin{cases} 1 & \text{if } \alpha_{n+1}(v_{n+1,1}\alpha_{n+1} - 1) = 0, \\ \text{sign}\left(\frac{\alpha_{n+1}}{v_{n+1,1}\alpha_{n+1} - 1}\right) & \text{otherwise,} \end{cases}$$

where $v_{n+1,1}$ is the first component of v_{n+1}. Let

$$\mathbf{p}' = \mathbf{p}(\hat{\beta}(t)) + \sqrt{\dot{L}_1(t)} C (v_{n+1,1}\mathbf{p}(\hat{\beta}(t)) - v_{n+1})$$

Suppose $\mathbf{p}' \notin U_{\hat{\beta}(t)}$. Note that since $||v_{n+1}|| \leq 1$ and $\left|\left|\mathbf{p}(\hat{\beta}(t))\right|\right| \leq \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||$,

$$\left|\left|\mathbf{p}(\hat{\beta}(t)) - \mathbf{p}'\right|\right|^2 = C_0^2 \dot{L}_1(t) \left|\left|v_{n+1,1}\mathbf{p}(\hat{\beta}(t)) - v_{n+1}\right|\right|^2 \leq C_0^2 \dot{L}_1(t)(1 + \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2).$$

Since $t > T_3$, this is a contradiction, and thus $\mathbf{p}' \in U_{\hat{\beta}(t)}$.

We now show that $\mathbf{p}'$ satisfies the constraint in (4.11). Observe that

$$\begin{aligned} \mathbf{p}'^T P(t)^{-1} \mathbf{p}' &\leq \lambda_{\max}(P(t)^{-1}) ||\mathbf{p}'||^2 \leq \lambda_{\min}(P(t))^{-1} \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2 \leq \text{tr}(P(t)^{-1}) \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2 \\ &\leq L_1(n+1)^{-1} \max_{\mathbf{p}\in\mathcal{P}} ||\mathbf{p}||^2 \end{aligned} \tag{4.26}$$

and

$$
\begin{aligned}
&\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2 \\
&= \left|\left| P(t)^{-1} \left(\sum_{i=1}^{n+1} \alpha_i v_i + \sqrt{\dot{L}_1(t)} C v_{n+1,1} \left(\sum_{i=1}^{n+1} \alpha_i v_i \right) - \sqrt{\dot{L}_1(t)} C v_{n+1} \right) \right|\right|^2 \\
&= \left|\left| P(t)^{-1} \begin{pmatrix} (\alpha_{n+1} + \sqrt{\dot{L}_1(t)} C v_{n+1,1} \alpha_{n+1} - \sqrt{\dot{L}_1(t)} C) v_{n+1} \\ + \sum_{i=1}^{n} (1 + \sqrt{\dot{L}_1(t)} C v_{n+1,1}) \alpha_i v_i \end{pmatrix} \right|\right|^2 \\
&= \left|\left| \begin{array}{l} (\alpha_{n+1} + \sqrt{\dot{L}_1(t)} C v_{n+1,1} \alpha_{n+1} - \sqrt{\dot{L}_1(t)} C) \lambda_{n+1}^{-1} v_{n+1} \\ + \sum_{i=1}^{n} (1 + \sqrt{\dot{L}_1(t)} C v_{n+1,1}) \alpha_i \lambda_i^{-1} v_i \end{array} \right|\right|^2 \\
&= (\alpha_{n+1} + \sqrt{\dot{L}_1(t)} C v_{n+1,1} \alpha_{n+1} - \sqrt{\dot{L}_1(t)} C)^2 \lambda_{n+1}^{-2} \\
&+ \sum_{i=1}^{n} (1 + \sqrt{\dot{L}_1(t)} C v_{n+1,1})^2 \alpha_i^2 \lambda_i^{-2} \\
&\geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)} (v_{n+1,1} \alpha_{n+1} - 1))^2 \lambda_{n+1}^{-2} \\
&\geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)} (v_{n+1,1} \alpha_{n+1} - 1))^2 \mathsf{tr}(P(t)^{-1})^2,
\end{aligned}
$$

and thus

$$
\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2 \geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)} (v_{n+1,1} \alpha_{n+1} - 1))^2 L_1(t+1)^{-2}. \tag{4.27}
$$

By construction of C,

$$
\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2 \geq (\alpha_{n+1}^2 + C^2 \dot{L}_1(t) (v_{n+1,1} \alpha_{n+1} - 1)^2) L_1(t+1)^{-2}. \tag{4.28}
$$

If $|v_{n+1,1}\alpha_{n+1} - 1| \geq 1/2$, then

$$
\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2 \geq \frac{1}{4} C^2 \dot{L}_1(t) L_1(t+1)^{-2}.
$$

If $|v_{n+1,1}\alpha_{n+1} - 1| < 1/2$, then $v_{n+1,1} \neq 0$, and $v_{n+1,1}\alpha_{n+1} > 1/2$, thus $\alpha_{n+1}^2 > (1/4) v_{n+1,1}^{-2} \geq 1/4$ since $v_{n+1,1} \leq 1$. We then also have

$$
\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2 \geq \alpha_{n+1}^2 L_1(t+1)^{-2} \geq \frac{1}{4} C_0^2 \dot{L}_1(t) L_1(t+1)^{-2}, \tag{4.29}
$$

since $C_0^2 \dot{L}_1(t) \leq 1$ for $t > T_2 \geq T_4$.

Using $L_1(t+1) \leq \frac{C_0}{2} L_1(t)$ for $t > T_4$, we have

$$
\frac{\left|\left|P(t)^{-1}\mathbf{p}'\right|\right|^2}{1 + \mathbf{p}'^T P(t)^{-1} \mathbf{p}'} \geq \frac{\frac{1}{4} C_0^2 \dot{L}_1(t) L_1(t+1)^{-2}}{1 + L_1(n+1)^{-1} \max_{\mathbf{p} \in \mathcal{P}} ||\mathbf{p}||^2} \geq \frac{\dot{L}_1(t)}{L_1(t)^2}.
$$

4.6.3 Proofs of Section 4.4.2

Proof of Proposition 4.4

The assertions on T_ρ are proven in Theorem 7.1, under the assumption $L_1(n) \geq cn^\alpha$, for some $c > 0$, $\frac{1}{2} < \alpha \leq 1$, $n_0 \in \mathbb{N}$ and all $n \geq n_0$. The proof of that theorem considers last-time variables of the form

$$\begin{aligned} T_{A[i]} &= \sup\{n \geq n_0 \mid \rho^{-1} A_n[i] - \tfrac{1}{d+2d^2}(c_2/2)L_1(n)\}, \\ T_{B[i,j]} &= \sup\{n \geq n_0 \mid B_n[i,j] - \tfrac{1}{d+2d^2}(c_2/2)L_1(n)\}, \\ T_{J[i,j]} &= \sup\{n \geq n_0 \mid \rho J_n[i,j] - \tfrac{1}{d+2d^2}(c_2/2)L_1(n)\}, \end{aligned} \tag{4.30}$$

where $(A_n[i])_{n\in\mathbb{N}}$, $(B_n[i,j])_{n\in\mathbb{N}}$, and $(J_n[i,j])_{n\in\mathbb{N}}$ are martingales with square-integrable differences. It is shown that these last-times are a.s. finite and have finite η-moments, using Proposition 7.1 which considers these properties for general last-times of the form

$$T = \sup\{n \in \mathbb{N} \mid |S_n| \geq cn^\alpha\},$$

for a martingale S_n and constants $c > 0$, $\frac{1}{2} < \alpha \leq 1$.

In the current Proposition, the assertions $T_\rho < \infty$ a.s. and $E\left[T_\rho^\eta\right] < \infty$ follow exactly as in Theorem 7.1, with the difference that we apply Proposition 4.5 instead of Proposition 7.1 on the last-times (4.30). In particular, for fixed $\rho > 0$, let $\sigma_C > 0$ such that

$$\begin{aligned} \sigma_C &\geq \rho^{-1} \max_{1\leq i\leq d}\{\sup_{n\in\mathbb{N}} E\left[(A_n[i] - A_{n-1}[i])^2 \mid \mathcal{F}_{n-1}\right]\}, \\ \sigma_C &\geq \max_{1\leq i,j\leq d}\{\sup_{n\in\mathbb{N}} E\left[(B_n[i,j] - B_{n-1}[i,j])^2 \mid \mathcal{F}_{n-1}\right]\}, \\ \sigma_C &\geq \rho \max_{1\leq i,j\leq d}\{\sup_{n\in\mathbb{N}} E\left[(J_n[i,j] - J_{n-1}[i,j])^2 \mid \mathcal{F}_{n-1}\right]\}. \end{aligned}$$

Furthermore, define $c_2 = \inf_{\mathbf{p}\in\mathcal{P},\beta\in B_\rho} \dot{h}(\mathbf{p}^T\beta)$ (cf. the definition of c_2 in (7.7)), and let $c_\rho^* = \frac{d+2d^2}{(c_2/2)} \cdot 2\sigma_C\sqrt{\eta}$.

Proposition 4.5 implies that the last-times

$$\begin{aligned} T_{A[i]} &= \sup\{n \geq n_0 \mid \rho^{-1} A_n[i] - \tfrac{1}{d+2d^2}(c_2/2)c\sqrt{n\log(n)}\}, \\ T_{B[i,j]} &= \sup\{n \geq n_0 \mid B_n[i,j] - \tfrac{1}{d+2d^2}(c_2/2)c\sqrt{n\log(n)}\}, \\ T_{J[i,j]} &= \sup\{n \geq n_0 \mid \rho J_n[i,j] - \tfrac{1}{d+2d^2}(c_2/2)c\sqrt{n\log(n)}\}, \end{aligned} \tag{4.31}$$

are all finite a.s., with finite η-th moment, provided $c > c_\rho^*$ and $0 < \eta < \frac{\gamma-1}{2}$.

The asymptotic existence, strong consistency, and mean square bounds for $\hat{\beta}_k(t)$, follow directly from Theorems 7.2 and 7.3.

4.6.4 Proofs of Section 4.4.3

Proof of Proposition 4.5

Let $0 < c' < c$ such that $c' > 2\sigma\sqrt{\eta}$. There exists an $n' \in \mathbb{N}$ such that for all $n > n'$,

$$c\sqrt{(n/2)\log(n/2)} - \sqrt{2\sigma^2 n} \geq c'\sqrt{(n/2)\log(n/2)}.$$

For all $n > n'$,

$$\begin{aligned}
P(T > n) &= P\left(\exists k > n : |S_k| \geq c\sqrt{k\log(k)}\right) \\
&\leq \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(\exists 2^{j-1} \leq k < 2^j : |S_k| \geq c\sqrt{k\log(k)}\right) \\
&\leq \sum_{j \geq \lfloor \log_2(n) \rfloor} P\Big(\sup_{1 \leq k \leq 2^j} |S_k| \geq c\sqrt{2^{j-1}\log(2^{j-1})}\Big) \\
&\overset{(1)}{\leq} 2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c\sqrt{2^{j-1}\log(2^{j-1})} - \sqrt{2\sigma^2 2^j}\right) \\
&\overset{(2)}{\leq} 2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c'\sqrt{2^{j-1}\log(2^{j-1})}\right),
\end{aligned}$$

where (1) follows from Lemma 7.4, and (2) from the definition of n'.

By Chebyshev's and Rosenthal's inequality (see e.g. Hall and Heyde, 1980, Theorem 2.12), there is a $C_2 > 0$ such that for all $k > e, c > 0$,

$$\begin{aligned}
P\left(|S_k| \geq c\sqrt{k\log(k)}\right) &\leq E\left[|S_k|^r\right] c^{-r} k^{-r/2} (\log(k))^{-r/2} \\
&\leq C_2 \left((\sigma^2 k)^{r/2} + k \sup_{i \in \mathbb{N}} E\left[|X_i|^r\right]\right) c^{-r} k^{-r/2} (\log(k))^{-r/2} \\
&\leq \left(C_2\sigma^p + C_2 C\right) c^{-r} (\log(k))^{-r/2}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c'\sqrt{2^{j-1}\log(2^{j-1})}\right) \\
\leq &2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq \frac{\sqrt{j-1}}{\sqrt{2j}} c'\sqrt{2^j \log(2^j)}\right) \\
\leq &2 \sum_{j \geq \lfloor \log_2(n) \rfloor} K j^{-r/2} < \infty,
\end{aligned}$$

for some $K > 0$, and thus $P(T = \infty) \leq \liminf_{n\to\infty} P(T > n) = 0$. This proves $T < \infty$ a.s.

For $t \in \mathbb{R}_+$ write $S_t = S_{\lfloor t \rfloor}$. Then

$$\sum_{j \geq \log_2(n)} P\left(|S_{2^j}| \geq c'\sqrt{2^{j-1}\log(2^{j-1})}\right) \tag{4.32}$$

$$= \int_{j \geq \log_2(n)} P\left(|S_{2^j}| \geq c'\frac{\sqrt{j-1}}{\sqrt{2j}}\sqrt{2^j\log(2^j)}\right) dj \tag{4.33}$$

$$= \int_{k \geq n} P\left(|S_k| \geq c'\sqrt{\frac{1}{2} - \frac{\log(2)}{2\log(k)}}\sqrt{k\log(k)}\right) \frac{1}{k\log(2)} dk \tag{4.34}$$

$$\leq \sum_{k \geq n} P\left(|S_k| \geq c'\sqrt{\frac{1}{2} - \frac{\log(2)}{2\log(n')}}\sqrt{k\log(k)}\right) \frac{1}{k\log(2)}, \tag{4.35}$$

using a variable substitution $k = 2^j$. Since

$$\begin{aligned} E[T^\eta] &\leq \eta \left[1 + \sum_{n \geq 1} n^{\eta-1} P(T > n) \right] \\ &\leq \eta \left[1 + n' \max\{1, (n')^{\eta-1}\} + \sum_{n > n'} n^{\eta-1} P(T > n) \right] \\ &\leq M \sum_{n > n'} n^{\eta-1} \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_k| \geq c' \sqrt{\frac{1}{2} - \frac{\log(2)}{2\log(n')}} \sqrt{k \log(k)} \right) k^{-1}, \end{aligned}$$

for some constant $M > 0$, it follows that $E[T^\eta] < \infty$ if

$$\sum_{n \geq 1} n^{\eta-1} \sum_{k \geq n} P\left(|S_k| \geq \delta \sqrt{2k \log(k)} \right) k^{-1} < \infty,$$

where we write $\delta = c' \sqrt{\frac{1}{4} - \frac{\log(2)}{4\log(n')}}$. By interchanging the sums, it suffices to show

$$\sum_{k \geq 1} k^{\eta-1} P\left(|S_k| \geq \delta \sqrt{2k \log(k)} \right) < \infty. \tag{4.36}$$

We can choose n' sufficiently large such that $\delta > \sigma\sqrt{\eta}$. Then (4.36) follows from Theorem 4.3 with $a = \eta - 1$

Proof of Lemma 4.1

Apply Lemma 1 of Stout (1970) on the sequence $\left(\frac{X_i}{\sigma\sqrt{n}} \right)_{1 \leq i \leq n}$, with $l = 0$. Then for all $0 \leq \lambda c \leq 1$,

$$\begin{aligned} 1 &\geq E[\exp(\lambda S_n/(\sigma\sqrt{n})) \exp(-(\lambda^2/2)(1 + \lambda c/2) \left[\sum_{i=1}^{n} E[X_i^2 \mid \mathcal{F}_{i-1}] \right] / (\sigma^2 n))] \\ &\geq E[\exp(\lambda S_n/(\sigma\sqrt{n})) \exp(-(\lambda^2/2)(1 + \lambda c/2)]. \end{aligned}$$

For $0 \leq \epsilon \leq c^{-1}$, we thus have

$$P\left(S_n/(\sigma\sqrt{n}) \geq \epsilon \right) \leq \frac{E\left[\exp(\lambda S_n/(\sigma\sqrt{n}))\right]}{\exp(\lambda\epsilon)} \leq \exp((\lambda^2/2)(1 + \lambda c/2) - \lambda\epsilon).$$

Take $\lambda = \epsilon$ to prove the assertion.

Proof of Theorem 4.3

The proof of Theorem 4.3 uses the concepts median and symmetrization. For a random variable Y, the symmetrization Y^s of Y is defined as $Y^s = Y - Y'$, where Y' is independent of Y and has the same distribution. A median $\text{med}(Y)$ of Y is a scalar $m \in \mathbb{R}$ such that $P(Y \geq m) \geq \frac{1}{2} \leq P(Y \leq m)$. A median always exists, but is not necessarily unique. Moreover, if $E[Y] < \infty$ then $|\text{med}(Y) - E[Y]| \leq \sqrt{2\text{Var}(Y)}$ (Loève, 1977a, 18.1.a).

The first step in the proof is to bound the tail-probabilities $P(S_n > \delta\sqrt{2n \log(n)})$ in terms of symmetrized random variables. To this end, choose $\delta_1 \in (\sigma\sqrt{1+a}, \delta)$ and $n_1 \in \mathbb{N}$ such that

$\delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n} \geq \delta_1\sqrt{2n\log(n)}$ for all $n \geq n_1$. The weak symmetrization inequalities (Loève, 1977a, 18.1.A(i)) state that $P(Y - \text{med}(Y) \geq \epsilon) \leq 2P(Y^s \geq \epsilon)$ for any random variable Y and all $\epsilon > 0$. Since $E[S_n] = 0$ and $\text{Var}(S_n) \leq \sigma^2 n$ for all $n \in \mathbb{N}$, it follows that for all $n \geq n_1$,

$$\begin{aligned} P(S_n > \delta\sqrt{2n\log(n)}) &\leq P(S_n - \text{med}(S_n) > \delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n}) \\ &\leq 2P(S_n^s > \delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n}) \\ &\leq 2P(S_n^s > \delta_1\sqrt{2n\log(n)}). \end{aligned} \tag{4.37}$$

As a next step, we consider a truncation of S_n^s. In particular, for all $i \in \mathbb{N}$, write $\tilde{X}_i^s = X_i^s \mathbf{1}_{|X_i^s| \leq g(i)}$, where $g(i) = \sigma^2\kappa\delta_1^{-1}\sqrt{i}/\sqrt{2\log(i)}$ and $0 < \kappa < 1$ is specified below, and write $\tilde{S}_n^s = \sum_{i=1}^n \tilde{X}_i^s$. Define $T_{\neq} = \sup\{i \in \mathbb{N} \mid X_i^s \neq \tilde{X}_i^s\} = \sup\{i \in \mathbb{N} \mid |X_i^s| > g(i)\}$. Then

$$\sum_{n=1}^{\infty} n^a P\left(S_n^s > \delta_1\sqrt{2n\log(n)}\right) \leq \sum_{n=1}^{\infty} n^a P\left(S_n^s > \delta_1\sqrt{2n\log(n)}, n > T_{\neq}\right) + \sum_{n=1}^{\infty} n^a P\left(n \leq T_{\neq}\right). \tag{4.38}$$

If $n > T_{\neq}$ then $S_n^s = \tilde{S}_n^s + (S_{T_{\neq}}^s - \tilde{S}_{T_{\neq}}^s)$. Let $\delta_2 \in (\sigma\sqrt{1+a}, \delta_1)$ and $n_2 \in \mathbb{N}$ such that $\delta_1\sqrt{2n\log(n)} - (S_T^s - \tilde{S}_T^s) \geq \delta_2\sqrt{2n\log(n)}$ for all $n \geq \max\{T, n_2\}$. Then

$$\begin{aligned} P(S_n^s > \delta_1\sqrt{2n\log(n)}, n > T_{\neq}) &= P(\tilde{S}_n^s > \delta_1\sqrt{2n\log(n)} - (S_T^s - \tilde{S}_T^s), n > T_{\neq}) \\ &\leq P(\tilde{S}_n^s > \delta_2\sqrt{2n\log(n)}). \end{aligned}$$

Note that $(\tilde{X}_i^s)_{i\in\mathbb{N}}$ is a martingale difference sequence w.r.t. $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$, with $\sup_{i\in\mathbb{N}} E[(\tilde{X}_i^s)^2 \mid \mathcal{F}_{i-1}] \leq \sigma^2 < \infty$ a.s.

Let $\epsilon_n = \sigma^{-1}\delta_2\sqrt{2\log(n)}$, choose $\kappa \in (0,1)$ such that $\delta_2 > (1-\frac{\kappa}{2})^{-1/2}\sigma\sqrt{1+a}$, and set $c_n = \kappa\epsilon_n^{-1}$. Then $0 \leq \epsilon_n c_n = \kappa \leq 1$ and $\max_{1\leq i\leq n} |\tilde{X}_i^s| \leq g(n) \leq \sigma\sqrt{n}c_n$, using $\delta_1^{-1} \leq \delta_2^{-1}$. By Lemma 4.1,

$$\begin{aligned} P\left(\tilde{S}_n^s > \delta_2\sqrt{2n\log(n)}\right) = P\left(\tilde{S}_n^s > \epsilon_n\sigma\sqrt{n}\right) &\leq \exp(-(\epsilon_n^2/2)(1-\epsilon_n c_n/2)) \\ &= \exp\left(-\frac{\delta_2^2}{\sigma^2}\left(1-\frac{\kappa}{2}\right)\log(n)\right). \end{aligned}$$

Since $\delta_2 > (1-\frac{\kappa}{2})^{-1/2}\sigma\sqrt{1+a}$, we have $-1 > a - \delta_2^2\sigma^{-2}(1-\frac{\kappa}{2})$ and thus

$$\sum_{n=1}^{\infty} n^a P\left(\tilde{S}_n^s > \delta_2\sqrt{2n\log(n))}\right) = O\left(\sum_{n=1}^{\infty} n^{a-\delta_2^2\sigma^{-2}(1-\kappa/2)}\right) < \infty \text{ a.s.} \tag{4.39}$$

This proves that the first term of the right hand side of (4.38) is finite a.s.

For the second term, we have

$$\begin{aligned}\sum_{n=1}^{\infty} n^a P(T_{\neq} \geq n) &= \sum_{n=1}^{\infty} n^a P\left(\exists k \geq n : |X_k|/g(k) > 1\right) \\ &\leq \sum_{n=1}^{\infty} n^a \sum_{j \geq \lfloor \log_{(2)}(n) \rfloor} P\left(\exists 2^j \leq k < 2^{j+1} : |X_k|/g(k) > 1\right) \\ &\leq \sum_{n=1}^{\infty} n^a \sum_{j \geq \lfloor \log_{(2)}(n) \rfloor} P\left(\sup_{1 \leq k \leq 2^{j+1}} |X_k| > g(2^{j+1})\right) \\ &\leq \sum_{n=1}^{\infty} n^a \sum_{j \geq \lfloor \log_{(2)}(n) \rfloor} \frac{C}{g(2^{j+1})^p}, \end{aligned} \tag{4.40}$$

by Doob's inequality for martingales. Furthermore,

$$\begin{aligned}\sum_{j \geq \lfloor \log_2(n) \rfloor} \frac{1}{g(2^{j+1})^p} &= O\left(\int_{\log_{(2)}(n)}^{\infty} \frac{1}{(\sigma^2 \delta_1^{-1} \kappa)^p (2^{j+1})^{p/2} (\log(2^{j+1}))^{-p/2}} dj\right) \\ &= O\left(\int_{2n}^{\infty} \frac{1}{k^{p/2} (\log(k))^{-p/2}} \cdot \frac{1}{k} dk\right) \\ &= O\left(n^{1-p/2} (\log(n))^{p/2}\right).\end{aligned}$$

where we applied a change of variables $k = 2^{j+1}$. Combining this with (4.40), it follows from the assumption $p > 2(a+2)$ that

$$\sum_{n=1}^{\infty} n^a P(T_{\neq} \geq n) = O\left(\sum_{n=1}^{\infty} n^{a+1-p/2} (\log(n))^{p/2}\right) < \infty. \tag{4.41}$$

It follows from (4.37), (4.38), (4.39) and (4.41) that

$$\sum_{n=1}^{\infty} n^a P(S_n > \delta \sqrt{2n \log(n)}) < \infty.$$

By replacing X_i and S_i with $-X_i$ and $-S_i$ in the proof, (4.20) follows. For (4.21), choose $\delta_3 \in (\sigma\sqrt{1+a}, \delta)$, and let $n_3 \in \mathbb{N}$ such that $\delta_3 \sqrt{2n \log(n)} \leq \delta \sqrt{2n \log(n)} - \sqrt{2\sigma^2 n}$ for all $n \geq n_3$. Then for all $n \geq n_3$,

$$\begin{aligned}P\Big(\sup_{1 \leq i \leq n} |S_i| > \delta \sqrt{2n \log(n)}\Big) &\leq 2P\Big(|S_n| \geq \delta \sqrt{2n \log(n)} - \sqrt{2\sigma^2 n}\Big) \\ &\leq 2P\Big(|S_n| \geq \frac{1}{2} \delta_3 \sqrt{2n \log(n)}\Big),\end{aligned}$$

using Lemma 7.4. Now (4.21) follows from (4.20).

Chapter 5

Dynamic pricing and learning for a single product with finite inventory

5.1 Introduction

In this chapter, we study dynamic pricing and learning in a setting where finitely many products are sold during finite selling periods, and where unsold inventory perishes at the end of each selling season. This setting is applicable for airline tickets, hotel rooms, concert tickets, car rental reservations, and many other products. The key feature that distinguishes this setting from the situation with infinite inventory is the nature of the optimal selling prices. In the single-product infinite-inventory setting, studied in Chapter 3, the optimal selling price is a single value that remains constant over time. In the current setting things are different: the optimal selling price is not a single value, but it changes over time, depending on the inventory level and the remaining length of the selling season (see, for example, Gallego and van Ryzin, 1994). It turns out that these price fluctuations drastically change the structural properties of dynamic pricing and learning problems.

Existing literature on dynamic pricing and learning with a finite inventory focuses on results for a single selling season, cf. Besbes and Zeevi (2009) or Wang et al. (2011). To asses the performance of proposed pricing strategies, they consider an asymptotic regime where the demand rate and the initial amount of inventory grow to infinity. Such an asymptotic regime may have practical value if demand, initial inventory, and the length of the selling season are relatively large. In many situations, however, this is not the case. For example, in the hotel rooms industry, a product may be modeled as a combination of arrival date and length-of-stay (Talluri and van Ryzin, 2004, section 10.2, Weatherford and Kimes, 2003). Different products may have different, overlapping selling periods, and similar demand characteristics. It would therefore be unwise to learn the consumer behavior for each product and selling period separately. In addition, the average demand, initial capacity and length of a selling period may be quite low, which makes the asymptotic regime not a suitable setting to study the performance of pricing strategies. This motivates the present chapter on dynamic pricing of perishable products with finite initial inventory, during multiple *consecutive* selling seasons of finite duration.

Similar to Chapters 3 and 4, we assume a parametric demand model with unknown parameters which can be estimated with maximum-likelihood estimation. We then propose a pricing policy that, apart from a finite number of time periods, essentially is a certainty equivalent policy: at each decision moment the price is chosen that is optimal w.r.t. the available parameter estimates. Somewhat surprisingly, the certainty equivalent policy has a very good performance. The pa-

rameter estimates converge to the correct values, and the selling prices converge to the optimal prices. The regret, which measures the expected amount of revenue loss due to not using the optimal prices, is $O(\log^2(T))$, where T denotes the number of selling seasons. This is considerably better than $\sqrt{T}$, which is the best achievable bound on the regret when inventory is infinite.

This difference in qualitative behavior of the regret can be explained as follows. In the infinite inventory model, prices and parameter estimates can get stuck in what Harrison et al. (2011a) call an "indeterminate equilibrium". This means that for some values of the parameter estimates, the expected observed demand at the certainty equivalent price is equal to what the parameter estimates predict; in other words, the observations confirm the correctness of the (incorrect) parameter estimates. As a result, certainty equivalent pricing induces insufficient dispersion in the chosen selling prices to eventually learn the true value of the parameters. Such cannot occur in the setting with finite inventories and finite selling seasons. An optimal price - optimal w.r.t. certain parameter estimates - is namely not a fixed number, but changes depending on the remaining inventory and the remaining length of the selling season. As a result, an optimal policy naturally induces endogenous price dispersion, and prices cannot get stuck in an "indeterminate equilibrium". On the contrary, the large amount of price dispersion implies that the unknown parameters are learned quite fast and that Regret(T) is only $O(\log^2(T))$.

The rest of this chapter is organized as follows. Section 5.2 discusses the mathematical model, the structure of the demand distribution, the full-information optimal solution, and the regret measure. Section 5.3 shows how the unknown parameters of the model can be estimated, and contains a result concerning the speed at which parameter estimates converge to the true value. The endogenous-learning property of the system is described in Section 5.4; our pricing policy and its performance bound in terms of the regret is contained in Section 5.5. A numerical illustration of the policy and its performance is provided in Section 5.6. All mathematical proofs are contained in Section 5.7.

5.2 Preliminaries

In this section we subsequently introduce the model, describe the characteristics of the demand distribution, discuss the optimal pricing policy under full information, and introduce the regret as quality measure of pricing policies.

5.2.1 Model formulation

We consider a monopolist seller of perishable products which are sold during consecutive selling seasons. Each selling season consists of $S \in \mathbb{N}$ discrete time periods: the i-th selling season starts at time period $1 + (i-1)S$, and lasts until period iS, for all $i \in \mathbb{N}$. We write $SS_t = 1 + \lfloor (t-1)/S \rfloor$ to denote the selling season corresponding to period t, and $s_t = t - (SS_t - 1)S$ to denote the relative time in the selling period. At the start of each selling season, the seller has $C \in \mathbb{N}$ discrete units of inventory at his disposal, which can only be sold during that particular selling season. At the end of a selling season, all unsold inventory perishes.

In each time period $t \in \mathbb{N}$ the seller has to determine a selling price $p_t \in [p_l, p_h]$. Here $0 < p_l < p_h$ denote the lowest and highest price admissible to the firm. After setting the price the seller observes a realization of demand, which takes values in $\{0, 1\}$, and collects revenue. We let c_t, $(t \in \mathbb{N})$, denote the capacity or inventory level at the beginning of period $t \in \mathbb{N}$, and d_t the

demand in period t. The dynamics of $(c_t)_{t\in\mathbb{N}}$ are given by

$$\begin{aligned} c_t &= C, && \text{if } s_t = 1, \\ c_t &= \max\{c_{t-1} - d_{t-1}, 0\}, && \text{if } s_t \neq 1. \end{aligned}$$

The pricing decisions of the seller are allowed to depend on previous prices and demand realizations, but not on future ones. More precisely, for each $t \in \mathbb{N}$ we define the set of possible histories $\mathcal{H}_t$ as

$$\mathcal{H}_t = \{(p_1, \ldots, p_t, d_1, \ldots, d_t) \in [p_l, p_h]^t \times \{0, 1\}^t\},$$

with $\mathcal{H}_0 = \{\emptyset\}$. A pricing strategy $\psi = (\psi_t)_{t\in\mathbb{N}}$ is a collection of functions $\psi_t : \mathcal{H}_{t-1} \to [p_l, p_h]$, such that for each $t \geq 2$, the seller chooses the price $p_t = \psi_t(p_1, \ldots, p_{t-1}, d_1, \ldots, d_{t-1})$, and $p_1 = \psi_1(\emptyset)$.

The revenue collected in period t equals $p_t \min\{c_t, d_t\}$. The purpose of the seller is to find a pricing strategy ψ that maximizes the cumulative expected revenue earned after T selling seasons, $\sum_{i=1}^{TS} E_\psi[p_i \min\{d_i, c_i\}]$. Here we write E_ψ to emphasize that this expectation depends on the pricing strategy ψ.

5.2.2 Demand distribution

The demand in a single time period against selling price p is a realization of the random variable $D(p)$. We assume that $D(p)$ is Bernoulli distributed with mean $E[D(p)] = h(\beta_0 + \beta_1 p)$, for all $p \in [p_l, p_h]$, some $(\beta_0, \beta_1) \in \mathbb{R}^2$, and some function h. The true value of β is denoted by $\beta^{(0)}$, and is unknown to the seller. If we wish to emphasize the dependence of the demand on the unknown parameters β, we write $D(p, \beta)$. Conditionally on selling prices, the demand in any two different time periods are independent.

To ensure existence and uniqueness of revenue-maximizing selling prices, we make a number of assumptions on h and β. First, we assume that $\beta^{(0)}$ lies in the interior of a compact known set $B \subset \mathbb{R}^2$, and assume that $\beta_1 < 0$ for all $\beta \in B$. Second, we assume that h is three times continuously differentiable, log-concave, $h(\beta_0 + \beta_1 p) \in (0, 1)$ for all $\beta \in B$ and $p \in [p_l, p_h]$, and the derivative $\dot{h}(z)$ of $h(z)$ is strictly positive. This last assumption, together with $\beta_1 < 0$ for all $\beta \in B$, implies that expected demand is decreasing in p, for all $\beta \in B$.

Write $r^* = \max_{p\in[p_l,p_h]} p \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p)$, and for $(a, \beta, p) \in \mathbb{R} \times B \times [p_l, p_h]$, define

$$g_{a,\beta}(p) = -(p - a)\beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)}.$$

We assume that $g_{a,\beta^{(0)}}(p_l) < 1$, $g_{a,\beta^{(0)}}(p_h) > 1$, and $g_{a,\beta^{(0)}}(p)$ is strictly increasing in p, for all $0 \leq a \leq r^*$. These conditions, which for $a = 0$ coincide with the assumptions in Lariviere (2006, page 602), ensure that $p \mapsto (p - a)h(\beta_0^{(0)} + \beta_1^{(0)} p)$ has a unique maximizer in (p_l, p_h).

Examples of functions h that satisfy the assumptions (with appropriate conditions on B and $[p_l, p_h]$), are $h(z) = \exp(z)$, $h(z) = z$, and $h(z) = \text{logit}(z) = \exp(z)/(1 + \exp(z))$.

5.2.3 Full-information optimal solution

If the value of β is known, the optimal prices can be determined by solving a Markov decision problem (MDP). Since each selling season corresponds to the same MDP, the optimal pricing

strategy for the infinite-horizon average reward criterion is to repeatedly use the optimal policy for a single selling season. The state space of this MDP is $\mathcal{X} = \{(c, s) \mid c = 0, \ldots, C, s = 1, \ldots, S\}$, where (c, s) means that there are c units of remaining inventory at the beginning of the s-th period of the selling season, and the action space is the interval $[p_l, p_h]$. If action p is used in state (c, s), $s < S$, then with probability $h(\beta_0 + \beta_1 p)$ a state transition $(c, s) \to ((c-1)^+, s+1)$ occurs and reward $ph(\beta_0 + \beta_1 p)\mathbf{1}_{c>0}$ is obtained; with probability $1 - h(\beta_0 + \beta_1 p)$ a state transition $(c, s) \to (c, s+1)$ occurs and zero reward is obtained. If action p is used in state (c, S), then with probability one a state transition $(c, s) \mapsto (C, 1)$ occurs; the obtained reward equals $ph(\beta_0 + \beta_1 p)\mathbf{1}_{c>0}$ with probability $h(\beta_0 + \beta_1 p)$, and zero with probability $1 - h(\beta_0 + \beta_1 p)$.

A (stationary deterministic) policy π is a matrix $(\pi(c, s))_{0 \le c \le C, 1 \le s \le S}$ in the policy space $\Pi = [p_l, p_h]^{(C+1)\times S}$. Given a policy $\pi \in \Pi$, let $V_\beta^\pi(c, s)$ be the expected revenue-to-go function starting in state $(c, s) \in \mathcal{X}$ and using the actions of π. Then $V_\beta^\pi(c, s)$ satisfies the following recursion:

$$\begin{aligned} V_\beta^\pi(c, s) &= P(D(\pi(c, s), \beta) = 0) \cdot V_\beta^\pi(c, s+1) \\ &\quad + P(D(\pi(c, s), \beta) = 1) \cdot (\pi(c, s) + V_\beta^\pi(c-1, s+1)), \quad (c \ge 1), \qquad (5.1) \\ V_\beta^\pi(0, s) &= 0, \qquad (5.2) \end{aligned}$$

for all $1 \le s \le S$, where we write $V_\beta^\pi(c, S+1) = 0$ for all $0 \le c \le C$.

By Proposition 4.4.3 of Puterman (1994), for each $\beta \in B$ there is a corresponding optimal policy $\pi_\beta^* \in \Pi$. This policy can be calculated using backward induction. Write $V_\beta(c, s) = V_\beta^{\pi_\beta^*}(c, s)$ for the optimal revenue-to-go function. Then $V_\beta(c, s)$ and $\pi_\beta^*(c, s)$, for $(c, s) \in \mathcal{X}$, satisfy the following recursion:

$$\begin{aligned} V_\beta(c, s) &= \max_{p \in [p_l, p_h]} \big[p - \Delta V_\beta(c, s+1)\big] h(\beta_0 + \beta_1 p) + V_\beta(c, s+1), \\ \pi_\beta^*(c, s) &\in \arg\max_{p \in [p_l, p_h]} \big[p - \Delta V_\beta(c, s+1)\big] h(\beta_0 + \beta_1 p), \end{aligned} \qquad (5.3)$$

where we define $\Delta V_\beta(c, s) = V_\beta(c, s) - V_\beta(c-1, s)$, and $\Delta V_\beta(0, s) = 0$ for all $1 \le s \le S$.

The optimal average reward of the MDP is equal to $V_\beta(C, 1)$, and the true optimal average reward is equal to $V_{\beta^{(0)}}(C, 1)$.

5.2.4 Regret measure

To asses the quality of the pricing decisions of the seller, we define the regret. This quantity measures the expected amount of money lost due to not using the optimal prices. The regret of pricing strategy ψ after the first T selling seasons is defined as

$$\text{Regret}(\psi, T) = T \cdot V_{\beta^{(0)}}(C, 1) - \sum_{i=1}^{TS} E_\psi[p_i \min\{d_i, c_i\}], \qquad (5.4)$$

where $(p_i)_{i \in \mathbb{N}}$ denote the prices generated by the pricing strategy ψ.

Maximizing the cumulative expected revenue is equivalent to minimizing the regret, but observe that the regret cannot directly be used by the seller to find the optimal strategy, since it depends on the unknown $\beta^{(0)}$. Also note that we calculate the regret over a number of selling seasons, and not over a number of time periods. The reason is that the optimal policy $\pi_{\beta^{(0)}}^*$ is optimized over an entire selling season, and not over each individual state of the underlying MDP: a chosen price p_t may induce a higher instant reward in a certain state (c_t, s_t) than the optimal price $\pi_{\beta^{(0)}}^*(c_t, s_t)$.

This effect is averaged out by looking at the optimal expected reward in an entire selling season.

5.3 Parameter estimation

5.3.1 Maximum-likelihood estimation

The value of $\beta^{(0)}$ can be estimated with maximum-likelihood estimation. In particular, given a sample of prices $p_1, \ldots, p_t$ and demand realizations $d_1, \ldots d_t$, the log-likelihood function $L_t(\beta)$ equals

$$L_t(\beta) = \sum_{i=1}^{t} \log \left[h(\beta_0 + \beta_1 p_i)^{d_i} (1 - h(\beta_0 + \beta_1 p_i))^{1-d_i} \right].$$

The score function, the derivative of $L_t(\beta)$ with respect to β, equals

$$l_t(\beta) = \sum_{i=1}^{t} \frac{\dot{h}(\beta_0 + \beta_1 p_i)}{h(\beta_0 + \beta_1 p_i)(1 - h(\beta_0 + \beta_1 p_i))} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (d_i - h(\beta_0 + \beta_1 p_i)). \tag{5.5}$$

We let $\hat{\beta}_t$ be a solution to $l_t(\beta) = 0$. If no solution exists, we define $\hat{\beta}_t = \beta^{(1)}$, for some predefined $\beta^{(1)} \in B$. If a solution to $l_t(\beta) = 0$ exists but lies outside B, we define $\hat{\beta}_t$ as the projection of this solution on B. For most choices of h there is no explicit formula for the solution of $l_t(\beta) = 0$, and numerical methods have to be deployed to calculate it.

5.3.2 Convergence rates of parameter estimates

Define the design matrix

$$P_t = \sum_{i=1}^{t} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (1, p_i), \quad (t \in \mathbb{N}). \tag{5.6}$$

The smallest eigenvalue of P_t, denoted by $\lambda_{\min}(P_t)$, is a measure for the amount of price dispersion in $p_1, \ldots, p_t$. The growth rate of $\lambda_{\min}(P_t)$ plays an important role in establishing bounds on the speed at which the parameter estimates $\hat{\beta}_t$ converge to the true value $\beta^{(0)}$. In particular, the higher the growth rate of $\lambda_{\min}(P_t)$, the faster $E[||\hat{\beta}_t - \beta^{(0)}||^2]$ converges to zero. This is formalized in the next proposition. To state the result, we define the last-time random variable

$$T_\rho = \sup \left\{ t \in \mathbb{N} \mid \text{there is no } \beta \in B \text{ with } \left|\left| \beta - \beta^{(0)} \right|\right| \leq \rho \text{ and } l_t(\beta) = 0 \right\}, \tag{5.7}$$

for $\rho > 0$.

Proposition 5.1. *Suppose L is a non-random function on $\mathbb{N}$ such that $\lambda_{min}(P_t) \geq L(t) > 0$ a.s., for all $t \geq t_0$ and some non-random $t_0 \in \mathbb{N}$, and such that $\inf_{t \geq t_0} L(t) t^{-\alpha} > 0$, for some $\alpha > 1/2$. Then there exists a $\rho_1 > 0$ such that for all $0 < \rho \leq \rho_1$ we have $T_\rho < \infty$ a.s., $E[T_\rho] < \infty$, and $E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t > T_\rho} \right] = O(\log(t)/L(t))$.*

This follows directly from Theorem 7.1, Theorem 7.2, and Remark 7.2 in Chapter 7.

5.4 Endogenous learning

Proposition 5.1 shows that the growth rate of $\lambda_{\min}(P_t)$ influences the speed at which the parameter estimates converge to the true value. The main result of this section is that $\lambda_{\min}(P_t)$ strictly increases if, during a selling season, prices are used that are close to that prescribed by $\pi^*_{\beta^{(0)}}$. This means that a continuous use of prices close to $\pi^*_{\beta^{(0)}}$ leads to a linear growth rate of $\lambda_{\min}(P_t)$, which by Proposition 5.1 implies that the parameter estimates converges very fast to the true value, in particular with rate $E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t>T_\rho}\right] = O\left(\log(t)/t\right)$.

This result can be interpreted as the system having an endogenous learning property: the unknown parameters are learned very fast when a policy close to the optimal policy is used.

Theorem 5.1. *Let $1 < C < S$ and $k \in \mathbb{N}$. There exist a constant $v_0 > 0$, depending on $\beta^{(0)}$, and an open neighborhood $\mathcal{U} \subset B$ containing $\beta^{(0)}$, such that, if*

$$p_{s+(k-1)S} = \pi^*_{\beta(s)}(c_{s+(k-1)S}, s)$$

for all $s = 1, \ldots, S$ and some sequence $\beta(1), \ldots, \beta(S) \in \mathcal{U}$, then

$$\lambda_{min}(P_{kS}) - \lambda_{min}(P_{(k-1)S}) \geq \frac{1}{8} v_0^2 (1 + p_h^2)^{-1},$$

and

$$\min_{1 \leq s, s' \leq S} |p_{s+(k-1)S} - p_{s'+(k-1)S}| \geq v_0/2.$$

In Theorem 5.1, the requirement $C < S$ is crucial. If $C > S$, then clearly $C - S$ items cannot be sold during the selling season. The selling of the remaining S items can be interpreted as that each item can be sold only in a single, dedicated period. There is then no interaction between individual items, and the pricing problem is equivalent to S repeated problems with a single item and a single selling period. This setting does not satisfy an endogenous learning property. On the contrary, the seller needs to actively experiment with selling prices in order to learn the unknown parameters. A pricing strategy for this setting is elaborated in Chapter 3.

The proof of Theorem 5.1 makes use of the following auxiliary lemmas. Lemma 5.1 shows that the assumptions we impose on $g_{a,\beta}(p)$ do not only hold for $a \in [0, r^*]$ and $\beta = \beta^{(0)}$, but also on an open neighborhood around $[0, r^*] \times \{\beta^{(0)}\}$. This result, which follows directly from the continuity assumptions on h, enables us in later proofs to apply the implicit function theorem. Lemma 5.2 considers the optimization problem underlying (5.3), and shows uniqueness, differentiability, and sensitivity properties. These results are applied in Lemma 5.3 to conclude that $(\pi^*_\beta)_{1 \leq c \leq C, 1 \leq s \leq S}$ is uniquely defined and continuous in β, on an open neighborhood around $\beta^{(0)}$. Lemma 5.4 relates price differences to the growth of $\lambda_{\min}(P_t)$ during a selling season.

Lemma 5.1. *There are open sets $U_a \subset \mathbb{R}$ containing $[0, r^*]$, and $U_B \subset B$ containing $\beta^{(0)}$, with*

$$\sup_{\beta \in U_B} \max_{p \in [p_l, p_h]} p \cdot h(\beta_0 + \beta_1 p) \in U_a, \tag{5.8}$$

and such that

$$g_{a,\beta}(p_l) < 1,\ g_{a,\beta}(p_h) > 1 \text{ and } g_{a,\beta}(p) \text{ strictly increasing in } p, \tag{5.9}$$

holds for all $(a, \beta) \in U_a \times U_B$.

Lemma 5.2. *Let U_a and U_B be as in Lemma 5.1, and for all $(a, \beta) \in U_a \times U_B$ define the function $f_{a,\beta}(p) = (p - a)h(\beta_0 + \beta_1 p)$. Write $\dot{f}_{a,\beta}(p)$ and $\ddot{f}_{a,\beta}(p)$ for the first and second derivative of $f_{a,\beta}(p)$*

with respect to p, and let $p^*_{a,\beta} = \underset{p\in[p_l,p_h]}{\arg\max}\, f_{a,\beta}(p)$. Then:

(i) $p^*_{a,\beta}$ is the unique solution to $\dot{f}_{a,\beta}(p) = 0$, lies in (p_l, p_h), and in addition satisfies $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$.

(ii) $p^*_{a,\beta}$ is continuously differentiable in a and β, strictly increasing in a, and $f_{a,\beta}(p^*_{a,\beta})$ is strictly decreasing in a.

(iii) There is a $K_0 > 0$ such that for all $(a,\beta) \in U_a \times U_B$ and $p \in [p_l, p_h]$,

$$f_{a,\beta}(p^*_{a,\beta}) - f_{a,\beta}(p) \le K_0(p - p^*_{a,\beta})^2.$$

Lemma 5.3. *Let U_B be as in Lemma 5.1. For each $\beta \in U_B$ and $(c,s) \in \mathcal{X}$ with $c > 0$, $\pi^*_\beta(c,s)$ is uniquely defined and continuous in β.*

Lemma 5.4. *Let $k \in \mathbb{N}$. If there are $s, s' \in \{1,\dots,S\}$ such that $|p_{s+(k-1)S} - p_{s'+(k-1)S}| \ge \delta$, then $\lambda_{min}(P_{kS}) \ge \lambda_{min}(P_{(k-1)S}) + \frac{1}{2}\delta^2(1+p_h^2)^{-1}$.*

5.5 Pricing strategy

We propose a pricing strategy based on the following principle: in each period, estimate the unknown parameters, and subsequently use the action from the policy that is optimal with respect to this estimate.

Pricing strategy $\Phi(\epsilon)$

Initialization: Choose $0 < \epsilon < (p_h - p_l)/4$, and initial prices $p_1, p_2 \in [p_l, p_h]$, with $p_1 \ne p_2$.
For all $t \ge 2$: if $c_{t+1} = 0$, set $p_{t+1} \in [p_l, p_h]$ arbitrary. If $c_{t+1} > 0$:
Estimation: Determine $\hat{\beta}_t$, and let $p_{\text{ceqp}} = \pi^*_{\hat{\beta}_t}(c_{t+1}, s_{t+1})$.
Pricing:
I) If

(a) $|p_i - p_j| < \epsilon$ for all $1 \le i, j \le t$ with $SS_i = SS_{t+1}$, and

(b) $|p_i - p_{\text{ceqp}}| < \epsilon$ for all $1 \le i \le t$ with $SS_i = SS_{t+1}$, and

(c) $c_{t+1} = 1$ or $s_{t+1} = S$,

then choose $p_{t+1} \in \big(\{p_{\text{ceqp}} + 2\epsilon, p_{\text{ceqp}} - 2\epsilon\} \cap [p_l, p_h]\big)$.
II) Else, set $p_{t+1} = p_{\text{ceqp}}$.

Given a positive inventory level, the pricing strategy $\Phi(\epsilon)$ sets the price p_{t+1} equal to the price that is optimal according to the available parameter estimates $\hat{\beta}_t$, except possibly when the state (c_{t+1}, s_{t+1}) is in the set $\{(c,s) \mid c = 1 \text{ or } s = S\}$. This set contains all states that, with positive probability, are the last states in the selling season in which products are sold (either because the selling season almost finishes, or because the inventory consists of only a single product). In these states, the price p_{t+1} deviates from the certainty equivalent price p_{ceqp} if otherwise $\max\{|p_i - p_j| \mid SS_i = SS_{t+1}\} < \epsilon$.

The endogenous learning property described in Section 5.4 implies that if $\hat{\beta}_t$ is sufficiently close to $\beta^{(0)}$ and ϵ is sufficiently small, then I) does not occur. As $\hat{\beta}_t$ converges to $\beta^{(0)}$, the pricing strategy $\Phi(\epsilon)$ eventually acts as a certainty equivalent pricing strategy. The pricing decisions in

II) are driven by optimizing instant revenue, and do not reckon with the objective of optimizing the quality of the parameter estimates $\hat{\beta}_t$. The endogenous learning property makes sure that learning the parameter values happens on the fly, without active effort.

As a result, the parameter estimates converge quickly to their true values, and the pricing decisions quickly to the optimal pricing decisions. The following theorem shows that the regret of the strategy $\Phi(\epsilon)$ is $O(\log^2(T))$ in the number of selling seasons T.

Theorem 5.2. *Let $1 < C < S$, v_0 as in Theorem 5.1, and assume $\epsilon < v_0/2$. Then*

$$\textit{Regret}(\Phi(\epsilon), T \cdot S) = O(\log^2(T)).$$

5.6 Numerical illustration

To illustrate the analytical results that we have derived, we provide a small numerical illustration. We let $C = 10$, $S = 20$, $p_l = 1$, $p_h = 20$, $\beta_0^{(0)} = 2$, $\beta_1^{(0)} = -0.4$, and $h(z) = \text{logit}(z)$. The optimal expected revenue per selling season, $V_{\beta^{(0)}}(C, 1)$, is equal to 47.79. We consider a time span of 100 selling periods. Figure 5.1 shows a sample path of $||\hat{\beta}_t - \beta^{(0)}||$ for $t = 1, \ldots, 100S$, Regret(T), and the relative regret $\frac{\text{Regret}(T)}{T \cdot V_{\beta^{(0)}}(C,1)} \times 100\%$, for $T = 1, \ldots, 100$.

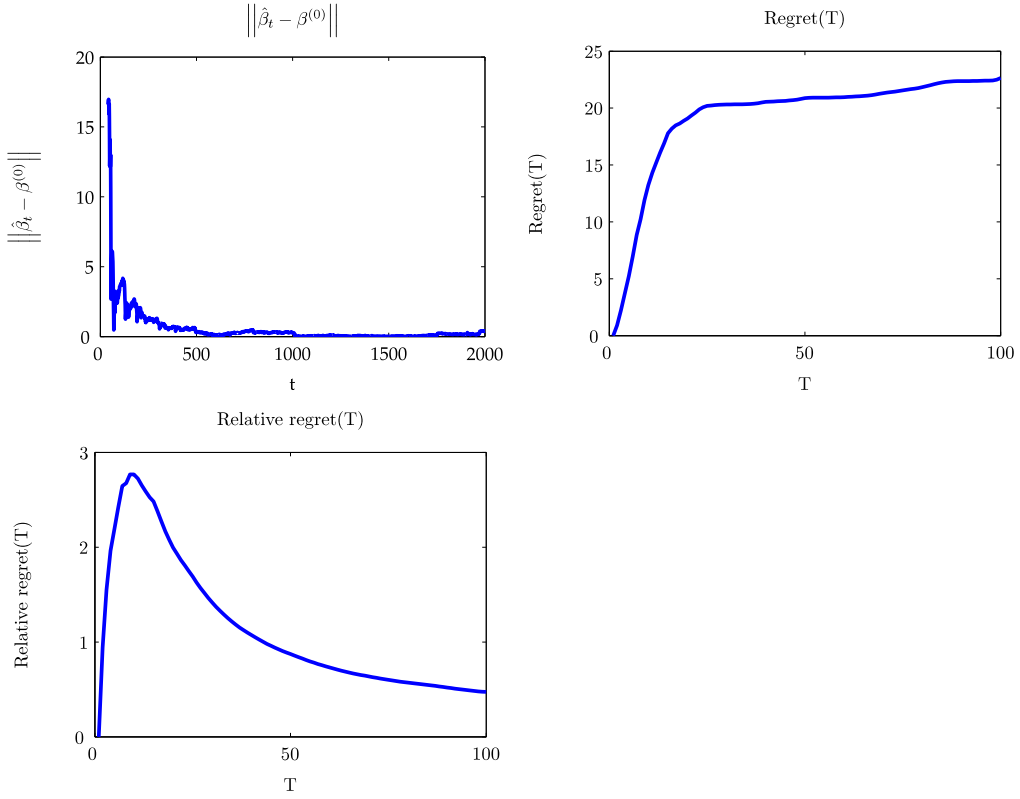

Figure 5.1: Sample path of estimation error, regret, and relative regret

5.7 Proofs

Proof of Theorem 5.1

Consider the k-th selling season, and write $c(1) = c_{1+(k-1)S}$, $c(2) = c_{2+(k-1)S}$, ..., $c(S) = c_{kS}$. First we show that there is a $v_0 > 0$ such that if prices $\pi^*_{\beta^{(0)}}(c(s), s)$ are used in state $(c(s), s)$, for all $s = 1, \ldots, S$, then there are $1 \le s, s' \le S$ with $|\pi^*_{\beta^{(0)}}(c(s), s) - \pi^*_{\beta^{(0)}}(c(s'), s')| > v_0$. Since π^*_β is continuous in β around $\beta^{(0)}$ (Lemma 5.3), this implies that there is an open neighborhood $\mathcal{U} \subset U_B$ around $\beta^{(0)}$ such that, if price $\pi^*_{\beta(s)}(c(s), s)$ is used in state $(c(s), s)$, for all $s = 1, \ldots, S$, then there are $1 \le s, s' \le S$ with $|\pi^*_{\beta(s)}(c(s), s) - \pi^*_{\beta(s')}(c(s'), s')| > v_0/2$.

Application of Lemma 5.4 with $K_1 = (1 + p_h^2)^{-1} \frac{v_0^2}{8}$ proves the theorem.

Define

$$\triangleleft = \{(c, s) \mid S + 1 - C \le s \le S, S + 1 - s \le c \le C\}. \tag{5.10}$$

See Figure 5.2 for an illustration of $\triangleleft$ in the state space $\mathcal{X}$. Notice that since $(C, 1) \notin \triangleleft$ (by the assumption $C < S$), the path $(c(s), s)_{1 \le s \le S}$ may or may not hit $\triangleleft$. We show that in both cases, at least two different selling prices occur on the path $(c(s), s)_{1 \le s \le S}$.

Case 1. The path $(c(s), s)_{1 \le s \le S}$ hits $\triangleleft$. Then there is an s such that $(c(s), s) \in \triangleleft$ and $(c(s), s-1) \notin \triangleleft$. In particular, $(c(s), s-1) \in (L\triangleleft) = \{(1, S-1), (2, S-2), \ldots, (C-1, S-C+1), (C, S-C)\}$, where $(L\triangleleft)$ denotes the points (c, s) immediately left to $\triangleleft$ in Figure 5.2. The sets $\triangleleft$ and $(L\triangleleft)$ satisfy the following properties:

(P.1) If $(c, s) \in \triangleleft$ then $\Delta V_{\beta^{(0)}}(c, s+1) = 0$, $\pi^*_{\beta^{(0)}}(c, s) = \arg\max_{p \in [p_l, p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)} p)$, and $V_{\beta^{(0)}}(c, s) = (S - s + 1) \cdot V_{\beta^{(0)}}(1, S)$.

(P.2) If $(c, s) \in (L\triangleleft)$, then $\pi^*_{\beta^{(0)}}(c, s) \ne \pi^*_{\beta^{(0)}}(c, s+1)$ and $\Delta V_{\beta^{(0)}}(c+1, S-c) \ne 0$ (provided $c < C$).

Proof of (P.1): Backward induction on s. If $s = S$ and $(c, s) \in \triangleleft$, then the assertions follow immediately. Let $s < S$. Then $\Delta V_{\beta^{(0)}}(c, s+1) = V_{\beta^{(0)}}(c, s+1) - V_{\beta^{(0)}}(c-1, s+1) = 0$, $\pi^*_{\beta^{(0)}}(c, s) = \arg\max_{p \in [p_l, p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)} p)$ and $V_{\beta^{(0)}}(c, s) = \max_{p \in [p_l, p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(c, s+1) = (S - s + 1) \cdot V_{\beta^{(0)}}(1, S)$, by (5.3) and the induction hypothesis. This proves (P.1).

Proof of (P.2). Induction on c. If $c = 1$ and $(c, s) \in (L\triangleleft)$, then $(c, s) = (1, S-1)$. Since $\Delta V_{\beta^{(0)}}(1, S) = V_{\beta^{(0)}}(1, S) > 0$, Lemma 5.2 and equation (5.3) imply $\pi^*_{\beta^{(0)}}(1, S-1) \ne \pi^*_{\beta^{(0)}}(1, S)$. In addition,

$$V_{\beta^{(0)}}(2, S-1) = \max_{p \in [p_l, p_h]} \left((p - \Delta V_{\beta^{(0)}}(2, S)) h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(2, S) \right), \tag{5.11}$$

$$V_{\beta^{(0)}}(1, S-1) = \max_{p \in [p_l, p_h]} \left((p - \Delta V_{\beta^{(0)}}(1, S)) h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(1, S) \right). \tag{5.12}$$

Property $(P.1)$ implies $V_{\beta^{(0)}}(2, S) = V_{\beta^{(0)}}(1, S)$ and $\Delta V_{\beta^{(0)}}(2, S) = 0$. Furthermore, $\Delta V_{\beta^{(0)}}(1, S) = V_{\beta^{(0)}}(1, S) > 0$, and thus by Lemma 5.2, $\Delta V_{\beta^{(0)}}(2, S-1) = V_{\beta^{(0)}}(2, S-1) - V_{\beta^{(0)}}(1, S-1) \ne 0$.

Let $c > 1$ and $(c, s) \in (L\triangleleft)$. Then $(c, s) = (c, S - c)$. By the induction hypothesis we have

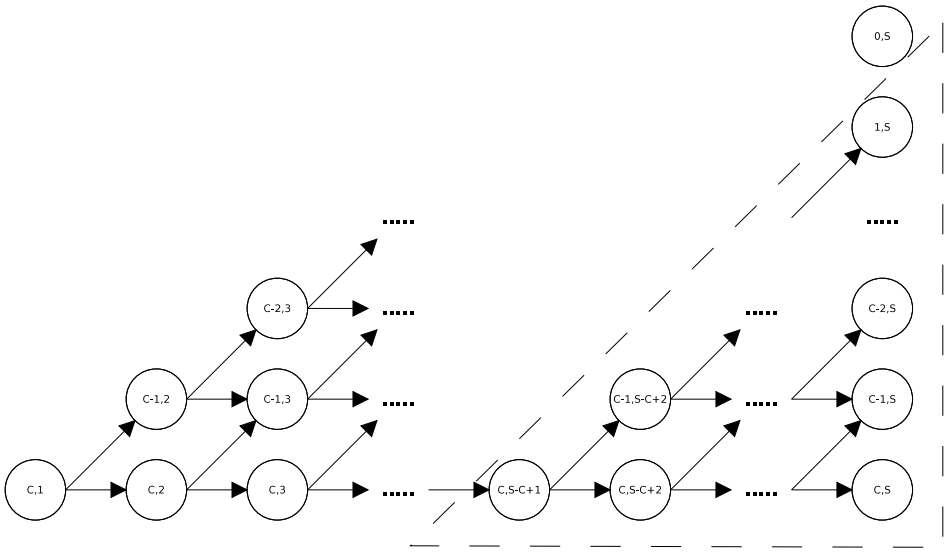

Figure 5.2: Schematic picture of $\triangleleft$

$\Delta V_{\beta^{(0)}}(c, S-c+1) \neq 0$, and thus

$$\pi^*_{\beta^{(0)}}(c, S-c) = \underset{p\in[p_l,p_h]}{\arg\max} \big(p - \Delta V_{\beta^{(0)}}(c, S-c+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) \tag{5.13}$$

$$\neq \underset{p\in[p_l,p_h]}{\arg\max}\, ph(\beta_0^{(0)} + \beta_1^{(0)} p) = \pi^*_{\beta^{(0)}}(c, S-c+1), \tag{5.14}$$

where we used Lemma 5.2 for the first inequality, and (P.1) for the second equality. It remains to show $\Delta V_{\beta^{(0)}}(c+1, S-c) \neq 0$, when $c < C$. Note that

$$\begin{aligned} V_{\beta^{(0)}}(c+1, S-c) = &\max_{p\in[p_l,p_h]} (p - \Delta V_{\beta^{(0)}}(c+1, S-c+1)) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) \\ &+ V_{\beta^{(0)}}(c+1, S-c+1), \\ V_{\beta^{(0)}}(c, S-c) = &\max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(c, S-c+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(c, S-c+1). \end{aligned}$$

Since $(c+1, S-c+1) \in \triangleleft$ and $(c, S-c+1) \in \triangleleft$, (P.1) implies $V_{\beta^{(0)}}(c+1, S-c+1) = V_{\beta^{(0)}}(c, S-c+1)$. In addition, $c < C$ implies $(c+1, S-c) \in \triangleleft$, and thus $\Delta V_{\beta^{(0)}}(c+1, S-c+1) = 0$ by (P.1). The induction hypothesis implies $\Delta V_{\beta^{(0)}}(c, S-c+1) \neq 0$. Then Lemma 5.2 implies $V_{\beta^{(0)}}(c+1, S-c) \neq V_{\beta^{(0)}}(c, S-c)$. This proves (P.2), and shows that a price-change occurs when $\triangleleft$ is entered.

This concludes case 1.

Case 2. The path $(c(s), s)_{1\leq s\leq S}$ does not hit $\triangleleft$. Then there is an s such that $c(s) = 2$ and $c(s+1) = 1$. We show $\pi^*_{\beta^{(0)}}(2, s) \neq \pi^*_{\beta^{(0)}}(1, s+1)$, for all $1 \leq s \leq S-2$.

$$\pi^*_{\beta^{(0)}}(2, s) = \underset{p\in[p_l,p_h]}{\arg\max} \big(p - \Delta V_{\beta^{(0)}}(2, s+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p), \tag{5.15}$$

$$\pi^*_{\beta^{(0)}}(1, s+1) = \underset{p\in[p_l,p_h]}{\arg\max} \big(p - \Delta V_{\beta^{(0)}}(1, s+2)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p), \tag{5.16}$$

By Lemma 5.2, and the fact that $\pi^*_{\beta^{(0)}}(2, s)$ and $\pi^*_{\beta^{(0)}}(1, s+1)$ are both contained in (p_l, p_h), it suffices to show $\Delta V_{\beta^{(0)}}(2, s+1) \neq \Delta V_{\beta^{(0)}}(1, s+2)$. We show by backward induction that

$V_{\beta^{(0)}}(2,s) - V_{\beta^{(0)}}(1,s) \neq V_{\beta^{(0)}}(1,s+1)$ for all $2 \leq s \leq S-1$. Let $2 \leq s \leq S-1$.

$$V_{\beta^{(0)}}(2,s) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(2,s+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(2,s+1), \tag{5.17}$$

$$V_{\beta^{(0)}}(1,s) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(1,s+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(1,s+1), \tag{5.18}$$

$$V_{\beta^{(0)}}(1,s+1) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(1,s+2)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(1,s+2). \tag{5.19}$$

Using

$$V_{\beta^{(0)}}(1,s+1) \geq \Big[(\pi^*_{\beta^{(0)}}(2,s) - \Delta V_{\beta^{(0)}}(1,s+2))h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(2,s)) + V_{\beta^{(0)}}(1,s+2)\Big],$$

we have

$$\begin{aligned}
&V_{\beta^{(0)}}(2,s) - V_{\beta^{(0)}}(1,s) - V_{\beta^{(0)}}(1,s+1)\\
\leq&(\pi^*_{\beta^{(0)}}(2,s) - \Delta V_{\beta^{(0)}}(2,s+1))h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(2,s)) + V_{\beta^{(0)}}(2,s+1)\\
&-\Big[(\pi^*_{\beta^{(0)}}(1,s) - \Delta V_{\beta^{(0)}}(1,s+1))h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(1,s)) + V_{\beta^{(0)}}(1,s+1)\Big]\\
&-\Big[(\pi^*_{\beta^{(0)}}(2,s) - \Delta V_{\beta^{(0)}}(1,s+2))h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(2,s)) + V_{\beta^{(0)}}(1,s+2)\Big]\\
=&-\pi^*_{\beta^{(0)}}(1,s)h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(1,s))\\
&+\Big[V_{\beta^{(0)}}(2,s+1) - V_{\beta^{(0)}}(1,s+1) - V_{\beta^{(0)}}(1,s+2)\Big]\Big[1 - h(\beta_0^{(0)} + \beta_1^{(0)}(\pi^*_{\beta^{(0)}}(2,s))\Big]\\
&+V_{\beta^{(0)}}(1,s+1)h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(1,s))\\
\leq&-\pi^*_{\beta^{(0)}}(1,s)h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(1,s)) + V_{\beta^{(0)}}(1,s+1)h(\beta_0^{(0)} + \beta_1^{(0)}\pi^*_{\beta^{(0)}}(1,s))\\
=&V_{\beta^{(0)}}(1,s+1) - V_{\beta^{(0)}}(1,s).
\end{aligned}$$

The last inequality is implied by $V_{\beta^{(0)}}(2,s+1) - V_{\beta^{(0)}}(1,s+1) - V_{\beta^{(0)}}(1,s+2) \leq 0$, which for $s = S-1$ follows from (P.1), and for $s < S-1$ follows from the induction hypothesis. The proof of Lemma 5.3 implies $p_h - V_{\beta^{(0)}}(1,s+1) > 0$, and thus $V_{\beta^{(0)}}(1,s) \geq (p_h - V_{\beta^{(0)}}(1,s+1)) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p_h) + V_{\beta^{(0)}}(1,s+1) > V_{\beta^{(0)}}(1,s+1)$. This proves $V_{\beta^{(0)}}(2,s) - V_{\beta^{(0)}}(1,s) - V_{\beta^{(0)}}(1,s+1) < 0$.

This concludes case 2.

We have shown that in case 1, $\pi^*_{\beta^{(0)}}(c, S-c) \neq \pi^*_{\beta^{(0)}}(c, S-c+1)$, and in case 2, $\pi^*_{\beta^{(0)}}(2,s) \neq \pi^*_{\beta^{(0)}}(1,s+1)$. This implies that on any path $(c(s),s)_{1\leq s\leq S}$ in $\mathcal{X}$, starting at $(C,1)$, the policy $\pi^*_{\beta^{(0)}}$ induces a price-change somewhere along the path $(c(s),s)_{1\leq s\leq S}$. Since the number of possible paths is finite, it follows that there exists a $v_0 > 0$ such that for all paths $(c(s),s)_{1\leq s\leq S}$,

$$|\pi^*_{\beta^{(0)}}(c(s),s) - \pi^*_{\beta^{(0)}}(c(s'),s')| \geq v_0.$$

Proof of Theorem 5.2

Consider the k-th selling season, for some arbitrary fixed $k \in \mathbb{N}$. The prices generated by $\Phi(\epsilon)$ are based on the estimates $\hat{\beta}_t$, which are determined by the historical prices and demand realizations. Now, different demand realizations can lead to the same state (c,s) of the MDP. For example, a sale in the first period of a selling season and no sale in the second period leads to state $(C-2,3)$, but this state is also reached if there is no sale in the first period and a sale in the second period of the selling season. These two "routes" may lead to different estimates $\hat{\beta}_t$, and to different

pricing decisions in state $(C-2,3)$. Thus, with $\Phi(\epsilon)$, the prices in the k-th selling season are not determined by a stationary policy for the Markov decision problem described in Section 5.2.3.

To be able to compare the optimal revenue in a selling season with that obtained by $\Phi(\epsilon)$, we define a new Markov decision problem, in which the states are sequences of demand realizations in the selling season. Conditional on all prices and demand realizations from before the selling season, the policy $\Phi(\epsilon)$ is then a stationary policy for this new MDP: each state is associated with a unique price prescribed by $\Phi(\epsilon)$. This enables us to calculate bounds on the regret obtained in a single selling season.

We define this new MDP for any $\beta \in B$. The state space $\tilde{\mathcal{X}}$ consists of all sequences of possible demand realizations in the selling season:

$$\tilde{\mathcal{X}} = \{(x_1, \dots, x_s) \in \{0,1\}^s \mid 0 \leq s \leq S\},$$

where we denote the empty sequence by $(\emptyset)$. The action space is $[p_l, p_h]$. Using action p in state $(x_1, \dots, x_s)$, for $0 \leq s < S$, induces a state transition from $(x_1, \dots x_s)$ to $(x_1, \dots, x_s, 1)$ with probability $h(\beta_0 + \beta_1 p)$ (corresponding to a sale, and inducing immediate reward $ph(\beta_0 + \beta_1 p)\mathbf{1}_{\sum_{i=1}^s x_i < C}$), and from $(x_1, \dots x_s)$ to $(x_1, \dots, x_s, 0)$ with probability $1 - h(\beta_0 + \beta_1 p)$ (corresponding to no sale, and inducing zero reward). There are no state transitions in the terminal states $(x_1, \dots, x_S) \in \tilde{\mathcal{X}}$.

It is easily seen that the MDP described in section 5.2.3 is the same as the one described here, except that there states are aggregated: all states $(x_1, \dots, x_s)$ and $(x'_1, \dots, x'_{s'})$ with $s = s'$ and $\sum_{i=1}^s x_i = \sum_{i=1}^{s'} x'_i$ are there taken together.

Let $\tilde{\pi} = (\tilde{\pi}(x))_{x \in \tilde{\mathcal{X}}}$ be a stationary deterministic policy for this MDP with augmented state space, and let $\tilde{V}_\beta^{\tilde{\pi}}(x)$ be the corresponding value function, for $\beta \in B$. For $x = (x_1, \dots, x_s) \in \tilde{\mathcal{X}}$ with $s < S$ we write $(x;1) = (x_1, \dots, x_s, 1)$ and $(x;0) = (x_1, \dots, x_s, 0)$. Then, for any $x = (x_1, \dots, x_s) \in \tilde{\mathcal{X}}$ and $\beta \in B$, $\tilde{V}_\beta^{\tilde{\pi}}(x)$ satisfies the backward recursion

$$\tilde{V}_\beta^{\tilde{\pi}}(x) = (\tilde{\pi}(x)\mathbf{1}_{\sum_{i=1}^s x_i < C} + \tilde{V}_\beta^{\tilde{\pi}}(x;1))h(\beta_0 + \beta_1 \tilde{\pi}(x)) + \tilde{V}_\beta^{\tilde{\pi}}(x;0)(1 - h(\beta_0 + \beta_1 \tilde{\pi}(x))),$$

where we write $\tilde{V}_\beta^{\tilde{\pi}}(x;1) = \tilde{V}_\beta^{\tilde{\pi}}(x;0) = 0$ for all terminal states $(x_1, \dots, x_S) \in \tilde{\mathcal{X}}$.

Let $\tilde{\pi}_\beta^*$ be the optimal policy corresponding to $\beta \in B$, and write $\tilde{V}_\beta(x) = \tilde{V}_\beta^{\tilde{\pi}_\beta^*}(x)$. Then

$$\tilde{V}_\beta(x) = \max_{p \in [p_l, p_h]} \left[p\mathbf{1}_{\sum_{i=1}^s x_i < C} - \left(\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1)\right) \right] h(\beta_0 + \beta_1 p) + \tilde{V}_\beta(x;0), \tag{5.20}$$

$$\tilde{\pi}_\beta^*(x) = \arg\max_{p \in [p_l, p_h]} \left[p\mathbf{1}_{\sum_{i=1}^s x_i < C} - \left(\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1)\right) \right] h(\beta_0 + \beta_1 p). \tag{5.21}$$

Using the same line of reasoning as Lemma 5.2 and 5.3, it can easily be shown that $\tilde{\pi}_\beta^*((x_1, \dots, x_s))$ is unique if and only if $\sum_{i=1}^s x_i < C$. For all x with $\sum_{i=1}^s x_i \geq C$, choose $\tilde{\pi}_\beta^*(x) = p_h$. In this way $\tilde{\pi}_\beta^*(x)$ is uniquely defined for all $x \in \tilde{\mathcal{X}}$.

Let $\mathcal{U}$ and v_0 be as in Theorem 5.1, ρ_1 as in Proposition 5.1, and choose $\rho \in (0, \rho_1)$ such that $\beta \in \mathcal{U}$ whenever $||\beta - \beta^{(0)}|| \leq \rho$.

If $(k-1)S > T_\rho$, then $\hat{\beta}_t \in \mathcal{U}$ for all $t = 1 + (k-1)S, \dots, S(k-1)S$, and Theorem 5.1 implies $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) \geq \frac{1}{8} v_0^2 (1 + p_h^2)^{-1}$. If $(k-1)S \leq T_\rho$, then I) of the pricing strategy $\Phi(\epsilon)$ guarantees that there are $1 \leq s, s' \leq S$ such that $|p_{s+(k-1)S} - p_{s'+(k-1)S}| \geq \epsilon$. By Lemma 5.4 this implies $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) \geq \frac{1}{2}\epsilon^2 (1 + p_h^2)^{-1}$. Since $\epsilon^2 \leq v_0^2/4$, this means that

$\lambda_{\min}(P_{kS}) \geq k \cdot \frac{1}{2}\epsilon^2(1+p_h^2)^{-1}$ for all $k \in \mathbb{N}$, and thus for all $t > S$,

$$\lambda_{\min}(P_t) \geq \lambda_{\min}(P_{(SS_t-1)S}) \geq (SS_t - 1) \cdot \frac{1}{2}\epsilon^2(1+p_h^2)^{-1} \geq t \cdot \frac{1}{4S}\epsilon^2(1+p_h^2)^{-1},$$

using $SS_t - 1 \geq t\frac{(SS_t-1)}{S \cdot SS_t} \geq \frac{t}{2S}$. By application of Proposition 5.1 with $t_0 = S$ and $L(t) = t \cdot \frac{1}{4S}\epsilon^2(1+p_h^2)^{-1}$, we have $T_\rho < \infty$ a.s., $E_{\Phi(\epsilon)}[T_\rho] < \infty$, and $E_{\Phi(\epsilon)}[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t>T_\rho}] = O(\log(t)/t)$.

In addition, $v_0/2 > \epsilon$ implies that I) of the pricing strategy $\Phi(\epsilon)$ does not occur for all t with $(SS_t - 1)S > T_\rho$. In particular, if $(k-1)S > T_\rho$, then

$$p_{1+s+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{s+(k-1)S}}(d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S}), \tag{5.22}$$

for all $1 \leq s \leq S-1$, and

$$p_{1+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{(k-1)S}}(\emptyset). \tag{5.23}$$

Let $H = (p_1, \ldots, p_{(k-1)S}, d_1, \ldots, d_{(k-1)S})$ denote the history of prices and demand up to and including time period $(k-1)S$. Conditional on H, and given that $(k-1)S > T_\rho$, the parameter estimates $\hat{\beta}_{s+(k-1)S}$ in (5.22) and (5.23) are completely determined by the state $(d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S})$. Thus, for each state $x \in \tilde{\mathcal{X}}$ there is a uniquely associated price prescribed by $\Phi(\epsilon)$. Consequently, there is a stationary deterministic policy, denoted by $\tilde{\pi}^H$, such that

$$\begin{aligned} p_{1+s+(k-1)S} &= \tilde{\pi}^H(x), \quad \text{when } x = (d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S}),\ 1 \leq s \leq S-1, \\ p_{1+(k-1)S} &= \tilde{\pi}^H(\emptyset). \end{aligned}$$

This enables us to bound the regret in the k-th selling season:

$$\begin{aligned} & V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i \min\{d_i, c_i\}] \\ =& E_{\Phi(\epsilon)}\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right)\mathbf{1}_{(k-1)S \leq T_\rho}\right] \\ &+ E_{\Phi(\epsilon)}\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right)\mathbf{1}_{(k-1)S > T_\rho}\right] \\ \leq& \tilde{V}_{\beta^{(0)}}(\emptyset) P((k-1)S \leq T_\rho) \\ &+ E_{\Phi(\epsilon)}\left[E_{\Phi(\epsilon)}\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right)\mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right] \\ \leq& \tilde{V}_{\beta^{(0)}}(\emptyset)\frac{E_{\Phi(\epsilon)}[T_\rho]}{(k-1)S} \qquad (5.24) \\ &+ E_{\Phi(\epsilon)}\left[E_{\Phi(\epsilon)}\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \tilde{V}^{\tilde{\pi}^H}_{\beta^{(0)}}(\emptyset)\right)\mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right]. \qquad (5.25) \end{aligned}$$

The term (5.24) is finite because $E[T_\rho] < \infty$. To obtain an upper bound on the term (5.25), we need a number of sensitivity results:

(S.0) For all $\beta \in U_B$ and x such that $(x;0),(x;1) \in \tilde{\mathcal{X}}$, we have

$$0 \leq \tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \leq \max_{p\in[p_l,p_h]} p \cdot h(\beta_0 + \beta_1 p). \tag{5.26}$$

(S.1) Write $Y_s = (d_{1+(k-1)S}, \ldots, d_{s+(k-1)S})$ for $1 \leq s \leq S-1$, and $Y_0 = (\emptyset)$. Let K_0 be as in Lemma 5.3(iii). Then for all stationary deterministic policies $\tilde{\pi}$ and all $0 \leq s \leq S-1$,

$$(\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s))\mathbf{1}_{(k-1)S>T_\rho} \leq K_0 \sum_{\sigma=s}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}(Y_\sigma))^2 \mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.} \tag{5.27}$$

(S.2) There is a $K_3 > 0$ such that for all β with $||\beta - \beta^{(0)}|| \leq \rho$, and all $x \in \tilde{\mathcal{X}}$,

$$|\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x)| \leq K_3 ||\beta - \beta^{(0)}||. \tag{5.28}$$

The proof of these three sensitivity properties is given below.

Application of (S.1), (S.2), and Proposition 5.1 now gives

$$\begin{aligned}
&E_{\Phi(\epsilon)}\left[E_{\Phi(\epsilon)}\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \tilde{V}^{\tilde{\pi}^H}_{\beta^{(0)}}(\emptyset)\right)\mathbf{1}_{(k-1)S>T_\rho} \mid H\right]\right] \\
\leq &E_{\Phi(\epsilon)}\left[E_{\Phi(\epsilon)}\left[K_0 \sum_{\sigma=0}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}^H(Y_\sigma))^2 \mathbf{1}_{(k-1)S>T_\rho} \mid H\right]\right] \\
= &E_{\Phi(\epsilon)}\left[K_0 \sum_{\sigma=0}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}^*_{\hat{\beta}_{\sigma+(k-1)S}}(Y_\sigma))^2 \mathbf{1}_{(k-1)S>T_\rho}\right] \\
\leq &E_{\Phi(\epsilon)}\left[K_0 K_3^2 \sum_{\sigma=0}^{S-1} \left|\left|\beta^{(0)} - \hat{\beta}_{\sigma+(k-1)S}\right|\right|^2 \mathbf{1}_{(k-1)S>T_\rho}\right] \\
\leq &K_4 \sum_{\sigma=0}^{S-1} \frac{\log(\sigma+(k-1)S)}{\sigma+(k-1)S},
\end{aligned}$$

for some K_4 independent of k and S.

We then have

$$\begin{aligned}
&V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E_{\Phi(\epsilon)}[p_i \min\{d_i, c_i\}] \\
\leq &\tilde{V}_{\beta^{(0)}}(\emptyset) E_{\Phi(\epsilon)}[T_\rho] \frac{1}{(k-1)S} + K_4 \sum_{\sigma=0}^{S-1} \frac{\log(\sigma+(k-1)S)}{\sigma+(k-1)S} \\
\leq &K_5 \sum_{t=1+(k-1)S}^{kS} \frac{\log(t)}{t},
\end{aligned}$$

for some $K_5 > 0$, independent of k and S.

The proof of the theorem is complete by observing

$$\begin{aligned}\text{Regret}(\Phi(\epsilon), T\cdot S) &= \sum_{k=1}^{T}\Big[V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E_{\Phi(\epsilon)}[p_i \min\{d_i, c_i\}]\Big]\\ &\le \sum_{k=1}^{T} K_5 \sum_{t=1+(k-1)S}^{kS} \frac{\log(t)}{t} = K_5 \sum_{t=1}^{TS}\frac{\log(t)}{t}\\ &= O(\log^2(T)).\end{aligned}$$

Proof of (S.0)

We prove the assertion for all $(x_1,\ldots,x_{s-1}) \in \tilde{\mathcal{X}}$, $s = 1,\ldots,S$, by backward induction on s. If $x = (x_1,\ldots,x_{S-1}) \in \tilde{\mathcal{X}}$ then $\tilde{V}_\beta(x;0) = \tilde{V}_\beta(x;1) = 0$.

Let $x \in \mathcal{X}$. If $\sum_{i=1}^{s} x_i \ge C$ then $\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) = 0$. If $\sum_{i=1}^{s} x_i < C$ then the induction hypothesis implies

$$\begin{aligned}\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) &= \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0)\\ &\quad - \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0)\\ &\ge \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;1)) + \tilde{V}_\beta(x;0;0)\\ &\quad - \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0)\\ &= \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)(1 - h(\beta_0 + \beta_1\pi^*_\beta(x;1)))\\ &\quad + \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)h(\beta_0 + \beta_1\pi^*_\beta(x;1))\\ &\ge 0,\end{aligned}$$

and

$$\begin{aligned}\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) &= \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0)\\ &\quad - \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0)\\ &\le \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0)\\ &\quad - \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big] h(\beta_0 + \beta_1\pi^*_\beta(x;0)) - \tilde{V}_\beta(x;1;0)\\ &= \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)(1 - h(\beta_0 + \beta_1\pi^*_\beta(x;0)))\\ &\quad + \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)h(\beta_0 + \beta_1\pi^*_\beta(x;0))\\ &\le \max_{p\in[p_l,p_h]} p\cdot h(\beta_0 + \beta_1 p).\end{aligned}$$

Proof of (S.1)

Backward induction on s. If $s = S-1$ then Lemma 5.3(iii) implies

$$\begin{aligned}&\tilde{V}_{\beta^{(0)}}(Y_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{S-1})\\ =&\max_{p\in[p_l,p_h]} p\mathbf{1}_{\sum_{i=1}^{S-1} Y_i < C} h(\beta_0^{(0)} + \beta_1^{(0)} p) - \tilde{\pi}(Y_{S-1})\mathbf{1}_{\sum_{i=1}^{S-1} Y_i < C} h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_{S-1}))\\ \le& K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_{S-1}))^2 \text{ a.s.},\end{aligned}$$

and thus

$$(\tilde{V}_{\beta^{(0)}}(Y_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{S-1})) \cdot \mathbf{1}_{(k-1)S>T_\rho} \leq K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_{S-1}))^2 \cdot \mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.}$$

If $0 \leq s < S-1$, then

$$\begin{aligned}
&\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s) \\
=&\max_{p\in[p_l,p_h]} \big[p\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}p) \\
&-\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_s)) \\
&+\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_s)) \\
&-\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_s)) \\
&+\tilde{V}(Y_s;0) - \tilde{V}^{\tilde{\pi}}(Y_s;0) \\
\leq&K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_s))^2 \\
&+\big(\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;0)\big) \cdot \big(1 - h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_s))\big) \\
&+\big(\tilde{V}_{\beta^{(0)}}(Y_s;1) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;1)\big) \cdot (h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(Y_s)) \\
=&K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_s))^2 + \big[\tilde{V}_{\beta^{(0)}}(Y_{s+1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{s+1})\big] \text{ a.s.}
\end{aligned}$$

Here the first inequality follows from Lemma 5.3(iii), observing that (S.0) implies $\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1) \in U_a$. The induction hypothesis now implies

$$(\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s))\mathbf{1}_{(k-1)S>T_\rho} \leq K_0 \sum_{\sigma=s}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}(Y_\sigma))^2 \mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.}$$

Proof of (S.2)

If $\sum_{i=1}^s x_i \geq C$ then $\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x) = p_h - p_h = 0$. If $\sum_{i=1}^s x_i < C$, then

$$\begin{aligned}
\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x) &= p^*_{\tilde{V}_\beta(x;0)-\tilde{V}_\beta(x;1),\beta} - p^*_{\tilde{V}_{\beta^{(0)}}(x;0)-\tilde{V}_{\beta^{(0)}}(x;1),\beta^{(0)}} \\
&= p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}},
\end{aligned} \tag{5.29}$$

in the notation of Lemma 5.2, with $a = \tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1)$ and $a^{(0)} = \tilde{V}_{\beta^{(0)}}(x;0) - \tilde{V}_{\beta^{(0)}}(x;1)$.

By (S.0) we have $a, a^{(0)} \in U_a$ and $\beta \in U_B$, and thus by Lemma 5.2, $\tilde{\pi}^*_\beta(x)$ is continuously differentiable. The set $\{\beta \in B \mid ||\beta - \beta^{(0)}|| \leq \rho\}$ is compact, and so is the set $\{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \mid ||\beta - \beta^{(0)}|| \leq \rho, \beta \in B\}$. As a result, the derivative of $p^*_{a,\beta}$ w.r.t. (a,β) is bounded on the set $(a,\beta) \in \{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \mid ||\beta - \beta^{(0)}|| \leq \rho, \beta \in B\} \times \{\beta \in B \mid ||\beta - \beta^{(0)}|| \leq \rho\}$. It follows by a first-order Taylor expansion that there is a $K_6 > 0$ such that for all such (a,β),

$$|p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}}| \leq K_6(|a - a^{(0)}| + ||\beta - \beta^{(0)}||). \tag{5.30}$$

It is not difficult to show by backward induction that for all $x \in \tilde{\mathcal{X}}$ there is a $K_x > 0$ such that, for all β with $||\beta - \beta^{(0)}|| \leq \rho$,

$$\left|\tilde{V}_\beta(x) - \tilde{V}_{\beta^{(0)}}(x)\right| \leq K_x \left|\left|\beta - \beta^{(0)}\right|\right|. \tag{5.31}$$

Combining (5.29), (5.30), and (5.31), we obtain

$$\begin{aligned}
&|\tilde{\pi}^*_{\beta}(x) - \tilde{\pi}^*_{\beta^{(0)}}(x)| \\
\leq& K_6(|a - a^{(0)}| + ||\beta - \beta^{(0)}||) \\
\leq& K_6(|\tilde{V}_{\beta}(x;0) - \tilde{V}_{\beta^{(0)}}(x;0)| + |\tilde{V}_{\beta}(x;1) - \tilde{V}_{\beta^{(0)}}(x;1)| + ||\beta - \beta^{(0)}||) \\
\leq& K_6(1 + 2\max_{x\in\tilde{\mathcal{X}}} K_x)||\beta - \beta^{(0)}||.
\end{aligned}$$

This proves (S.2).

5.8 Proofs of auxiliary lemmas

Proof of Lemma 5.2

Since

$$\dot{f}_{a,\beta}(p) = h(\beta_0 + \beta_1 p)\left[1 + (p-a)\beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)}\right] = h(\beta_0 + \beta_1 p)\left[1 - g_{a,\beta}(p)\right],$$

and $h(\beta_0 + \beta_1 p) > 0$ for all $\beta \in U_B$, $p \in [p_l, p_h]$, we have $\dot{f}_{a,\beta}(p) = 0$ if and only if $g_{a,\beta}(p) = 1$. By Lemma 5.1, for all $(a,\beta) \in U_a \times U_B$ there is a unique $p^*_{a,\beta} \in (p_l, p_h)$ such that $g_{a,\beta}(p^*_{a,\beta}) = 1$. From

$$\begin{aligned}
\ddot{f}_{a,\beta}(p) &= \frac{\partial}{\partial p}\Big[h(\beta_0 + \beta_1 p)(1 - g_{a,\beta}(p))\Big] \\
&= \beta_1 \dot{h}(\beta_0 + \beta_1 p)(1 - g_{a,\beta}(p)) - h(\beta_0 + \beta_1 p)\frac{\partial}{\partial p} g_{a,\beta}(p)
\end{aligned}$$

follows

$$\ddot{f}_{a,\beta}(p^*_{a,\beta}) = -h(\beta_0 + \beta_1 p^*_{a,\beta})\frac{\partial}{\partial p} g_{a,\beta}(p^*_{a,\beta}) < 0,$$

since by Lemma 5.1, $g_{a,\beta}$ is strictly increasing in p. This proves (i).

For all $(a,\beta) \in U_a \times U_B$, $p^*_{a,\beta}$ is the unique solution in (p_l, p_h) to $g_{a,\beta}(p) - 1 = 0$, and

$$\frac{\partial g_{a,\beta}(p)}{\partial p}\Big|_{p=p^*_{a,\beta}} > 0.$$

The implicit function theorem (see e.g. Duistermaat and Kolk, 2004) then implies that $p^*_{a,\beta}$ is continuously differentiable at every $(a,\beta) \in U_a \times U_B$.

Furthermore, for all $(a,\beta) \in U_a \times U_B$ and $p \in [p_l, p_h]$ we have

$$\frac{\partial g_{a,\beta}(p)}{\partial a} = \beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)} < 0.$$

This implies that for all $a \in U_a$, $a' \in U_a$, with $a < a'$, and all $p \in [p_l, p_h]$ with $p \leq p^*_{a,\beta}$, we have $g_{a',\beta}(p) \leq g_{a,\beta}(p) \leq 1$. Therefore $p^*_{a',\beta} > p^*_{a,\beta}$ for all $a < a'$, and thus $p^*_{a,\beta}$ is strictly monotone increasing in a.

Using $g_{a,\beta}(p^*_{a,\beta}) = 1$ and thus $(p^*_{a,\beta} - a) = (-\beta_1^{-1})\frac{h(\beta_0+\beta_1 p^*_{a,\beta})}{\dot{h}(\beta_0+\beta_1 p^*_{a,\beta})}$, we have

$$f_{a,\beta}(p^*_{a,\beta}) = (p^*_{a,\beta} - a)h(\beta_0 + \beta_1 p^*_{a,\beta}) = (-\beta_1^{-1})\frac{h(\beta_0 + \beta_1 p^*_{a,\beta})^2}{\dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})},$$

and thus

$$\frac{\partial}{\partial a} f_{a,\beta}(p^*_{a,\beta})) = (-\beta_1^{-1}) \left(\frac{\partial}{\partial z} \frac{h(z)^2}{\dot{h}(z)} \Big|_{z=\beta_0+\beta_1 p^*_{a,\beta}} \right) \beta_1 \frac{\partial}{\partial a} p^*_{a,\beta}. \tag{5.32}$$

Log-concavity of h implies $\frac{\partial^2 \log(h(z))}{\partial z^2} = \frac{h(z)\ddot{h}(z) - \dot{h}(z)^2}{h(z)^2} \leq 0$, and thus

$$\begin{aligned}\frac{\partial}{\partial z}\frac{h(z)^2}{\dot{h}(z)} &= \frac{2h(z)\dot{h}(z)^2 - h(z)^2\ddot{h}(z)}{\dot{h}(z)^2} = h(z)\left[2 - \frac{h(z)\ddot{h}(z)}{h(z)^2}\frac{h(z)^2}{\dot{h}(z)^2}\right] \\ &\geq h(z)\left[2 - \frac{\dot{h}(z)^2}{h(z)^2}\frac{h(z)^2}{\dot{h}(z)^2}\right] = h(z).\end{aligned}$$

Since $\frac{\partial}{\partial a} p^*_{a,\beta} > 0$, it follows that $f_{a,\beta}(p^*_{a,\beta})$ is strictly decreasing in a. This completes the proof of (ii).

Let $K_0 = \sup_{(a,\beta,p)\in U_a \times U_B \times [p_l, p_h]} -\ddot{f}_{a,\beta}(p)/2$. Since $(a, \beta, p) \mapsto f_{a,\beta}(p)$ is twice continuously differentiable on $\mathbb{R} \times B \times [p_l, p_h]$ and $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$, it follows that $0 < K_0 < \infty$. By a Taylor expansion, there is a $\tilde{p}_{a,\beta}$ on the line segment between p and $p^*_{a,\beta}$, such that

$$\begin{aligned}f_{a,\beta}(p) &= f_{a,\beta}(p^*_{a,\beta}) + \dot{f}_{a,\beta}(p^*_{a,\beta})(p - p^*_{a,\beta}) + \frac{1}{2}\ddot{f}_{a,\beta}(\tilde{p}_{a,\beta})(p - p^*_{a,\beta})^2 \\ &\geq f_{a,\beta}(p^*_{a,\beta}) - K_0(p - p^*_{a,\beta})^2,\end{aligned}$$

using $\dot{f}_{a,\beta}(p^*_{a,\beta}) = 0$. This proves (iii).

Proof of Lemma 5.3

Let $\beta \in U_B$. We show $0 \leq \Delta V_\beta(c, s) \leq \max_{p\in[p_l,p_h]} ph(\beta_0 + \beta_1 p)$, for all $(c, s) \in \mathcal{X}$. By (5.8), this implies $\Delta V_\beta(c, s) \in U_a$. In view of (5.3), uniqueness and continuity of π^*_β then follow from repeated application of Lemma 5.2(i, ii), for each $(c, s) \in \mathcal{X}$.

If $s = S$ then $\Delta V_\beta(c, S) = 0$ for $c > 1$ or $c = 0$, and $V_\beta(1, S) = \max_{p\in[p_l,p_h]} ph(\beta_0 + \beta_1 p)$. If $s < S$, then by backward induction,

$$\begin{aligned}\Delta V_\beta(c, s) &= (\pi^*_\beta(c, s) - \Delta V_\beta(c, s + 1))h(\beta_0 + \beta_1\pi^*_\beta(c, s)) + V_\beta(c, s + 1) \\ &\quad - (\pi^*_\beta(c - 1, s) - \Delta V_\beta(c - 1, s + 1))h(\beta_0 + \beta_1\pi^*_\beta(c - 1, s)) - V_\beta(c - 1, s + 1) \\ &\geq (\pi^*_\beta(c - 1, s) - \Delta V_\beta(c, s + 1))h(\beta_0 + \beta_1\pi^*_\beta(c - 1, s)) + V_\beta(c, s + 1) \\ &\quad - (\pi^*_\beta(c - 1, s) - \Delta V_\beta(c - 1, s + 1))h(\beta_0 + \beta_1\pi^*_\beta(c - 1, s)) - V_\beta(c - 1, s + 1) \\ &= \Delta V_\beta(c, s + 1))(1 - h(\beta_0 + \beta_1\pi^*_\beta(c - 1, s))) \\ &\quad + \Delta V_\beta(c - 1, s + 1))h(\beta_0 + \beta_1\pi^*_\beta(c - 1, s)) \\ &\geq 0,\end{aligned}$$

and

$$\begin{aligned}
\Delta V_\beta(c,s) &= (\pi^*_\beta(c,s) - \Delta V_\beta(c,s+1))h(\beta_0+\beta_1\pi^*_\beta(c,s)) + V_\beta(c,s+1) \\
&\quad - (\pi^*_\beta(c-1,s) - \Delta V_\beta(c-1,s+1))h(\beta_0+\beta_1\pi^*_\beta(c-1,s)) - V_\beta(c-1,s+1) \\
&\le (\pi^*_\beta(c,s) - \Delta V_\beta(c,s+1))h(\beta_0+\beta_1\pi^*_\beta(c,s)) + V_\beta(c,s+1) \\
&\quad - (\pi^*_\beta(c,s) - \Delta V_\beta(c-1,s+1))h(\beta_0+\beta_1\pi^*_\beta(c,s)) - V_\beta(c-1,s+1) \\
&= \Delta V_\beta(c,s+1))(1-h(\beta_0+\beta_1\pi^*_\beta(c,s))) \\
&\quad + \Delta V_\beta(c-1,s+1))h(\beta_0+\beta_1\pi^*_\beta(c,s)) \\
&\le \max_{p\in[p_l,p_h]} ph(\beta_0+\beta_1 p).
\end{aligned}$$

Proof of Lemma 5.4

For any 2×2 positive definite matrix A with eigenvalues $0<\lambda_1\le\lambda_2$, we have $\lambda_2 \le \lambda_1+\lambda_2 = \mathrm{tr}(A)$, $\det(A)=\lambda_1\lambda_2$, and consequentially $\lambda_1 = \det(A)/\lambda_2 \ge \det(A)/\mathrm{tr}(A)$. If in addition $a,b\le p_h$, then

$$\lambda_{\min}\begin{pmatrix} 2 & a+b \\ a+b & a^2+b^2 \end{pmatrix} \ge \frac{2a^2+2b^2-(a+b)^2}{2+a^2+b^2)} \ge \frac{(a-b)^2}{2(1+p_h^2)}.$$

Since $\lambda_{\min}(P_t)\ge\lambda_{\min}(P_r)+\lambda_{\min}(P_{r'})$ for all $r,r',t\in\mathbb{N}$ with $r+r'=t$ (Bhatia, 1997, Corollary III.2.2, page 63), we have

$$\begin{aligned}
\lambda_{\min}(P_{kS}) &\ge \lambda_{\min}(P_{(k-1)S}) + \lambda_{\min}\left(\sum_{1\le i\le S, i\notin\{s,s'\}}\begin{pmatrix}1\\ p_{i+(k-1)S}\end{pmatrix}(1,p_{i+(k-1)S})\right) \\
&\quad + \lambda_{\min}\left(\begin{pmatrix}1\\ p_{s+(k-1)S}\end{pmatrix}(1,p_{s+(k-1)S}) + \begin{pmatrix}1\\ p_{s'+(k-1)S}\end{pmatrix}(1,p_{s'+(k-1)S})\right) \\
&\ge \lambda_{\min}(P_{(k-1)S}) + \frac{(p_{s+(k-1)S}-p_{s'+(k-1)S})^2}{2(1+p_h^2)} \\
&\ge \lambda_{\min}(P_{(k-1)S}) + \frac{\delta^2}{2(1+p_h^2)}.
\end{aligned}$$

Chapter 6

Dynamic pricing and learning in a changing environment

6.1 Introduction

The preceding Chapters 3, 4 and 5 study dynamic pricing and learning in a stationary environment: the parameters that describe the market behavior do not change over time. These chapters study in different pricing-and-learning settings the question if price experimentation is necessary, and how that should be conducted in an optimal manner.

The assumption of a stationary environment is a strong condition. Markets are generally not stable, but may vary over time, without the seller immediately being aware of it (cf. Dolan and Jeuland (1981), Wildt and Winer (1983), and Section 2 of Elmaghraby and Keskinocak (2003)). These changes may have various causes: shifts in consumer tastes, competition (Wildt and Winer, 1983), appearance of technological innovations (Chen and Jain, 1992), market saturation and product diffusion effects, which are related to the life cycle of a product (Bass, 1969, Dolan and Jeuland, 1981, Raman and Chatterjee, 1995), marketing and advertisement efforts (Horsky and Simon 1983), competitors entering or exiting the market, appearance of new sales channels, and many more.

Wildt and Winer (1983, page 365) argued already in 1983 that "constant-parameter models are not capable of adequately reflecting such changing market environments". In fact, the historical literature on statistical economics show that this issue has been known since longtime, as illustrated by the following quotation of Schultz (1925) on the law of demand:

> "The validity of the theoretical law [of demand] is limited to a point in time. But in order to derive *concrete, statistical* laws our observations must be numerous; and in order to obtain the requisite number of observations, data covering a considerable period must be used. During the interval, however, important dynamic changes take place in the condition of the market. In the case of a commodity like sugar, the principal dynamic changes that need be considered are the changes in our sugar-consuming habits, fluctuations in the purchasing power of money, and the increase of population." (page 409 of Schultz, 1925).

Although the literature on dynamic pricing and learning has increased rapidly in recent years, cf. Chapter 2, models with a varying market have hardly been considered. This motivates the current study of dynamic pricing and learning in a changing environment.

The combination of dynamic pricing and learning in a changing market, is a rather unexplored area. Besbes and Zeevi (2011) study a pricing problem where the willingness-to-pay (WtP) distribution of the customers changes at some unknown point in time. They do however assume that the WtP distribution before and after the change is fully known to the firm. Chen and Jain (1992) consider optimal pricing policies in models where the demand not only depends on the selling price, but also on the cumulative amount of sales; in this way diffusion effects are modeled. In addition, the demand is influenced by an observable state variable, which models unpredictable events that change the demand function, and whose dynamics are driven by a Poisson process. Apart from these random events, the demand again is fully deterministic and known to the firm, and learning by the firm is not considered. Hanssens et al. (2001) and Leeflang et al. (2009, Section 2.3) discuss several dynamic market models, as well as estimation methods, but do not integrate this with the problem of optimal dynamic pricing. A relevant study from the control literature is from Godoy et al. (2009). They consider an estimation problem in a linear system, where the parameters are subject to shock changes, and analyze the performance of a sliding-window linear regression method. A major assumption is that the controls are deterministic. This differs from pricing problems, where the prices (the controls) usually depend in a non-trivial way on all previously observed sales realizations.

In the present chapter, we study the problem of dynamic pricing and learning in a changing environment. We consider an additive demand model, where the expected demand is the sum of a stochastic process and a known function depending on the selling price. This stochastic process models the size of the market, and its characteristics are initially unknown to the firm. Its value at a certain point in time may be estimated from accumulated sales data; however, since the market may be changing over time, estimation methods are needed that are designed for time-varying systems. We deploy two such estimators, namely estimating with a forgetting factor, and estimation based on a "sliding window" approach. For both estimators we derive an upper bound on the expected estimation error.

Next, we propose a simple, intuitive pricing policy: at each decision moment, the firm estimates the market size with one of the just mentioned estimators, and subsequently sets the next selling price equal to the optimal price, given that the current market estimate is correct. This is a so-called myopic or certainty equivalent policy, since at each decision moment the firm acts as if being certain about its estimates. To measure the quality of this pricing policy, we define AverageRegret(T), which measures the expected costs of not choosing optimal prices in the first T periods, and LongRunAverageRegret, which equals the limit superior of AverageRegret(T) as T grows large. We derive upper bounds on AverageRegret(T) and LongRunAverageRegret. These bounds are not only stated in terms of the variables associated with the used estimation method (the forgetting factor, or the size of the sliding window), but also in terms of a measure of the *impact* that market fluctuations have on the estimation error. Clearly, if the market is very unstable and inhibits very large and frequent fluctuations, the impact may become extremely large, which negatively affects the obtained revenue.

The novel, key idea of this study is that (i) this impact can be bounded, using assumptions on the market process that the firm makes a priori, (ii) the resulting upper bounds on AverageRegret(T) and LongRunAverageRegret can be used by the firm to determine the optimal estimator of the market (i.e. the optimal value of the forgetting factor or window size), (iii) this provides the firm explicit guarantees on the maximum expected revenue loss. This framework enables the firm to *hedge against change*: the firm is certain that the expected regret does not exceed a certain known value, provided the market process satisfies the posed assumptions. These assumptions may be very general, and cover many important cases; for example, bounds on the probability that the market value changes in a certain period, bounds on the maximum difference between two consecutive market values, or bounds on the maximum and minimum value that the market process

may attain. We provide numerical examples to illustrate the methodology, in three practically relevant settings: in the first we assume that the market value is continuously changing; in the second, we make use of the well-known Bass model to model the diffusion of an innovative products; and in the third, we consider an oligopoly, where price changes by competitors causes occasional changes in the market. The application of our methodology on the Bass model makes this the first study that incorporates learning with this widely used product-diffusion model; so-far, only deterministic settings (Robinson and Lakhani, 1975, Dolan and Jeuland, 1981, Kalish, 1983), or random settings where no learning is present (Chen and Jain, 1992, Raman and Chatterjee, 1995, Kamrad et al., 2005) have been considered in the literature.

Summarizing, in one of the first studies on dynamic pricing and learning in a changing environment, our contributions are as follows:

(i) We introduce a model of dynamic pricing and learning in a changing market environment, using a very generic description of the market process.

(ii) We discuss two estimators of time-varying processes, and prove upper bounds on the estimation error.

(iii) We propose a methodology that enables the decision maker to hedge against change. This results in explicit guarantees on the regret, and guides the choice of the optimal estimator.

(iv) We show the application of the methodology in several concrete cases, and offer several numerical examples to illustrate its use and performance.

The rest of this chapter is organized as follows. The model is introduced in Section 6.2. In Section 6.3 we discuss estimation with forgetting factor or with sliding window, and provide upper bounds on the expected estimation error. The myopic pricing policy is described in Section 6.4, together with upper bounds on the regret. Section 6.5 discusses our proposed methodology of hedging against change, and examines its use in several concrete cases. Numerical illustrations are provided in Section 6.6. All mathematical proofs are contained in Section 6.7.

6.2 Model

We consider a monopolist firm selling a single type of product. In each time period $t \in \mathbb{N}$, the firm decides on a selling price $p_t \in [p_l, p_h]$, where $0 \leq p_l < p_h < \infty$ denote the lowest and highest admissible price. After choosing the price, the seller observes demand d_t, which is a realization of the random variable $D_t(p_t)$. Conditional on the selling prices, the demand in different time periods is independent. The expected demand in period t, against a price p, is of the form

$$E\left[D_t(p)\right] = M(t) + g(p). \tag{6.1}$$

Here $(M(t))_{t\in\mathbb{N}}$ is a nonnegative stochastic process called the *market process*. The dynamics of this process are unknown to the seller, and may be quite general: trends, seasonal effects, and shock changes, can all be captured by the market process. The process may even depend on historical sales data. In particular, let $\mathcal{F}_t$ be the σ-algebra generated by $d_1, p_1, M(1), \ldots, d_t, p_t, M(t)$, $\mathcal{F}_0$ the trivial σ-algebra, and write $\epsilon_t = d_t - g(p_t) - M(t)$; then we assume that $M(t)$, ϵ_t and p_t are all $\mathcal{F}_{t-1}$-measurable, for all $t \in \mathbb{N}$. In addition we impose the following mild conditions on the moments of $M(t)$ and ϵ_t: there are positive constants σ_M and σ, such that

$$\sup_{t\in\mathbb{N}} E\left[M(t)^2 \mid \mathcal{F}_{t-1}\right] \leq \sigma_M^2 \text{ a.s.} \qquad \text{and} \qquad \sup_{t\in\mathbb{N}} E\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right] \leq \sigma^2 \text{ a.s.} \tag{6.2}$$

The function g in (6.1) models the dependence of expected demand on selling price. It is assumed to be known by the seller. A typical example that is widely used in practice is the linear demand function $g(p) = -bp$ for some $b > 0$. After observing demand, the seller collects revenue $p_t d_t$, and proceeds to the next period. The purpose of the seller is to maximize expected revenue.

Let $r(p, M) = p \cdot (M + g(p))$ denote the expected revenue in a single period, when the market size equals M and the selling price is set at p. The price that generates the highest amount of expected revenue, given that the current market size equals M, is denoted by $p^*(M)) = \arg\max_{p \in [p_l, p_h]} r(p, M)$.

We impose some mild conditions to ensure that this optimal price exists and is uniquely defined. In particular, we assume that for all admissible prices p, $g(p)$ is decreasing in p, and twice continuously differentiable with first and second derivative denoted by $g'(p)$ and $g''(p)$. These two properties immediately carry over to the expected demand, and in fact are quite natural conditions for demand functions to hold. In addition, we assume that for all $M \geq 0$ the revenue function $r(p, M)$ is unimodal with unique optimum $p^\#(M) \geq 0$ satisfying $r'(p^\#(M), M) = 0$, and in addition $\sup_{M \geq 0 : p^\#(M) \in [p_l, p_h]} r''(p^\#(M), M) < 0$; here $r'(p, M)$ and $r''(p, M)$ denote the first and second derivative of $r(p, M)$ w.r.t. p. These assumptions on g and r are fairly standard conditions of demand and revenue functions, and ensure that the revenue function is locally strictly concave around the optimum. Clearly, if $p^\#(M)$ lies in the interval $[p_l, p_h]$ then $p^*(M) = p^\#(M)$, and if $p^\#(M) \notin [p_l, p_h]$, then $p^*(M)$ is the projection of $p^\#(M)$ on the interval $[p_l, p_h]$.

Remark 6.1. It is not difficult to show that the conditions on g are satisfied for the linear demand model $g(p) = -bp$, $b > 0$. For nonlinear demand functions like $g(p) = -bp^c$ with $b > 0, c > 0, c \neq 1$, or $g(p) = -b \log(p)$, $b > 0$, the conditions are satisfied if the market process is bounded: $\sup_{t \in \mathbb{N}} M(t) \leq M_{\text{max}}$ a.s. for some $M_{\text{max}} > 0$. This additional condition ensures that the condition $\sup_{M \geq 0 : p^\#(M) \in \mathcal{P}} r''(p^\#(M), M) < 0$ is satisfied. Note that for practical applications, it is generally not restrictive to assume a bound on the market.

Remark 6.2. In (6.1) we assume an additive demand model: the expected demand is the sum of a known and an unknown part. Another approach is to assume a multiplicative demand model, where the expected demand is the product of the two parts: $E[D_t(p)] = M(t) \cdot g(p)$. However, such a demand model has a remarkable property. By differentiating the revenue function w.r.t. p, one can easily show that the optimal price is the solution to the equation $pg'(p)/g(p) = -1$. This equation is independent of the market process, and as a result, the firm does not need to know or estimate the market in order to determine the optimal selling price. Intuitively it is clear that for many products such a model does not accurately reflect reality.

The value of the market process and the corresponding optimal price are unknown to the seller. As a result, the decision maker might choose sub-optimal prices, which incurs a loss of revenue relative to someone who would know the market process and the optimal price. The goal of the seller is to determine a pricing policy that minimizes this loss of revenue. With a pricing policy we here mean a sequence of prices $(p_t)_{t \in \mathbb{N}}$ in $[p_l, p_h]$ which is predictable w.r.t. $(\mathcal{F}_t)_{t \in \mathbb{N}}$; in other words, each price p_t may depend on all previously chosen prices $p_1, \ldots, p_{t-1}$, demand realizations $d_1, \ldots, d_{t-1}$, and market values $M(1), \ldots, M(t-1)$.

To asses the quality of a pricing policy Φ, we define the following two quantities.

$$\text{AverageRegret}(\Phi, T) = \frac{1}{T-1} \sum_{t=2}^{T} E\left[r(p^*(M(t)), M(t)) - r(p_t, M(t))\right], \tag{6.3}$$

$$\text{LongRunAverageRegret}(\Phi) = \limsup_{T \to \infty} \text{AverageRegret}(\Phi, T). \tag{6.4}$$

Each term in the summand of (6.3) measures the expected revenue loss caused by not using the

optimal price in period t. The expectation operator is because both p_t and $M(t)$ may be random variables. We start "measuring" the average regret from the second period. This simplifies several expressions that appear in further sections; in addition, in the first period, no data is available to estimate $M(1)$, and minimizing the instantaneous regret encountered in the first period is not possible. Furthermore, note that AverageRegret(Φ, T) and LongRunAverageRegret(Φ) are not observed by the seller, and thus can not directly be used to determine an optimal pricing policy.

6.3 Estimation of market process

Estimating the value of the market process gives vital information that is needed to determine the selling price. Since the market may change over time, the firm needs an estimation method that can handle such changes. In this section we describe two such methods: (I) estimation with forgetting factor, and (II) estimation with a sliding window.

(I) Estimation of $M(t)$ with forgetting factor. Let $\lambda \in [0, 1]$ be the forgetting factor, to be determined by the decision maker. The estimate $\hat{M}_\lambda(t)$, with forgetting factor λ, based on demand realizations $d_1, \ldots, d_t$ and prices $p_1, \ldots, p_t$, is equal to

$$\hat{M}_\lambda(t) = \arg\min_{M>0} \sum_{i=1}^{t} (d_i - M - g(p_i))^2 \lambda^{t-i}. \tag{6.5}$$

The factor λ^{t-i} acts as a weight on the data $(p_i, d_i)_{1 \le i \le t}$. Data that lies further in the past gets a lower weight; data from the recent past receives more weight (unless $\lambda = 1$, in which case all available data gets equal weight, or $\lambda = 0$, in which case only the most recent observation is taken into account). This captures the idea that the longer ago data has been generated, the likelier it is that the corresponding value of the market process differs from its current value. Accordingly, data from longer ago is assigned a smaller weight than data from the more recent past. Whether this intuition is true depends of course on the specific characteristics of $M(t)$.

By differentiating the righthandside of (6.5) w.r.t. M, we obtain the following explicit expression for $\hat{M}_\lambda(t)$:

$$\hat{M}_\lambda(t) = \frac{\sum_{i=1}^{t} (d_i - g(p_i)) \lambda^{t-i}}{\sum_{i=1}^{t} \lambda^{t-i}}. \tag{6.6}$$

(II) Estimation of $M(t)$ with a sliding window. Let $N \in \mathbb{N}_{\ge 2} \cup \{\infty\}$ be the window size, determined by the decision maker. The estimate $\hat{M}_N(t)$, with sliding window size N, based on demand realizations $d_1, \ldots, d_t$ and prices $p_1, \ldots, p_t$, is equal to

$$\hat{M}_N(t) = \arg\min_{M>0} \sum_{i=\max\{t-N+1,1\}}^{t} (d_i - M - g(p_i))^2. \tag{6.7}$$

Here only data from the N most recent observations is used to form an estimate. All data that is generated longer than N time periods ago, is neglected (if $N = \infty$, then all available data is taken into account). Similar to the estimate with forgetting factor, the rationale behind the estimate $\hat{M}_N(t)$ is the idea that for data generated long ago, it is more likely that the corresponding market value differs from its current value. This is captured in the fact that only the N most recent observations are used to estimate $M(t)$. Whether this idea is correct depends again on the specifics of $M(t)$.

Differentiating the righthandside of (6.7) w.r.t. M, we obtain the following expression:

$$\hat{M}_N(t) = \frac{1}{\min\{N,t\}} \sum_{i=\max\{t-N+1,1\}}^{t} (d_i - g(p_i)). \tag{6.8}$$

Both estimation methods (I) and (II) depend on a decision variable, (λ resp. N), that can be interpreted as a measure for the responsiveness to changes in the market. A high value of λ resp. N means that much information from the historical data is used to form estimates; this is advantageous in case of a stable market, but disadvantageous in case of many or large recent changes in the market process. Similarly, a low value of λ resp. N implies that the estimate of $M(t)$ is mainly determined by recent data; naturally, this is more beneficial in a volatile market than in a stable market. In Section 6.5 we offer several guidelines how to choose the values of λ or N.

Remark 6.3. The estimation methods (I) and (II) are designed for a very generic market process $M(t)$, that can capture various 'forms' such as trends, shock changes, seasonal effects, Markov-modulated processes, et cetera. Instead, if one imposes a certain structure on $M(t)$, estimators for $M(t)$ can be tailored according to this specific structure. For example, the firm could assume that $M(t) = ct + \epsilon_t$, where $c \in \mathbb{R}$ is a non-random constant and $(\epsilon_t)_{t\in\mathbb{N}}$ is a sequence of zero-mean i.i.d. random variables. Then a more sensible estimator of $M(t)$ would be $\hat{M}(t) = 2(t+1)^{-1}\sum_{i=1}^{t}(d_i - g(p_i))$, which satisfies $E\left[\hat{M}(t)\right] = \frac{2}{t+1}\sum_{i=1}^{t} c\cdot i = ct$, and thus is an unbiased estimator. However, assuming such additional structure comes with the cost of misspecification: if $M(t)$ is not of the form assumed by the firm, but for example c suffers from shock changes, this estimator may perform quite badly. We avoid such misspecifications by using estimators that are applicable in a very generic setting.

Market fluctuations influence the accuracy of the estimates $\hat{M}_\lambda(t)$ and $\hat{M}_n(t)$. The following quantities $I_\lambda(t)$ and $I_N(t)$ measure this *impact* or *influence* of market variations on the estimates. Observe that this impact is not solely determined by the market process, but also by the choice of λ and N:

$$I_\lambda(t) = E\left[\left|\left(\frac{1-\lambda}{1-\lambda^t}\mathbf{1}(\lambda<1) + \frac{1}{t}\mathbf{1}(\lambda=1)\right)\sum_{i=1}^{t}(M(i)-M(t+1))\lambda^{t-i}\right|^2\right],$$

$$I_N(t) = E\left[\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{t}(M(i)-M(t+1))\right|^2\right].$$

The following proposition gives a bound on the expected estimation error of (I) and (II), in terms of λ, N, and the impact measures $I_\lambda(t)$ and $I_N(t)$.

Proposition 6.1. *For all $t \in \mathbb{N}$,*

$$E\left[\left|\hat{M}_\lambda(t) - M(t+1)\right|^2\right] \le 2\sigma^2\left[\frac{(1-\lambda)}{(1+\lambda)}\frac{(1+\lambda^t)}{(1-\lambda^t)}\mathbf{1}(\lambda<1) + \frac{1}{t}\mathbf{1}(\lambda=1)\right] + 2I_\lambda(t) \tag{6.9}$$

and

$$E\left[\left|\hat{M}_N(t) - M(t+1)\right|^2\right] \le 2\frac{\sigma^2}{\min\{N,t\}} + 2I_N(t). \tag{6.10}$$

If the processes $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent, then

$$E\left[\left|\hat{M}_\lambda(t) - M(t+1)\right|^2\right] \leq \sigma^2\left[\frac{(1-\lambda)}{(1+\lambda)}\frac{(1+\lambda^t)}{(1-\lambda^t)}\mathbf{1}(\lambda<1) + \frac{1}{t}\mathbf{1}(\lambda=1)\right] + I_\lambda(t) \tag{6.11}$$

and

$$E\left[\left|\hat{M}_N(t) - M(t+1)\right|^2\right] \leq \frac{\sigma^2}{\min\{N,t\}} + I_N(t), \tag{6.12}$$

with equality in (6.11), (6.12) *if the disturbance terms are homoscedastic, i.e. $E[\epsilon_t^2 \mid \mathcal{F}_{t-1}] = \sigma^2$ for all $t \in \mathbb{N}$.*

The first terms of the righthandsides of (6.9) - (6.12) are related to the natural fluctuations in demand. The lower these fluctuations, measured by σ^2, the lower this part of the estimation error becomes. The second terms of the righthandsides of (6.9) - (6.12) relate to the impact that market fluctuations have on the quality of the estimate of $M(t)$.

6.4 Pricing policy and performance bounds

The two estimation methods (I) and (II), introduced in Section 6.3, can be used by the seller to determine the selling prices. We study the situation where the seller uses the following simple, myopic pricing policy: at each decision moment, the seller estimates the market value with one of the two estimation methods described in Section 6.3. Subsequently, s/he chooses the selling price that is optimal w.r.t. this estimate. In other words, the seller always act as if the current estimate of the market is correct.

We denote this policy by Φ_λ if the market is estimated by method (I), with forgetting factor λ, and by Φ_N if the market is estimated by method (II), with sliding window of size N. The formal description of Φ_λ and Φ_N is as follows.

Myopic pricing policy Φ_λ / Φ_N

Initialization: Choose $\lambda \in [0,1]$ or $N \in \mathbb{N}_{\geq 2} \cup \{\infty\}$.
Set $p_1 \in [p_l, p_h]$ arbitrarily.
For all $t \in \mathbb{N}$:
Estimation: Let $\hat{M}_\cdot(t)$ denote either $\hat{M}_\lambda(t)$ (for policy Φ_λ) or $\hat{M}_N(t)$ (for policy Φ_N).
Pricing: Set $p_{t+1} = p^*(\max\{0, \hat{M}_\cdot(t)\})$.

The following theorem provides upper bounds on the (long run) average regret, for both myopic pricing policies Φ_λ and Φ_N:

Theorem 6.1. *There is a $K_0 > 0$, only depending on g, such that for all $T \geq 2$,*

$$\begin{aligned} \mathit{AverageRegret}(\Phi_\lambda, T) \leq\ & 2K_0\sigma^2\left[\frac{1-\lambda}{1+\lambda} + \frac{2}{T-1}\left(\frac{\lambda\log(\lambda) + (1-\lambda)\log(1-\lambda)}{(1+\lambda)\log(\lambda)}\right)\right]\mathbf{1}(\lambda<1) \\ & + 2K_0\sigma^2\left[\frac{1+\log(T-1)}{T-1}\right]\mathbf{1}(\lambda=1) \\ & + 2K_0\frac{1}{T-1}\sum_{t=1}^{T-1} I_\lambda(t), \end{aligned}$$

and

$$\textit{AverageRegret}(\Phi_N, T) \leq 2K_0\sigma^2 \left[\frac{\log(\min\{T-1, N\})}{T-1} + \frac{1}{\min\{N, T-1\}}\right] + \frac{2K_0}{T-1}\sum_{t=1}^{T-1} I_N(t).$$

Consequentially,

$$\textit{LongRunAverageRegret}(\Phi_\lambda) \leq 2K_0 \left[\sigma^2 \frac{1-\lambda}{1+\lambda} + \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} I_\lambda(t)\right], \tag{6.13}$$

for all $\lambda \in [0, 1]$*, and*

$$\textit{LongRunAverageRegret}(\Phi_N) \leq 2K_0 \left[\sigma^2 \frac{1}{N} + \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} I_N(t)\right], \tag{6.14}$$

for all $N \in \mathbb{N}_{\geq 2} \cup \{\infty\}$*, where we write* $1/\infty = 0$*.*

The main idea of the proof is to show that there is a $K_0 > 0$ such that for any M and M', we have $r(p^*(M), M) - r(p^*(M'), M) \leq K_0(M - M')^2$. Subsequently we apply the bounds derived in Proposition 6.1.

Remark 6.4. By (6.11) and (6.12), if the processes $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent, then all four inequalities of Theorem 6.1 are still valid if all righthandsides are divided by 2.

Remark 6.5. An explicit expression for K_0 is derived in the proof of Theorem 6.1. To obtain the most sharp bounds, one could also define K_0 directly as
$K_0 = \inf_{M,M'>0, M\neq M'} (r(p^*(M), M) - r(p^*(M'), M))/(M - M')^2$. For the important special case of a linear demand function, $g(p) = -bp$ for some $b > 0$, it is not difficult to show $p^*(M) = \min\{\max\{M/(2b), p_l\}, p_h\}$ and $K_0 = 1/(4b)$.

Remark 6.6. In dynamic pricing and learning studies that assume a stable market, one often considers the asymptotic behavior of $\text{Regret}(\Phi, T) = (T-1) \cdot \text{AverageRegret}(\Phi, T)$, where Φ the pricing policy that is used. Typically one proves bounds on the growth rate of $\text{Regret}(\Phi, T)$ for a certain policy, e.g. $\text{Regret}(\Phi, T) = O(\sqrt{T})$ or $\text{Regret}(\Phi, T) = O(\log(T))$. A policy is considered 'good' if the speed of convergence of the regret is close the best achievable rate, cf. Chapters 3, 4, 5. In the setting with a changing market, a simple example makes clear that one cannot do better than $\text{Regret}(\Phi, T) = O(T)$ or $\text{AverageRegret}(\Phi, T) = O(1)$. Suppose $M(t)$ is a Markov process taking values in $\{M_1, M_2\}$, with $M_1 \neq M_2$, and suppose $P(M(t+1) = M_i \mid M(t) = M_j) = \frac{1}{2}$, for all $i, j \in \{1, 2\}$ and $t \in \mathbb{N}$. Let $g(p) = -bp$, for some $b > 0$, and choose $[p_l, p_h]$ such that $p^\#(M_i) = M_i/(2b) \in (p_l, p_h)$, for $i = 1, 2$. Then for all $t \in \mathbb{N}$, the instantaneous regret incurred in period $t+1$ satisfies

$$\begin{aligned}
&E\left[r(p^*(M(t+1)), M(t+1)) - r(p_{t+1}, M(t+1))\right] \\
&\geq \inf_{p\in[p_l,p_h]} \left[\frac{1}{2}(r(p^*(M_1), M_1) - r(p, M_1)) + \frac{1}{2}(r(p^*(M_2), M_2) - r(p, M_2))\right] \\
&\geq \frac{b}{2} \inf_{p\in[p_l,p_h]} \left[(p^*(M_1) - p)^2 + (p^*(M_2) - p)^2\right] \\
&\geq \frac{b}{4}(p^*(M_1) - p^*(M_2))^2 \\
&\geq \frac{1}{16b}(M_1 - M_2)^2 > 0,
\end{aligned}$$

which implies that no policy can achieve a sub-linear $\text{Regret}(\Phi, T) = o(T)$. In fact, any pricing policy achieves the optimal growth rate $\text{Regret}(\Phi, T) = O(T)$. Thus, the challenge of dynamic

pricing and learning in such a changing environment is not to find a policy with optimal asymptotic growth rate, but rather to make the (long run) average regret as small as possible.

In view of the remark above, the question raises whether the bounds from Theorem 6.1 are sharp. The following proposition answers this question for the case of a linear demand function with homoscedastic disturbance terms independent of the market process.

Proposition 6.2. *Suppose $g(p) = -bp$ for some $b > 0$, $E[\epsilon_t^2 \mid \mathcal{F}_{t-1}] = \sigma^2$ for all $t \in \mathbb{N}$, the processes $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent, and $M(t) \in [2bp_l, 2bp_h]$ a.s. for all $t \in \mathbb{N}$. Then with $K_0 = 1/(4b)$,*

$$\mathit{LongRunAverageRegret}(\Phi_\lambda) = K_0 \left[\sigma^2 \frac{1-\lambda}{1+\lambda} + \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} I_\lambda(t) \right], \tag{6.15}$$

for all $\lambda \in [0, 1]$, and

$$\mathit{LongRunAverageRegret}(\Phi_N) = K_0 \left[\sigma^2 \frac{1}{N} + \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} I_N(t) \right], \tag{6.16}$$

for all $N \in \mathbb{N}_{\geq 2} \cup \{\infty\}$, where we write $1/\infty = 0$.

6.5 Hedging against changes

The bounds on the regret that we derive in Theorem 6.1 can be used by the firm to hedge against changes in the market. To this end, the firm should first decide what type of changes it is anticipating on, in the form of assumptions on the market process. Examples of such assumptions are (i) the market process is contained in an interval of known size, (ii) the maximum change in the market between two consecutive periods is bounded, (iii) market changes are bounded and occur only with a small probability, (iv) the market process is a sequence of i.i.d. realizations of a random variable. Given such assumptions on the market process, the firm can determine the corresponding optimal choice of λ (in case policy Φ_λ is used) or N (in case of Φ_N).

This is done as follows. First, one needs to translate the assumptions on the market process into upper bounds on the impact measures $I_\lambda(t)$ and $I_N(t)$. In particular, the goal is to find (non-random) functions $c_T(\lambda)$ and $c_T(N)$, such that for all $\lambda \in [0, 1]$ and $N \in \mathbb{N}_{\geq 2} \cup \{\infty\}$,

$$\frac{1}{T-1} \sum_{t=1}^{T-1} I_\lambda(t) \leq c_T(\lambda), \tag{6.17}$$

$$\frac{1}{T-1} \sum_{t=1}^{T-1} I_N(t) \leq c_T(N). \tag{6.18}$$

By plugging these bounds into Theorem 6.1, we obtain bounds on AverageRegret(Φ_λ, T) and AverageRegret(Φ_N, T) in terms of λ and N. The optimal choices of λ and N are then determined by simply minimizing these bounds with respect to λ and N. In some cases an explicit expression for the optimal choice may exist, otherwise numerical methods are needed to determine the optimum.

The resulting optimal optimal λ and N may depend on the length of the time horizon T. This may be undesirable to the firm, for instance because T is not known in advance, or because the time horizon is infinite. In this case it is more appropriate to minimize the LongRunAverageRegret. If

$c(\lambda)$ and $c(N)$ are functions that satisfy

$$\limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} I_\lambda(t) \leq c(\lambda), \tag{6.19}$$

$$\limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} I_N(t) \leq c(N), \tag{6.20}$$

then (6.13) and (6.14) imply

$$\text{LongRunAverageRegret}(\Phi_\lambda) \leq 2K_0\big[\sigma^2 \frac{1-\lambda}{1+\lambda} + c(\lambda)\big], \tag{6.21}$$

$$\text{LongRunAverageRegret}(\Phi_N) \leq 2K_0\big[\sigma^2 \frac{1}{N} + c(N)\big], \tag{6.22}$$

and the optimal choices of λ and N can be determined by minimizing the righthandsides of (6.21) and (6.22).

Remark 6.7. It is interesting to observe that the optimal choices of λ and N are *independent* of the function g. The relevant properties of g are captured by the constant K_0, but its value does not influence the optimal λ and N. In a way this separates optimal estimation and optimal pricing: the first is determined by the impact of the market process, while only the latter involves the function g. On the other hand, the variance of the demand distribution, related to σ^2, does influence the optimal λ and N. In addition, note that by Remark 6.4, the factor 2 on the righthandsides of (6.21) and (6.22) can be removed if the processes $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent. In practice, it may not always be known to the decision maker whether this condition is satisfied; but fortunately, this does not influence the optimal choice of λ and N.

To illustrate the methodology, we now look in more detail to the market scenarios (i) - (iv) mentioned in the beginning of this section. We derive the bound functions $c(\lambda)$ and $c(N)$ for each scenario, and show how to choose λ and N in order to minimize the righthandsides of (6.21) and (6.22).

Scenario (i). Bounds on the range of the market process. Here the firm assumes that for all $t \in \mathbb{N}$, $M(t)$ is a.s. contained in an interval with size $d > 0$. Then $|M(i) - M(t+1)| \leq d$ for all $i = 1, \ldots, t$, and it follows that we can take $c(\lambda) = c(N) = d^2$.

The righthandsides of (6.21) and (6.22) are minimized by taking $\lambda = 1$ and $N = \infty$. This is true regardless the value of d.

At first sight it may seem somewhat surprising that it is beneficial to take into account all available sales data to estimate the market, including 'very old' data. This can be explained by noting that in a period $t+1$, *all* preceding values of the market $M(1), \ldots, M(t)$ may differ by d from the current value $M(t+1)$. In such a volatile market situation, it is best to 'accept' an unavoidable error caused by market fluctuations, and instead focus on minimizing the estimation error caused by natural fluctuations $\epsilon_1, \ldots, \epsilon_t$ in the demand distribution. This is best done when all available data is taken into account; hence the optimality of choosing $\lambda = 1$ and $N = \infty$, even when d is very small.

Scenario (ii). Bounds on one-step market changes. Here the firm assumes that for all $t \in \mathbb{N}$, $|M(t) - M(t+1)| \leq d$ a.s., for some $d > 0$. At the end of Section 6.7, we derive

$$c(\lambda) = \begin{cases} d^2(1-\lambda)^{-2} & \text{if } \lambda \in [0,1) \\ \infty & \text{if } \lambda = 1 \end{cases},$$

and

$$c(N) = \begin{cases} \frac{1}{4}d^2(N+1)^2 & \text{if } N \in \mathbb{N}_{\geq 2} \\ \infty & \text{if } N = \infty \end{cases} .$$

Both $c(\lambda)$ and $c(N)$ are increasing in d. This can be interpreted as larger deviations in the market process having more impact on the estimates. In addition, $c(\lambda)$ and $c(N)$ are both increasing in its variables λ, N. This means that in this scenario, the impact increases if more data is taken into account. This can be explained intuitively by observing that in this scenario, older data may have been influenced by more changes in the market process than recent data.

Consider the upper bound (6.21). The derivative of $\sigma^2 \frac{(1-\lambda)}{(1+\lambda)} + d^2(1-\lambda)^{-2}$ w.r.t. $\lambda \in (0,1)$ is zero if and only if $(\sigma/d)^2(1-\lambda)^3 = (1+\lambda)^2$; this follows from basic algebraic manipulations. Since $(1-\lambda)^3$ is decreasing and $(1+\lambda)^2$ is increasing in λ, we have the following possibilities:

1. $(\sigma/d)^2 \leq 1$. Then $\sigma^2 \frac{(1-\lambda)}{(1+\lambda)} + d^2(1-\lambda)^{-2}$ is increasing on $\lambda \in (0,1)$, and the optimal choice is $\lambda = 0$.

2. $(\sigma/d)^2 > 1$. Then there is a unique $\lambda^* \in (0,1)$ that minimizes $\sigma^2 \frac{(1-\lambda)}{(1+\lambda)} + d^2(1-\lambda)^{-2}$. Although an explicit expression exists for λ^*, it is rather complicated, and it is not informative to state it here. Computing λ^* numerically is easy, and only involves solving a cubic equation.

Now consider the upper bound (6.22). The expression $\frac{\sigma^2}{N} + \frac{1}{4}d^2(N+1)^2$ is minimized by choosing N as the solution to $N^2(N+1) = 2(\sigma/d)^2$, which follows by taking the derivative w.r.t. N and some basic algebraic manipulations. It can easily be shown that there is a unique solution $N^* \in \mathbb{R}_{++}$, at which the minimum is attained, and that $\frac{\sigma^2}{N} + c(N)$ is minimized by choosing N as either $\lfloor N^* \rfloor$ or $\lceil N^* \rceil$. If $(\sigma/d)^2 \leq 10/4$ then the optimal N equals 1, if $(\sigma/d)^2 > 10/4$ then the optimal N is strictly larger than 1.

In this scenario, the quantity $(\sigma/d)^2$ serves as a proxy for the volatility of the market process $(M(t))_{t\in\mathbb{N}}$ relative to the variance of the disturbance terms $(\epsilon_t)_{t\in\mathbb{N}}$. Both for Φ_λ and Φ_N one can show that the optimal choice of λ and N is monotone increasing in this quantity $(\sigma/d)^2$. The larger the volatility of the market compared to the variance of the disturbance terms, the fewer data should be used to estimate the market. If $(\sigma/d)^2$ is sufficiently small, then the market fluctuations are quite large relative to the variance of the disturbance terms, and it is optimal to take only the most recent data point into account to estimate the market.

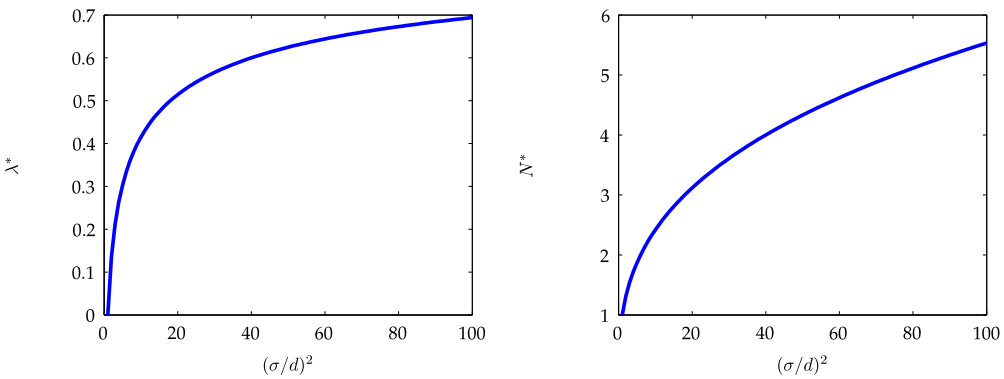

Figure 6.1: Relation between $(\sigma/d)^2$ and λ^*, N^*, for scenario (ii)

Scenario (iii). Bounded jump probabilities for the market process This scenario does not only involve assumptions on the maximum amount of change in the market, but also on the probabil-

ity of change. In particular, the firm assumes that for all $t \in \mathbb{N}$, $P\left(M(t+1) \neq M(t)\right) \leq \epsilon$, and that all $M(t)$ are contained in an interval of size d, for some $\epsilon \in (0,1)$ and $d > 0$. At the end of Section 6.7, we derive

$$c(\lambda) = \begin{cases} d^2\epsilon(1-\lambda^2)^{-1} & \text{if } \lambda \in [0,1) \\ \infty & \text{if } \lambda = 1 \end{cases},$$

$$c(N) = \begin{cases} d^2\epsilon\frac{(N+1)(2N+1)}{6N} & \text{if } N \in \mathbb{N}_{\geq 2} \\ \infty & \text{if } N = \infty \end{cases}.$$

Both $c(\lambda)$ and $c(N)$ are increasing in d and ϵ: the impact increases if market changes occur more often (ϵ increases), or if the magnitude of the jumps increases (d increases). Furthermore, $c(\lambda)$ and $c(N)$ are both increasing in λ and N, which means that taking more data into account increases the impact of market changes on the market estimate.

Consider the upper bound (6.21). The derivative of $\sigma^2\frac{(1-\lambda)}{(1+\lambda)} + d^2\epsilon(1-\lambda^2)^{-1}$ w.r.t. $\lambda \in (0,1)$ is zero if and only if $\frac{\sigma^2}{d^2\epsilon}(1-\lambda^2)^2 = \lambda(1+\lambda)^2$; this follows from basic algebraic manipulations. Since $(1-\lambda^2)^2$ is decreasing and $\lambda(1+\lambda)^2$ is increasing in λ, we have the following possibilities:

1. $\frac{\sigma^2}{d^2\epsilon} \leq 1$. Then $\sigma^2\frac{(1-\lambda)}{(1+\lambda)} + d^2\epsilon(1-\lambda^2)^{-1}$ is increasing on $\lambda \in (0,1)$, and the optimal choice is $\lambda = 0$.
2. $\frac{\sigma^2}{d^2\epsilon} > 1$. Then there is a unique $\lambda^* \in (0,1)$ that minimizes $\sigma^2\frac{(1-\lambda)}{(1+\lambda)} + d^2\epsilon(1-\lambda^2)^{-1}$. It is the unique solution in $(0,1)$ of the quartic equation $\frac{\sigma^2}{d^2\epsilon}(1-\lambda^2)^2 = \lambda(1+\lambda)^2$, which can easily be solved numerically.

Now consider the upper bound (6.22). The expression $\frac{\sigma^2}{N} + d^2\epsilon\frac{(N+1)(2N+1)}{6N}$ is minimized on $\mathbb{R}_{++}$ by choosing $N^* = \sqrt{\frac{3\sigma^2}{d^2\epsilon} + \frac{1}{2}}$, and the optimal N is equal to either $\lfloor N^* \rfloor$ or $\lceil N^* \rceil$. In addition, one can easily show that the optimal N equals 1 if $\frac{\sigma^2}{d^2\epsilon} \leq \frac{1}{2}$, and is strictly larger than 1 if $\frac{\sigma^2}{d^2\epsilon} > \frac{1}{2}$.

The quantity $\frac{\sigma^2}{d^2\epsilon}$ serves as a proxy for the volatility of the market process $(M(t))_{t\in\mathbb{N}}$ relative to the variance of the disturbance terms $(\epsilon_t)_{t\in\mathbb{N}}$. A low value of $\frac{\sigma^2}{d^2\epsilon}$ means a high volatility, a high value means a small volatility. This quantity is decreasing in the jump probability ϵ and the maximum market jump d, and increasing in the variance σ^2 of the disturbance terms $(\epsilon_t)_{t\in\mathbb{N}}$. The effect of $\frac{\sigma^2}{d^2\epsilon}$ on λ^* and N^* is shown in Figure 6.2. It clearly shows that the smaller the volatility of the market (e.g. the larger $\frac{\sigma^2}{d^2\epsilon}$), the more data should be taken into account to estimate the market process.

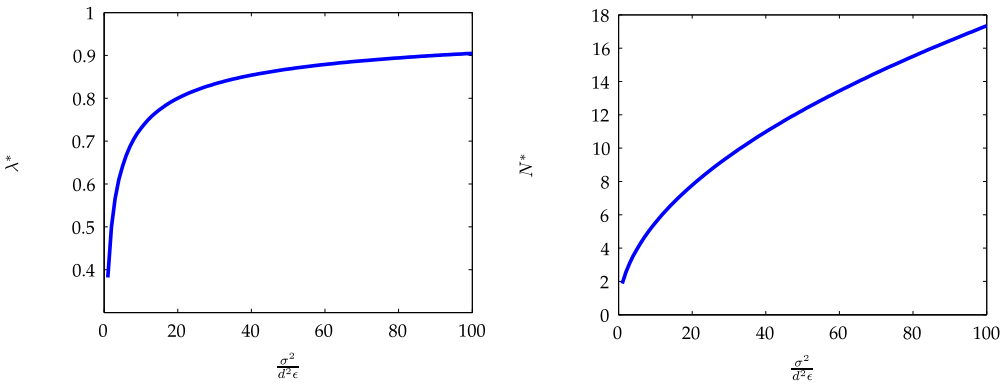

Figure 6.2: Relation between $\frac{\sigma^2}{d^2\epsilon}$ and λ^*, N^*, for scenario (iii)

Scenario (iv). Market process as i.i.d. sequence. In this last scenario we consider, the firm assumes that $(M(t))_{t\in\mathbb{N}}$ is a sequence of i.i.d. realizations of a random variable X on $[0,\infty)$, with finite variance denoted by $\mathrm{Var}(X)$. At the end of Section 6.7, we show

$$c(\lambda) = \mathrm{Var}(X)\frac{2}{1+\lambda},$$

and

$$c(N) = \mathrm{Var}(X)\Big(1+\frac{1}{N}\Big),$$

where we write $1/\infty = 0$.

These bounds are increasing in $\mathrm{Var}(X)$: the higher the variance, the larger the impact of market fluctuations on the estimates. Both $c(\lambda)$ and $c(N)$ are decreasing in its variables λ resp. N, and the righthandsides of (6.22) and (6.21) are minimized by taking $\lambda = 1$ and $N = \infty$; in other words, in this scenario it is optimal to take into account all available data to estimate the market process.

Remark 6.8. The above presented methodology of hedging against change has some similarities with robust optimization. There, one usually considers optimization problems whose optimal solutions depend on some parameters. These parameters are not known exactly by the decision maker, but assumed to lie in a certain "uncertainty set" which is known in advance. The optimal decision is then determined by optimizing against the worst case of the possible parameter values. An improvement of our methodology compared to robust optimization, is that we allow for many different types of assumptions on the market process, as illustrated by the four scenarios described above. In contrast, robust optimization generally only assumes a setting of an uncertainty set (comparable to our scenario (i)); trends, or Markov-modulated processes, are generally not considered. In addition, in robust optimization there is often no learning of the unknown parameters, whereas our methodology allows using accumulating data to estimate the unknown process; in several instances this enables us to "track" the market process.

6.6 Numerical illustration

In this section, we perform several numerical experiments to illustrate the pricing policy and choice of variables λ, N, as described in Section 6.5.

Experiment 1: continuously fluctuating market process. For each t, $M(t)$ is an independent realization of a random variable, uniformly distributed on the interval $[21, 24]$. Take $g(p) = -p$, $p_l = 5$, $p_h = 15$, and let $(\epsilon_t)_{t\in\mathbb{N}}$ be i.i.d. realizations of a standard normal distribution. For each $\lambda \in \{0.81, 0.82, \ldots, 0.99, 1\}$ we run 1000 simulations of the policy Φ_λ, and for all $N \in \{5, 15, 25, \ldots, 195\}$, we run 1000 simulations of Φ_N.

Figure 6.3 shows a typical sample path of $M(t)$ and $\hat{M}(t)$, both for Φ_λ and Φ_N. The market, indicated by the solid line, is continuously fluctuating around its mean, 22.5. The market estimate, depicted by the dashed line, stays much closer to 22.5.

The solid lines in Figure 6.4 show the simulation average of AverageRegret at $t = 2000$ for both Φ_λ and Φ_N, at different values of λ. The dashed lines show the upper bounds $K_0(\sigma^2\frac{1-\lambda}{1+\lambda} + c(\lambda))$ for Φ_λ, and $K_0(\sigma^2/N + c(N))$ for Φ_N, where $c(\lambda)$ and $c(N)$ are as in Scenario (iv), $K_0 = 1/4$ and $\mathrm{Var}(M(t)) = 3/4$ for all $t \in \mathbb{N}$. Note that $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are here independent, and thus by Remark 6.7, the factor 2 in the righthandsides of (6.21) and (6.22) is not present. The dashed and solid lines practically coincide in these figures.

This figure shows exactly the structure predicted by Scenario (iv): the average regret decreases

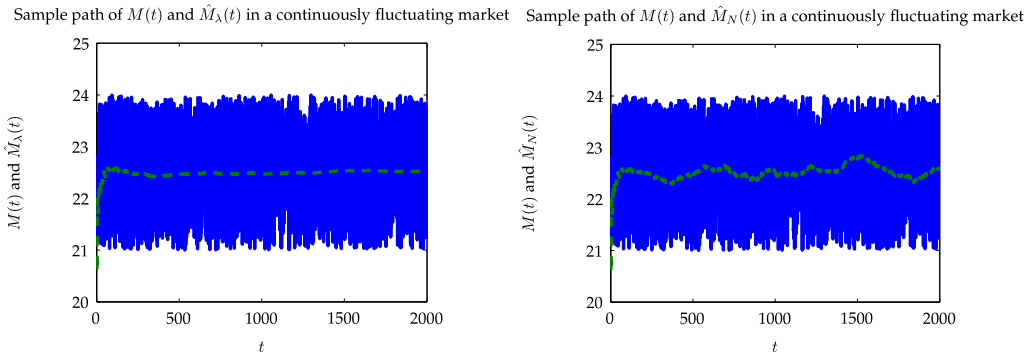

Figure 6.3: Sample path of $M(t)$ and $\hat{M}(t)$ in a continuously fluctuating market

as λ or N increases. The optimal choice of λ and N that minimize the long run average regret, is to set $\lambda = 1$ and $N = \infty$.

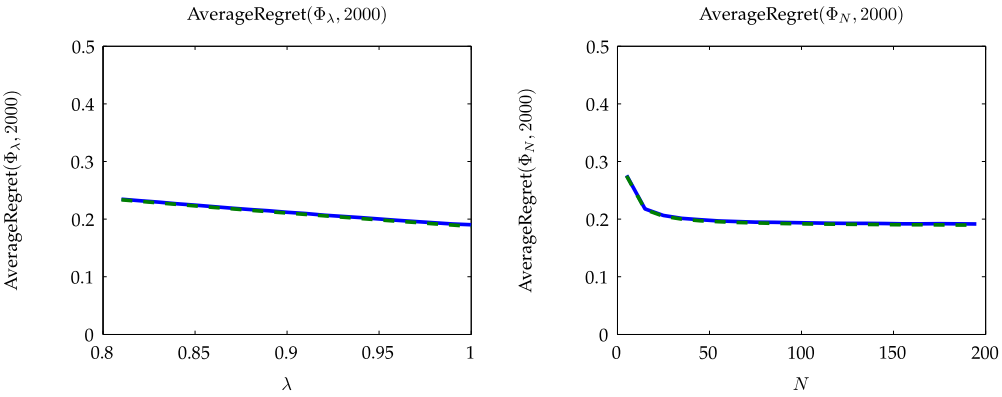

Figure 6.4: AverageRegret$(\Phi_\lambda, 2000)$ and AverageRegret$(\Phi_N, 2000)$ for experiment 1.

Experiment 2: Bass model for market process. The Bass model (Bass, 1969) is a widely-used model to describe the life-cycle or diffusion of an innovative product. An important property of this model is that the market process $M(t)$ is dependent on the realized cumulative sales up to time t. The model for $M(t)$ is

$$M(t) = \max\left\{0, a + b\sum_{i=1}^{t-1} d_i + c\Big(\sum_{i=1}^{t-1} d_i\Big)^2\right\},$$

cf. equation (4) of Dodds (1973). We choose $a = 33.6$, $c = -10^{-6}$ and $b = 0.0116$, and set $g(p) = -p$, $p_l = 1$ and $p_h = 50$. Let $(\epsilon_t)_{t\in\mathbb{N}}$ be i.i.d. realizations of a standard normal distribution. The characteristic shape of the market that arises from this model, is depicted in Figure 6.5. The solid lines denote a sample path of $M(t)$, the dashed lines a sample path of the estimates $\hat{M}_\lambda(t)$ and $\hat{M}_N(t)$.

For each $\lambda \in \{0.05, 0.10, 0.15, \ldots, 0.90\}$ we run 1000 simulations of the policy Φ_λ, and for all $N \in \{2, 3, 4, \ldots, 25\}$, we run 1000 simulations of Φ_N.

The solid lines in Figure 6.6 show the simulation-average of AverageRegret at $t = 500$ for both Φ_λ and Φ_N, at different values of λ. The dashed lines show the upper bounds $2K_0(\sigma^2\frac{1-\lambda}{1+\lambda} + c(\lambda))$ for Φ_λ, and $2K_0(\sigma^2/N + c(N))$ for Φ_N, where $c(\lambda)$ and $c(N)$ are as in Scenario (ii), $\sigma^2 = 1$, $K_0 = 1/4$,

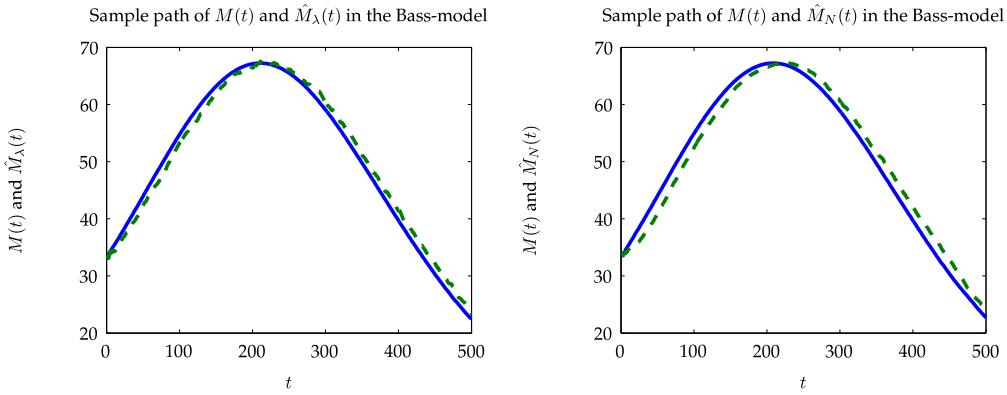

Figure 6.5: Sample path of $M(t)$ and $\hat{M}(t)$ in the Bass-model

and $d = 0.27$ (this was the largest observed value of $|M(t+1) - M(t)|$ over all t and all simulations. Of course, this quantity is in practice not observed by the seller, and a larger value of d just shifts the dashed lines upward in the figure).

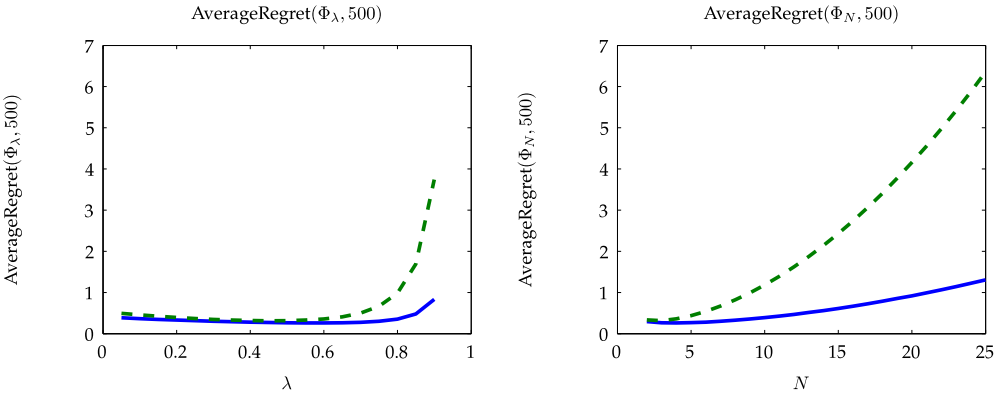

Figure 6.6: AverageRegret$(\Phi_\lambda, 500)$ and AverageRegret$(\Phi_N, 500)$ for experiment 2.

The optimal value of λ according to our upper bound equals $\lambda = 0.45$, with a corresponding upper bound on the regret of 0.31. The simulation average of AverageRegret$(\Phi_{0.45}, 500)$ was equal to 0.27. The optimal value of λ according to the simulations, was $\lambda = 0.60$, with a simulation average of AverageRegret$(\Phi_{0.60}, 500)$ equal to 0.26.

The optimal value of N according to our upper bound equals $N = 3$, with a corresponding upper bound on the regret of 0.32. The simulation average of AverageRegret$(\Phi_3, 500)$ was equal to 0.27. The optimal value of N according to the simulations, was $N = 4$, with a simulation average of AverageRegret$(\Phi_4, 500)$ equal to 0.26.

Figure 6.6 illustrates that taking into account all available data (i.e. $\lambda = 1$ or $N = \infty$) would lead to a very large regret. Thus, in this scenario, taking into account the changing nature of the market process improves the performance of the firm significantly.

Experiment 3: presence of price-changing competitors. Suppose the firm is acting in an environment where several competing companies are selling substitute products on the market. The firm knows that the competitors occasionally update their selling prices, but is not aware of the moments at which these changes occur. For this setting, the assumptions of scenario (iii) are appropriate.

In particular, consider the following case. The firm assumes that in each period, the probability that the market process changes because of the behavior of competitors, is not more than ϵ. If a change occurs, the maximum jump is assumed to be not more than d.

We choose $g(p) = -p$, $p_l = 1$ and $p_h = 50$, and let $\epsilon = 0.02$, $d = 5$. At each period t a realization z_t of a uniformly distributed random variable on $[0, 1]$ is drawn. If $z_t \geq 0.02$ then $M(t) = M(t-1)$; otherwise, $M(t)$ is drawn uniformly from the interval $[30, 35]$. Let $(\epsilon_t)_{t\in\mathbb{N}}$ be i.i.d. realizations of a standard normal distribution. (Note that these differ from the constant ϵ determined by the firm).

The characteristic the shape of the market that arises from this model, is depicted in Figure 6.7. The solid lines denote a sample path of $M(t)$, the dashed lines a sample path of the estimates $\hat{M}_\lambda(t)$ and $\hat{M}_N(t)$.

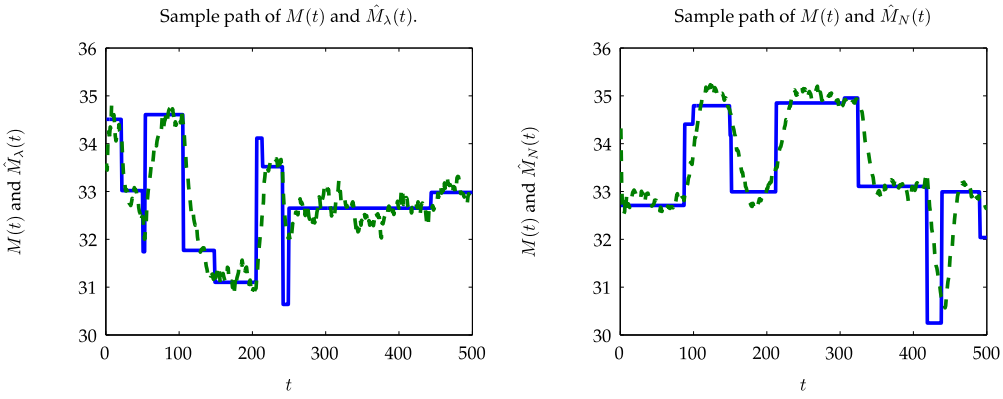

Figure 6.7: Sample path of $M(t)$ and $\hat{M}(t)$ in the model with price-changing competitors.

For each $\lambda \in \{0.10, 0.15, 0.20, \ldots, 0.95\}$ we run 1000 simulations of the policy Φ_λ, and for all $N \in \{2, 3, 4, \ldots, 25\}$, we run 1000 simulations of Φ_N.

The solid lines in Figure 6.8 show the simulation average of AverageRegret at $t = 500$ for both Φ_λ and Φ_N, at different values of λ. The dashed lines show the upper bounds $K_0(\sigma^2 \frac{1-\lambda}{1+\lambda} + c(\lambda))$ for Φ_λ, and $K_0(\sigma^2/N + c(N))$ for Φ_N, where $c(\lambda)$ and $c(N)$ are as in Scenario (iii), $\sigma^2 = 1$, $K_0 = 1/4$, $\epsilon = 0.02$, and $d = 5$. Note that $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are here independent, and thus by Remark 6.7, the factor 2 in the righthandsides of (6.21) and (6.22) is not present.

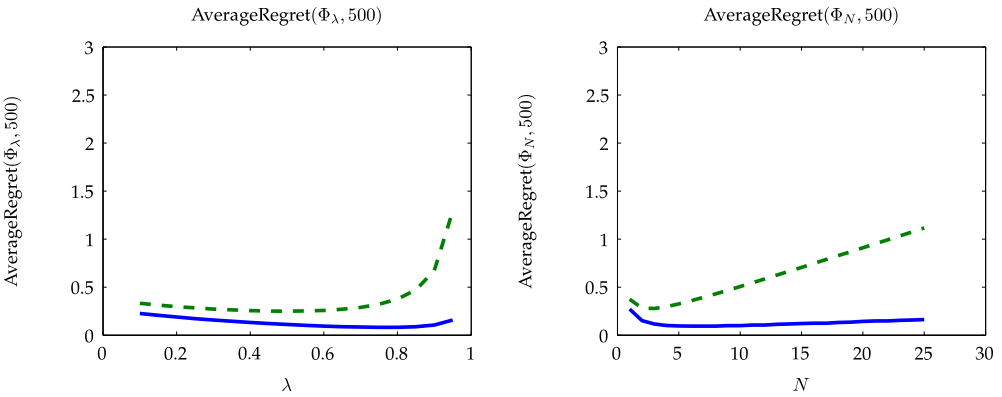

Figure 6.8: AverageRegret$(\Phi_\lambda, 500)$ and AverageRegret$(\Phi_N, 500)$ for experiment 3.

The optimal value of λ according to our upper bound equals $\lambda = 0.50$, with a corresponding upper bound on the regret of 0.25.The simulation average of AverageRegret$(\Phi_{0.50}, 500)$ was equal

to 0.11. The optimal value of λ according to the simulations, was $\lambda = 0.75$, with a simulation average of AverageRegret$(\Phi_{0.75}, 500)$ equal to 0.08.

The optimal value of N according to our upper bound equals $N = 3$, with a corresponding upper bound on the regret of 0.28. The simulation average of AverageRegret$(\Phi_3, 500)$ was equal to 0.12. The optimal value of N according to the simulations, was $N = 6$, with a simulation average of AverageRegret$(\Phi_6, 500)$ equal to 0.09.

Figure 6.8 illustrates that taking into account all available data (i.e. $\lambda = 1$ or $N = \infty$) would lead to much larger regret than obtained at the optimal λ and N. Thus, similar to scenario (ii), taking into account the changing nature of the market process leads to a significant profit improvement.

6.7 Proofs

Proof of Proposition 6.1

Equation (6.6) can be rewritten as

$$\hat{M}_\lambda(t) - M(t+1) = \frac{\sum_{i=1}^t \epsilon_i \lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}} + \frac{\sum_{i=1}^t (M(i) - M(t+1))\lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}}.$$

Note that $(\sum_{i=1}^t \lambda^{t-i})^{-1} = (1-\lambda^t)^{-1}(1-\lambda)\mathbf{1}(\lambda < 1) + \frac{1}{t}\mathbf{1}(\lambda = 1)$ and $E[\epsilon_i \epsilon_j] = E[\epsilon_i E[\epsilon_j \mid \mathcal{F}_i]] = 0$ whenever $i < j$. As a result,

$$E\left[\left|\sum_{i=1}^t \epsilon_i \lambda^{t-i}\right|^2\right] = \sum_{i=1}^t \lambda^{2(t-i)} E\left[\epsilon_i^2\right] \le \sigma^2 \left(\frac{1-\lambda^{2t}}{1-\lambda^2}\mathbf{1}(\lambda < 1) + t\mathbf{1}(\lambda = 1)\right),$$

and (6.9) follows using $|a+b|^2 \le 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$, and

$$\begin{aligned}&\left(\frac{1-\lambda^{2t}}{1-\lambda^2}\mathbf{1}(\lambda < 1) + t\mathbf{1}(\lambda = 1)\right)\left(\frac{1-\lambda}{1-\lambda^t}\mathbf{1}(\lambda < 1) + \frac{1}{t}\mathbf{1}(\lambda = 1)\right)^2 \\ =&\frac{1-\lambda}{1+\lambda}\frac{1+\lambda^t}{1-\lambda^t}\mathbf{1}(\lambda < 1) + \frac{1}{t}\mathbf{1}(\lambda = 1).\end{aligned}$$

If $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent, then $E[\epsilon_i M(j)] = 0$ for all $i, j \in \mathbb{N}$, and (6.11) follows from

$$\begin{aligned}E\left[\left|\hat{M}_\lambda(t) - M(t+1)\right|^2\right] &= E\left[\left|\frac{\sum_{i=1}^t \epsilon_i \lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}}\right|^2\right] + E\left[\left|\frac{\sum_{i=1}^t (M(i) - M(t+1))\lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}}\right|^2\right] \\ &\le \sigma^2 \left|\frac{\sum_{i=1}^t \lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}}\right|^2 + E\left[\left|\frac{\sum_{i=1}^t (M(i) - M(t+1))\lambda^{t-i}}{\sum_{i=1}^t \lambda^{t-i}}\right|^2\right],\end{aligned}$$

with equality if $(\epsilon_t)_{t\in\mathbb{N}}$ is homoscedastic.

Similarly, equation (6.8) can be rewritten as

$$\hat{M}_N(t) - M(t+1) = \frac{1}{\min\{N, t\}}\left(\sum_{i=1+(t-N)^+}^t \epsilon_i + \sum_{i=1+(t-N)^+}^t M(i) - M(t+1)\right).$$

Equation (6.10) follows using $|a+b|^2 \leq 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$, and by noting

$$E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} \epsilon_i\right|^2\right] = \frac{1}{\min\{N,t\}^2} \sum_{i=1+(t-N)^+}^{t} E\left[\epsilon_i^2\right] \leq \sigma^2 / \min\{N,t\}.$$

If $(\epsilon_t)_{t\in\mathbb{N}}$ and $(M(t))_{t\in\mathbb{N}}$ are independent, then $E[\epsilon_i M(j)] = 0$ for all $i, j \in \mathbb{N}$, and (6.12) follows from

$$\begin{aligned}
&E\left[\left|\hat{M}_N(t) - M(t+1)\right|^2\right] \\
&= E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} \epsilon_i\right|^2\right] + E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} M(i) - M(t+1)\right|^2\right] \\
&\leq \sigma^2 \left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t}\right|^2 + E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} M(i) - M(t+1)\right|^2\right],
\end{aligned}$$

with equality if $(\epsilon_t)_{t\in\mathbb{N}}$ is homoscedastic.

Proof of Theorem 6.1

We prove the theorem in two steps. In step 1, we show that there exists a $K_0 > 0$ such that for all $M \geq 0$, $M' \in \mathbb{R}$,

$$r(p^*(M), M) - r(p^*(M'), M) \leq K_0 (M - M')^2. \tag{6.23}$$

In step 2 we apply this result with $M = M(t)$, $M' = \hat{M}_\lambda(t)$ or $M' = \hat{M}_N(t)$, to obtain the regret bounds.

Step 1.

Fix $M \geq 0$, and let $r'(p, M)$ and $r''(p, M)$ denote the first and second derivative of $r(p, M)$ w.r.t. p. Let $M' \in \mathbb{R}$.

Case 1: $p^*(M) = p^\#(M)$. Then $r'(p^*(M), M) = 0$, and a Taylor series expansion yields

$$r(p, M) = r(p^*(M), M) + \frac{1}{2} r''(\tilde{p}, M)(p - p^*(M))^2,$$

for some $\tilde{p}$ on the line segment between p and $p^*(M)$. Let

$$K_1 = \sup_{p\in\mathcal{P}} |r''(p, M)| = \sup_{p\in\mathcal{P}} |2g'(p) + g''(p)|,$$

and note that K_1 is independent of M, and finite, because of the continuity of $g''(p)$. Then

$$r(p^*(M), M) - r(p, M) \leq \frac{K_1}{2}(p - p^*(M))^2 \quad \text{for all } p \in \mathcal{P}. \tag{6.24}$$

Write $h(p) = -g(p) - pg'(p)$, and note that $r'(p, M) = M - h(p)$. By assumption, for each $M \geq 0$ there is a unique $p^\#(M)$ such that $h(p) = M$, i.e. $p^\#(M) = h^{-1}(M)$ is well-defined. In addition,

for all $M \in h(\mathcal{P}) = \{h(p) \mid p \in \mathcal{P}\}$, we have $\frac{\partial}{\partial M} p^{\#}(M) = (h^{-1})'(M) = 1/h'(h^{-1}(M)) = -1/r''(p^*(M), M) > 0$. Thus, $p^{\#}(M)$ is continuous, differentiable, and monotone increasing on $M \in h(\mathcal{P})$. These properties imply the following: if there is an $M \geq 0$ s.t. $p^{\#}(M) > p_h$, then there is an $M_h > 0$ s.t. $h^{-1}(M) > p_h$ whenever $M > M_h$, $h^{-1}(M_h) = p_h$, and $h^{-1}(M) < p_h$ whenever $M < M_h$. Similarly, if there is an $M \geq 0$ s.t. $p^{\#}(M) < p_h$, then there is an $M_l \in (0, M_h)$ s.t. $h^{-1}(M) > p_l$ whenever $M > M_l$, $h^{-1}(M_l) = p_l$, and $h^{-1}(M) < p_l$ whenever $M < M_l$.

If $M' \geq 0$ and $p^*(M') = p^{\#}(M')$, then a Taylor expansion yields

$$|p^*(M') - p^*(M)| = |h^{-1}(M') - h^{-1}(M)| \leq |M' - M| K_2,$$

where $K_2 = \sup_{M \in h(\mathcal{P})} |(h^{-1})'(M)| = 1/\inf_{M \in h(\mathcal{P})} |r''(p^*(M), M)|$, which is finite by assumption.

If $M' \geq 0$ and $p^*(M') < p^{\#}(M')$, then $p^*(M') = p^*(M_h) = p_h$, $M' > M_h$, and

$$|p^*(M') - p^*(M)| = |p^*(M_h) - p^*(M)| \leq |M_h - M| K_2 \leq |M' - M| K_2.$$

If $M' \geq 0$ and $p^*(M') > p^{\#}(M')$, then $p^*(M') = p^*(M_l) = p_l$, $M' < M_l$, and

$$|p^*(M') - p^*(M)| = |p^*(M_l) - p^*(M)| \leq |M_l - M| K_2 \leq |M' - M| K_2,$$

It follows that $|p^*(M') - p^*(M)| \leq K_2 |M' - M|$ for all $M' \geq 0$, and thus by (6.24) we have

$$r(p^*(M), M) - r(p^*(M'), M) \leq \frac{1}{2} K_1 K_2^2 (M' - M)^2, \tag{6.25}$$

for all $M' \geq 0$, $M \in h(\mathcal{P})$. That this inequality is also valid when $M' < 0$, follows immediately from the observation

$$\begin{aligned} r(p^*(M), M) - r(p^*(M'), M) = r(p^*(M), M) - r(p^*(0), M) &\leq \frac{1}{2} K_1 K_2^2 (0 - M)^2 \\ &\leq \frac{1}{2} K_1 K_2^2 (M' - M)^2. \end{aligned}$$

Case 2: $p^*(M) \neq p^{\#}(M)$. Then $M \notin [M_l, M_h]$. Suppose $M > M_h$, the case $M < M_l$ is treated likewise. Suppose $M' \geq 0$. If $M' > M_h$ then $r(p^*(M), M) - r(p^*(M'), M) = 0$, suppose therefore $M' \leq M_h$. We have

$$\begin{aligned} & r(p^*(M), M) - r(p^*(M'), M) = r(p^*(M_h), M) - r(p^*(M'), M) \\ =& p^*(M_h)[M + g(p^*(M_h))] - p^*(M')[M + g(p^*(M'))] \\ =& r(p^*(M_h), M_h) - r(p^*(M'), M_h) + (p^*(M_h) - p^*(M'))(M - M_h) \\ \leq& \frac{1}{2} K_1 K_2^2 (M' - M_h)^2 + K_2 (M_h - M')(M - M_h) \\ \leq& \left(\frac{1}{2} K_1 K_2^2 + \frac{1}{4} K_2 \right) (M' - M)^2, \end{aligned}$$

where in the last inequality we use the fact $xy \leq \frac{1}{4}(x + y)^2$, $x, y \in \mathbb{R}$, with $x = M_h - M'$, $y = M - M_h$.

We have proven that for all $M \in \mathbb{R}$, $M' \geq 0$,

$$r(p^*(M), M) - r(p^*(M'), M) \leq \left(\frac{1}{2}K_1K_2^2 + \frac{1}{4}K_2\right)(M' - M)^2,$$

That this inequality is also valid when $M' < 0$, follows immediately from the observation

$$\begin{aligned} r(p^*(M), M) - r(p^*(M'), M) = r(p^*(M), M) - r(p^*(0), M) &\leq \left(\frac{1}{2}K_1K_2^2 + \frac{1}{4}K_2\right)(0 - M)^2 \\ &\leq \left(\frac{1}{2}K_1K_2^2 + \frac{1}{4}K_2\right)(M' - M)^2. \end{aligned}$$

This completes the proof of (6.23), with $K_0 = \frac{1}{2}K_1K_2^2 + \frac{1}{4}K_2$.

Step 2.

By Proposition 6.1, we obtain

$$\begin{aligned} &\text{AverageRegret}(\Phi_\lambda, T) \\ =&\frac{1}{T-1}\sum_{t=1}^{T-1} E\left[r(p^*(M(t+1)), M(t+1)) - r(p^*(\hat{M}_\lambda(t)), M(t+1))\right] \\ \leq&\frac{K_0}{T-1}\sum_{t=1}^{T-1} E\left[\left|\hat{M}_\lambda(t) - M(t+1)\right|^2\right] \\ \leq&\frac{2K_0}{T-1}\sum_{t=1}^{T-1}\left[\sigma^2\left[\frac{(1-\lambda)}{(1+\lambda)}\frac{(1+\lambda^t)}{(1-\lambda^t)}\mathbf{1}(\lambda < 1) + \frac{1}{t}\mathbf{1}(\lambda = 1)\right] + I_\lambda(t)\right]. \end{aligned}$$

Since

$$\begin{aligned} \sum_{t=1}^{T-1}\frac{\lambda^t}{1-\lambda^t} &= \frac{\lambda}{1-\lambda} + \sum_{t=2}^{T-1}\frac{\lambda^t}{1-\lambda^t} \leq \frac{\lambda}{1-\lambda} + \int_{t=1}^{T-2}\frac{\lambda^t}{1-\lambda^t}dt \\ &\leq \frac{\lambda}{1-\lambda} + \frac{-1}{\log(\lambda)}\int_{x=0}^{\lambda}\frac{1}{1-x}dx = \frac{\lambda}{1-\lambda} + \frac{\log(1-\lambda)}{\log(\lambda)}, \end{aligned}$$

we have for $\lambda < 1$,

$$\begin{aligned} \frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)}{(1+\lambda)}\frac{(1+\lambda^t)}{(1-\lambda^t)} &= \frac{1-\lambda}{1+\lambda} + \frac{2}{T-1}\frac{1-\lambda}{1+\lambda}\sum_{t=1}^{T-1}\frac{\lambda^t}{1-\lambda^t} \\ &\leq \frac{1-\lambda}{1+\lambda} + \frac{1}{T-1}\left(\frac{2\lambda}{1+\lambda} + 2\frac{1-\lambda}{1+\lambda}\frac{\log(1-\lambda)}{\log(\lambda)}\right), \end{aligned} \tag{6.26}$$

and thus

$$\begin{aligned}
&\text{AverageRegret}(\Phi_\lambda, T)\\
\leq& 2K_0\sigma^2\left[\frac{1-\lambda}{1+\lambda}+\frac{1}{T-1}\left(\frac{2\lambda}{1+\lambda}+2\frac{1-\lambda}{1+\lambda}\frac{\log(1-\lambda)}{\log(\lambda)}\right)\right]\mathbf{1}(\lambda<1)\\
&+2K_0\sigma^2\left[\frac{1+\log(T-1)}{T-1}\right]\mathbf{1}(\lambda=1)\\
&+\frac{2K_0}{T-1}\sum_{t=1}^{T-1}I_\lambda(t).
\end{aligned}$$

In addition, we have

$$\begin{aligned}
&\text{AverageRegret}(\Phi_N, T)\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1}E\left[r(p^*(M(t+1)),M(t+1))-r(p^*(\hat{M}_N(t)),M(t+1))\right]\\
\leq&\frac{K_0}{T-1}\sum_{t=1}^{T-1}E\left[\left|\hat{M}_N(t)-M(t+1)\right|^2\right]\\
\leq&\frac{2K_0}{T-1}\sum_{t=1}^{T-1}\left[\frac{\sigma^2}{\min\{N,t\}}+I_N(t)\right]\\
\leq&2K_0\sigma^2\left[\frac{\log(\min\{T-1,N\})}{T-1}+\frac{1}{\min\{N,T-1\}}\right]+\frac{2K_0}{T-1}\sum_{t=1}^{T-1}I_N(t),
\end{aligned}$$

where we used

$$\begin{aligned}
&\sum_{t=1}^{T-1}\frac{1}{\min\{N,t\}}=\sum_{t=1}^{N}\frac{1}{t}+\sum_{t=N+1}^{T-1}\frac{1}{N}\leq 1+\log(N)+\frac{T-1-N}{N}\quad\text{if } T-1\geq N,\\
&\sum_{t=1}^{T-1}\frac{1}{\min\{N,t\}}=\sum_{t=1}^{T-1}\frac{1}{t}\leq 1+\log(T-1)\quad\text{if } T-1<N,
\end{aligned}$$

and thus

$$\sum_{t=1}^{T-1}\frac{1}{\min\{N,t\}}\leq\log(\min\{T-1,N\})+\frac{T-1}{\min\{N,T-1\}}.$$

Proof of Proposition 6.2

The condition $M(t)\in[2bp_l,2bp_h]$ a.s., for all $t\in\mathbb{N}$, implies $p^*(M)=M/(2b)$ for all attainable values of M, and $r(p^*(M),M)-r(p^*(M'),M))=(M-M')^2/(4b)$ for all attainable values of M

and M'. By Proposition 6.1 we obtain

$$\begin{aligned}
&\text{LongRunAverageRegret}(\Phi_\lambda)\\
=&\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1}E\left[r(p^*(M(t+1)),M(t+1))-r(p^*(\hat{M}_\lambda(t)),M(t+1))\right]\\
=&\limsup_{T\to\infty}\frac{K_0}{T-1}\sum_{t=1}^{T-1}E\left[\left|\hat{M}_\lambda(t)-M(t+1)\right|^2\right]\\
=&\limsup_{T\to\infty}\frac{K_0}{T-1}\sum_{t=1}^{T-1}\left[\sigma^2\left[\frac{(1-\lambda)}{(1+\lambda)}\frac{(1+\lambda^t)}{(1-\lambda^t)}\mathbf{1}(\lambda<1)+\frac{1}{t}\mathbf{1}(\lambda=1)\right]+I_\lambda(t)\right]\\
=&K_0\left[\sigma^2\frac{(1-\lambda)}{(1+\lambda)}+\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}I_\lambda(t)\right],
\end{aligned}$$

and

$$\begin{aligned}
&\text{LongRunAverageRegret}(\Phi_N)\\
=&\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1}E\left[r(p^*(M(t+1)),M(t+1))-r(p^*(\hat{M}_N(t)),M(t+1))\right]\\
=&\limsup_{T\to\infty}\frac{K_0}{T-1}\sum_{t=1}^{T-1}E\left[\left|\hat{M}_N(t)-M(t+1)\right|^2\right]\\
=&\limsup_{T\to\infty}\frac{K_0}{T-1}\sum_{t=1}^{T-1}[\frac{\sigma^2}{\min\{N,t\}}+I_N(t)]\\
=&K_0\left[\sigma^2\frac{1}{N}+\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1}I_N(t)\right].
\end{aligned}$$

Calculation of $c_T(\lambda)$, $c(\lambda)$, $c_T(N)$ and $c(N)$ for scenario (2)

Let $\lambda \in [0,1)$. Then

$$\begin{aligned}
&\frac{1}{T-1}\sum_{t-1}^{T-1} I_\lambda(t)\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1} E\left[\left|\frac{1-\lambda}{1-\lambda^t}\sum_{i=1}^{t}(M(i)-M(t+1))\lambda^{t-i}\right|^2\right]\\
\leq&\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^2}{(1-\lambda^t)^2}\left|\sum_{i=1}^{t+1} d(t+1-i)\lambda^{t-i}\right|^2\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^2}{(1-\lambda^t)^2}d^2\left|(t+1)\sum_{i=1}^{t+1}\lambda^{t-i}-\sum_{i=1}^{t+1} i\lambda^{t-i}\right|^2\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^2}{(1-\lambda^t)^2}d^2\left|\frac{(t+1)(1-\lambda^{t+1})}{\lambda(1-\lambda)}-\frac{(t+1)}{\lambda(1-\lambda)}+\frac{(1-\lambda^{t+1})}{(1-\lambda)^2}\right|^2\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^2}{(1-\lambda^t)^2}d^2\left|-(t+1)(1-\lambda)^{-1}\lambda^t+(1-\lambda)^{-2}(1-\lambda^{t+1})\right|^2\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^{-2}}{(1-\lambda^t)^2}d^2\left|-(t+1)(1-\lambda)\lambda^t+(1-\lambda^{t+1})\right|^2\\
=&\frac{1}{T-1}(1-\lambda)^{-2}d^2\sum_{t=1}^{T-1}\left(\frac{\left(1-\lambda^t-t\lambda^t+t\lambda^{t+1}\right)}{(1-\lambda^t)}\right)^2\\
=&\frac{1}{T-1}(1-\lambda)^{-2}d^2\sum_{t=1}^{T-1}\left(1-t\lambda^t\frac{1-\lambda}{1-\lambda^t}\right)^2,
\end{aligned}$$

where we used $\sum_{i=1}^{t+1}(t+1)\lambda^{t-i}=(t+1)\lambda^{-1}(1-\lambda)^{-1}(1-\lambda^{t+1})$ and

$$\begin{aligned}
\sum_{i=1}^{t+1} i\lambda^{t-i}&=\sum_{i=1}^{t+1}\lambda^{t-i}\sum_{j=1}^{i}1=\sum_{j=1}^{t+1}\sum_{i=j}^{t+1}\lambda^{t-i}=\sum_{j=1}^{t+1}\big(\lambda^{-1}\sum_{i=1}^{t+1}\lambda^{t+1-i}-\lambda^{t-j+1}\sum_{i=1}^{j-1}\lambda^{j-1-i}\big)\\
&=\sum_{j=1}^{t+1}\big(\lambda^{-1}(1-\lambda)^{-1}(1-\lambda^{t+1})-\lambda^{t-j+1}(1-\lambda)^{-1}(1-\lambda^{j-1})\big)\\
&=(t+1)\lambda^{-1}(1-\lambda)^{-1}(1-\lambda^{t+1})-(1-\lambda)^{-1}\sum_{j=1}^{t+1}\big(\lambda^{t-j+1}-\lambda^t\big)\big)\\
&=(t+1)\lambda^{-1}(1-\lambda)^{-1}(1-\lambda^{t+1})-(1-\lambda)^{-1}\big((1-\lambda)^{-1}(1-\lambda^{t+1})-(t+1)\lambda^t\big)\\
&=(t+1)\lambda^{-1}(1-\lambda)^{-1}-(1-\lambda)^{-2}(1-\lambda^{t+1}).
\end{aligned}$$

We obtain

$$c_T(\lambda)=(1-\lambda)^{-2}d^2\frac{1}{T-1}\sum_{t=1}^{T-1}\left(1-t\lambda^t\frac{1-\lambda}{1-\lambda^t}\right)^2,$$

and taking the $\limsup_{T\to\infty}$ yields $c(\lambda) = (1-\lambda)^{-2}d^2$. For $\lambda = 1$,

$$\begin{aligned}
&\frac{1}{T-1}\sum_{t-1}^{T-1} I_\lambda(t)\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1} E\left[\left|\frac{1}{t}\sum_{i=1}^{t}(M(i)-M(t+1))\right|^2\right]\\
\leq&\frac{1}{T-1}\sum_{t=1}^{T-1}\left|\frac{1}{t}\sum_{i=1}^{t} d(t+1-i)\right|^2 = \frac{1}{T-1}\sum_{t=1}^{T-1} d^2\left|\frac{1}{t}\sum_{i=1}^{t} i\right|^2\\
=&\frac{1}{T-1}\sum_{t=1}^{T-1} d^2\frac{(t+1)^2}{4},
\end{aligned}$$

and $\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1} d^2\frac{(t+1)^2}{4} = \infty$.

Let $N \in \mathbb{N}_{\geq 2}$, then

$$\begin{aligned}
&\frac{1}{T-1}\sum_{t=1}^{T-1} I_N(t)\\
&= \frac{1}{T-1}\sum_{t=1}^{T-1} E\left[\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{t}(M(i)-M(t+1))\right|^2\right]\\
&\leq \frac{1}{T-1}\sum_{t=1}^{T-1}\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{t} d(t+1-i)\right|^2\\
&= \frac{1}{T-1}\sum_{t=1}^{T-1}\left|\frac{d}{\min\{N,t\}}\sum_{j=1}^{\min\{N,t\}} j\right|^2 = \frac{1}{T-1}\sum_{t=1}^{T-1}\frac{1}{4}d^2(\min\{N,t\}+1)^2\\
&= \frac{d^2}{4}\frac{1}{T-1}\sum_{t=1}^{T-1}(N+1)^2 + \frac{d^2}{4}\frac{1}{T-1}\sum_{t=1}^{\min\{T-1,N-1\}}[(t+1)^2-(N+1)^2]\\
&= \frac{d^2}{4}(N+1)^2 + \frac{d^2}{4}\frac{1}{T-1}\Big[-\min\{T-1,N-1\}(N+1)^2 + \sum_{t=2}^{\min\{T,N\}} t^2\Big]\\
&= \frac{d^2}{4}(N+1)^2 + \frac{d^2}{4}\frac{1}{T-1}\cdot\\
&\Big[(1-\min\{T,N\})(N+1)^2 - 1 + \min\{T,N\}(\min\{T,N\}+1)(2\min\{T,N\}+1)/6\Big],
\end{aligned}$$

where we used $\sum_{t=1}^{N} t^2 = N(N+1)(2N+1)/6$. After some algebraic manipulations, we derive that

$$c_T(N) = \begin{cases} \frac{1}{4}d^2\frac{-1+T(T+1)(2T+1)/6}{T-1} & \text{if } T < N\\ \frac{1}{4}d^2\Big[(N+1)^2 + \frac{1}{T-1}N(-4N^2-3N+7)/6\Big] & \text{if } T \geq N \end{cases}.$$

Taking $\limsup_{T\to\infty}$, we obtain $c(N) = \frac{1}{4}d^2(N+1)^2$. For $N = \infty$,

$$\frac{1}{T-1}\sum_{t=1}^{T-1} I_N(t) = \frac{1}{T-1}\sum_{t=1}^{T-1} E\left[\left|\frac{1}{t}\sum_{i=1}^{t}(M(i) - M(t+1))\right|^2\right]$$
$$\leq \frac{1}{T-1}\sum_{t=1}^{T-1}\left|\frac{1}{t}\sum_{i=1}^{t} d(t+1-i)\right|^2 = \frac{1}{T-1}\sum_{t=1}^{T-1} d^2\frac{(t+1)^2}{4},$$

and $\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1} d^2\frac{(t+1)^2}{4} = \infty$.

Calculation of $c_T(\lambda)$, $c(\lambda)$, $c_T(N)$ and $c(N)$ for scenario (3)

For $t \in \mathbb{N}$, define

$$X(t) = \min\{k \in \{1,\ldots,t+1\} \mid M(k) = M(k+1) = \ldots = M(t+1)\},$$

and note that $P(X(t) = k) \leq P(M(k-1) \neq M(k)) \leq \epsilon$. for all $k = 2,\ldots,t+1$.

For $\lambda \in [0,1)$,

$$\begin{aligned}
&E\left[\left|\sum_{i=1}^{t}(M(i) - M(t+1))\lambda^{t-i}\right|^2\right]\\
=&\sum_{k=1}^{t+1} E\left[\left|\sum_{i=1}^{t}(M(i) - M(t+1))\lambda^{t-i}\right|^2 \mid X(t) = k\right] P(X(t) = k)\\
\leq&\sum_{k=2}^{t+1} E\left[\left|\sum_{i=1}^{k-1}(M(i) - M(t+1))\lambda^{t-i}\right|^2 \mid X(t) = k\right]\epsilon\\
\leq&\sum_{k=2}^{t+1} d^2\left|\sum_{i=1}^{k-1}\lambda^{t-i}\right|^2\epsilon\\
=&\sum_{k=2}^{t+1} d^2\left|\lambda^{t-(k-1)}(1-\lambda)^{-1}(1-\lambda^{k-1})\right|^2\epsilon\\
=&\sum_{k=1}^{t} d^2\lambda^{2(t-k)}(1-\lambda)^{-2}(1-2\lambda^k+\lambda^{2k})\epsilon\\
=&d^2\epsilon(1-\lambda)^{-2}[(1-\lambda^2)^{-1}(1-\lambda^{2t}) - 2\lambda^t(1-\lambda)^{-1}(1-\lambda^t) + t\lambda^{2t}],
\end{aligned}$$

and thus

$$\begin{aligned}
c(\lambda) &= \limsup_{T\to\infty} \frac{1}{T-1}\sum_{t-1}^{T-1} I_\lambda(t) \\
&= \limsup_{T\to\infty} \frac{1}{T-1}\sum_{t=1}^{T-1} E\left[\left|\frac{1-\lambda}{1-\lambda^t}\sum_{i=1}^{t}(M(i)-M(t+1))\lambda^{t-i}\right|^2\right] \\
&\le \limsup_{T\to\infty} \frac{1}{T-1}\sum_{t=1}^{T-1} \frac{1}{(1-\lambda^t)^2} d^2\epsilon[(1-\lambda^2)^{-1}(1-\lambda^{2t}) - 2\lambda^t(1-\lambda)^{-1}(1-\lambda^t) + t\lambda^{2t}] \\
&= d^2\epsilon(1-\lambda^2)^{-1}.
\end{aligned}$$

For $\lambda = 1$,

$$\begin{aligned}
&\frac{1}{t}E\left[\left|\sum_{i=1}^{t}(M(i)-M(t+1))\right|^2\right] \\
=&\sum_{k=1}^{t+1}\frac{1}{t}E\left[\left|\sum_{i=1}^{t}(M(i)-M(t+1))\right|^2 \mid X(t)=k\right]P(X(t)=k) \\
\le&\sum_{k=2}^{t+1}\frac{1}{t}d^2(k-1)^2\epsilon = \sum_{k=1}^{t}\frac{1}{t}d^2k^2\epsilon = d^2\epsilon(t+1)(2t+1)/6,
\end{aligned}$$

and $\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1} d^2\epsilon(t+1)(2t+1)/6 = \infty$.

Let $N \in \mathbb{N}_{\ge 2}$, then

$$\begin{aligned}
I_N(t) &= E\left[\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{t}(M(i)-M(t+1))\right|^2\right] \\
&= \sum_{k=1}^{t+1} E\left[\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{t}(M(i)-M(t+1))\right|^2 \mid X(t)=k\right]P(X(t)=k) \\
&\le \sum_{k=2}^{t+1}\left|\frac{1}{\min\{N,t\}}\sum_{i=1+(t-N)^+}^{k-1} d\right|^2 \epsilon \\
&= d^2\epsilon\sum_{k=1+(t-N)^+}^{t}\left(\frac{k-(t-N)^+}{\min\{N,t\}}\right)^2 \\
&= d^2\epsilon\sum_{k=1}^{\min\{N,t\}}\left(\frac{k}{\min\{N,t\}}\right)^2 \\
&= d^2\epsilon\frac{(\min\{N,t\}+1)(2\min\{N,t\}+1)}{6\min\{N,t\}},
\end{aligned}$$

and thus

$$\begin{aligned} c(N) =& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} I_N(t) \\ \leq& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} d^2\epsilon \frac{(\min\{N,t\}+1)(2\min\{N,t\}+1)}{6\min\{N,t\}} \\ =& d^2\epsilon \frac{(N+1)(2N+1)}{6N}. \end{aligned}$$

If $N = \infty$, then

$$\begin{aligned} & \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} I_N(t) \\ =& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} E\left[\left|\frac{1}{t}\sum_{i=1}^{t}(M(i)-M(t+1))\right|^2\right] \\ =& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1}\sum_{k=1}^{t+1} E\left[\left|\frac{1}{t}\sum_{i=1}^{t}(M(i)-M(t+1))\right|^2 \mid X(t)=k\right] P(X(t)=k) \\ \leq& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1}\sum_{k=2}^{t+1} \left|\frac{1}{t}\sum_{i=1}^{k-1} d\right|^2 \epsilon \\ =& \infty. \end{aligned}$$

Calculation of $c_T(\lambda)$, $c(\lambda)$, $c_T(N)$ and $c(N)$ for scenario (4)

Let $\lambda \in [0,1)$. Then

$$\begin{aligned}
&E\left[\left|\sum_{i=1}^{t}(M(i)-M(t+1))\lambda^{t-i}\right|^2\right]\\
=&E\left[\sum_{i=1}^{t}(M(i)-M(t+1))^2\lambda^{2(t-i)}\right]\\
+&E\left[\sum_{i\neq j,1\leq i,j\leq t}(M(i)-M(t+1))\lambda^{t-i}(M(j)-M(t+1))\lambda^{t-j}\right]\\
=&\sum_{i=1}^{t}E\left[M(i)^2+M(t+1)^2-2M(i)M(t+1)\right]\lambda^{2(t-i)}\\
+&\sum_{i\neq j,1\leq i,j\leq t}E\left[M(i)M(j)-M(i)M(t+1)-M(t+1)M(j)+M(t+1)^2\right]\lambda^{2t-i-j}\\
=&(E\left[X^2\right]-E\left[X\right]^2)\sum_{i=1}^{t}2\lambda^{2(t-i)}+\sum_{i\neq j,1\leq i,j\leq t}(E\left[X^2\right]-E\left[X\right]^2)\lambda^{2t-i-j}\\
=&(E\left[X^2\right]-E\left[X\right]^2)\left(\sum_{i=1}^{t}\lambda^{2(t-i)}+\sum_{i=1}^{t}\lambda^{t-i}\sum_{j=1}^{t}\lambda^{t-j}\right)\\
=&\mathrm{Var}(X)\left((1-\lambda^2)^{-1}(1-\lambda^{2t})+(1-\lambda)^{-2}(1-\lambda^t)^2\right),
\end{aligned}$$

and thus

$$\begin{aligned}
c(\lambda)&=\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1}I_\lambda(t)\\
&=\limsup_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T-1}\frac{(1-\lambda)^2}{(1-\lambda^t)^2}\mathrm{Var}(X)\left((1-\lambda^2)^{-1}(1-\lambda^{2t})+(1-\lambda)^{-2}(1-\lambda^t)^2\right)\\
&=\mathrm{Var}(X)\left((1-\lambda)^2(1-\lambda^2)^{-1}+1\right)\\
&=\mathrm{Var}(X)\frac{2}{1+\lambda}.
\end{aligned}$$

If $\lambda = 1$ then

$$\begin{aligned}
&E\left[\left|\sum_{i=1}^{t}(M(i)-M(t+1))\lambda^{t-i}\right|^2\right]\\
=&(E\left[X^2\right]-E\left[X\right]^2)\left(\sum_{i=1}^{t}1+\sum_{i=1}^{t}\sum_{j=1}^{t}1\right)\\
=&\mathrm{Var}(X)(t+t^2),
\end{aligned}$$

and

$$\begin{aligned} c(1) &= \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} I_\lambda(t) \\ &= \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{1}{t^2} \mathrm{Var}(X)(t+t^2) \\ &= \mathrm{Var}(X). \end{aligned}$$

Let $N \in \mathbb{N}_{\geq 2} \cup \{\infty\}$, then

$$\begin{aligned} & E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} (M(i) - M(t+1))\right|^2\right] \\ =& E\left[\frac{1}{\min\{N,t\}^2} \sum_{i=1+(t-N)^+}^{t} (M(i) - M(t+1))^2\right. \\ &\left. + \frac{1}{\min\{N,t\}^2} \sum_{i\neq j, 1+(t-N)^+\leq i,j\leq t} (M(i) - M(t+1))(M(j) - M(t+1))\right] \\ =& \frac{1}{\min\{N,t\}^2} \sum_{i=1+(t-N)^+}^{t} E\left[M(i)^2 + M(t+1)^2 - 2M(i)M(t+1)\right] \\ &+ \frac{1}{\min\{N,t\}^2} \sum_{i\neq j, 1+(t-N)^+\leq i,j\leq t} E\left[M(i)M(j) - M(i)M(t+1) - M(t+1)M(j) + M(t+1)^2\right] \\ =& \frac{2}{\min\{N,t\}}(E\left[X^2\right] - E\left[X\right]^2) + \frac{\min\{N,t\}-1}{\min\{N,t\}}(E\left[X^2\right] - E\left[X\right]^2) \\ =& \left(1 + \frac{1}{\min\{N,t\}}\right) \mathrm{Var}(X), \end{aligned}$$

and thus

$$\begin{aligned} c(N) =& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} E\left[\left|\frac{1}{\min\{N,t\}} \sum_{i=1+(t-N)^+}^{t} (M(i) - M(t+1))\right|^2\right] \\ =& \limsup_{T\to\infty} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(1 + \frac{1}{\min\{N,t\}}\right) \mathrm{Var}(X) \\ =& \mathrm{Var}(X) \left(1 + \frac{1}{N}\right). \end{aligned}$$

Chapter 7

Mean square convergence rates for maximum quasi-likelihood estimators

7.1 Introduction

7.1.1 Motivation

We consider a statistical model of the form

$$E[Y(x)] = h(x^T\beta^{(0)}), \quad \text{Var}(Y(x)) = v(E[Y(x)]), \tag{7.1}$$

where $x \in \mathbb{R}^d$ is a design variable, $Y(x)$ is a random variable whose distribution depends on x, $\beta^{(0)} \in \mathbb{R}^d$ is an unknown parameter, and h and v are known functions on $\mathbb{R}$. Such models arise, for example, from generalized linear models (GLMs), where in addition to (7.1) one requires that the distribution of $Y(x)$ comes from the exponential family (cf. Nelder and Wedderburn (1972), McCullagh and Nelder (1983), Gill (2001)). We are interested in making inference on the unknown parameter $\beta^{(0)}$.

In GLMs, this is commonly done via maximum-likelihood estimation. Given a sequence of design variables $(x_i)_{1\leq i\leq n}$ and observed responses $(y_i)_{1\leq i\leq n}$, where each y_i is a realization of the random variable $Y(x_i)$, the maximum-likelihood estimator (MLE) $\hat{\beta}_n$ is a solution to the equation $l_n(\beta) = 0$, where $l_n(\beta)$ is defined as

$$l_n(\beta) = \sum_{i=1}^{n} \frac{\dot{h}(x_i^T\beta)}{v(h(x_i^T\beta))} x_i(y_i - h(x_i^T\beta)), \tag{7.2}$$

and where $\dot{h}$ denotes the derivative of h.

As discussed by Wedderburn (1974) and McCullagh (1983), if one drops the requirement that the distribution of $Y(x)$ is a member of the exponential family, and only assumes (7.1), one can still make inference on β by solving $l_n(\beta) = 0$. The solution $\hat{\beta}_n$ is then called a maximum quasi-likelihood estimator (MQLE) of $\beta^{(0)}$.

In this chapter, we are interested in the quality of the estimate $\hat{\beta}_n$ for models satisfying (7.1) by considering the expected value of $||\hat{\beta}_n - \beta^{(0)}||^2$, where $||\cdot||$ denotes the Euclidean norm. We derive conditions such that $\hat{\beta}_n$ converges a.s. to $\beta^{(0)}$ as t grows large, and obtain bounds on the mean square convergence rates. These bounds are key in proving regret bounds for the pricing policies considered in Chapters 3,4 and 5.

7.1.2 Literature

Although much literature is devoted to the (asymptotic) behavior of maximum (quasi-)likelihood estimators for models of the form (7.1), practically all of them focus on a.s. upper bounds on $||\hat{\beta}_n - \beta^{(0)}||$ instead of mean square bounds. The literature may be classified according to the following criteria:

1. Assumptions on (in)dependence of design variables and error terms.
   The sequence of vectors $(x_i)_{i\in\mathbb{N}}$ is called the design, and the error terms $(e_i)_{i\in\mathbb{N}}$ are defined as
   $$e_i = y_i - h(x_i^T\beta^{(0)}), \quad (i \in \mathbb{N}).$$
   Typically, one either assumes a fixed design, with all x_i non-random and the e_i mutually independent, or an adaptive design, where the sequence $(e_i)_{i\in\mathbb{N}}$ forms a martingale difference sequence w.r.t. its natural filtration and where the design variables $(x_i)_{i\in\mathbb{N}}$ are predictable w.r.t. this filtration. This last setting is appropriate for sequential decision problems under uncertainty, where decisions are made based on current parameter-estimates.

2. Assumptions on the dispersion of the design vectors.
   Define the design matrix
   $$P_n = \sum_{i=1}^{n} x_i x_i^T, \tag{7.3}$$
   and denote by $\lambda_{\min}(P_n)$, $\lambda_{\max}(P_n)$ the smallest and largest eigenvalues of P_n. Bounds on $||\hat{\beta}_n - \beta^{(0)}||$ are typically stated in terms of these two eigenvalues, which in some sense quantify the amount of dispersion in the sequence $(x_i)_{i\in\mathbb{N}}$.

3. Assumptions on the link function.
   In GLM terminology, h^{-1} is called the link function. It is called *canonical* or *natural* if $\dot{h} = v \circ h$, otherwise it is called a *general* or *non-canonical* link function. For canonical link functions, the quasi-likelihood equations (7.2) simplify to $l_n(\beta) = \sum_{i=1}^{n} x_i(y_i - h(x_i^T\beta)) = 0$.

To these three sets of assumptions, one usually adds smoothness conditions on h and v, and assumptions on the moments of the error terms.

An early result on the asymptotic behavior of solutions to (7.2), is from Fahrmeir and Kaufmann (1985). For fixed design and canonical link function, provided $\lambda_{\min}(P_n) = \Omega(\lambda_{\max}(P_n)^{1/2+\delta})$ a.s. for a $\delta > 0$ and some other regularity assumptions, they prove asymptotic existence and strong consistency of $(\hat{\beta}_n)_{n\in\mathbb{N}}$ (their Corollary 1; for the definition of $\Omega(\cdot)$, $O(\cdot)$ and $o(\cdot)$, see the next paragraph on notation). For general link functions, these results are proven assuming $\lambda_{\min}(P_n) = \Omega(\lambda_{\max}(P_n))$ a.s. and some other regularity conditions (their Theorem 5). Chen et al. (1999) consider only canonical link functions. In the fixed design case, they obtain strong consistency and convergence rates

$$||\hat{\beta}_n - \beta^{(0)}|| = o(\{(\log(\lambda_{\min}(P_n)))^{1+\delta}/\lambda_{\min}(P_n)\}^{1/2}) \text{ a.s.},$$

for any $\delta > 0$; in the adaptive design case, they obtain convergence rates

$$||\hat{\beta}_n - \beta^{(0)}|| = O(\{(\log(\lambda_{\max}(P_n))/\lambda_{\min}(P_n)\}^{1/2}) \text{ a.s.}$$

Their proof however is reported to contain a mistake, see Zhang and Liao (2008, page 1289). Chang (1999) extends these convergence rates for adaptive designs to general link functions, under the additional condition $\lambda_{\min}(P_n) = \Omega(n^{\alpha})$ a.s. for some $\alpha > 1/2$. His proof however also

appears to contain a mistake, see Remark 7.1. Yin et al. (2008) extends the setting of Chang (1999), with adaptive design and general link function, to multivariate response data. They obtain strong consistency and convergence rates

$$||\hat{\beta}_n - \beta^{(0)}|| = o(\{\lambda_{\max}(P_n)\log(\lambda_{\max}(P_n))\}^{1/2}\{\log(\log(\lambda_{\max}(P_n)))\}^{1/2+\delta}/\lambda_{\min}(P_n)) \text{ a.s.},$$

for $\delta > 0$, under assumptions on $\lambda_{\min}(P_n), \lambda_{\max}(P_n)$ that ensure that this asymptotic upper bound is $o(1)$ a.s. A recent study restricted to fixed designs and canonical link functions is Zhang and Liao (2008), who show $||\hat{\beta}_n - \beta^{(0)}|| = O_p(\lambda_{\min}(P_n)^{-1/2})$, provided $\lambda_{\min}(P_n) = \Omega(\lambda_{\max}(P_n)^{1/2})$ a.s. and other regularity assumptions.

7.1.3 Assumptions and contributions

In contrast with the above-mentioned literature, we study bounds for the expected value of $||\hat{\beta}_n - \beta^{(0)}||^2$. The design is assumed to be adaptive; i.e. the error terms $(e_i)_{i\in\mathbb{N}}$ form a martingale difference sequence w.r.t. the natural filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$, and the design variables $(x_i)_{i\in\mathbb{N}}$ are predictable w.r.t. this filtration. For applications of our results to sequential decision problems, where each new decision can depend on the most recent parameter estimate, this is the appropriate setting to consider. In addition, we assume $\sup_{i\in\mathbb{N}} E\left[e_i^2 \mid \mathcal{F}_{i-1}\right] \leq \sigma^2 < \infty$ a.s. for some $\sigma > 0$, and $\sup_{i\in\mathbb{N}} E\left[|e_i|^r\right] < \infty$ for some $r > 2$.

We consider general link functions, and only assume that h and v are thrice continuously differentiable with $\dot{h}(z) > 0$, $v(h(z)) > 0$ for all $z \in \mathbb{R}$. Concerning the design vectors $(x_i)_{i\in\mathbb{N}}$, we assume that they are contained in a bounded subset $X \subset \mathbb{R}^d$. Let $\lambda_1(P_n) \leq \lambda_2(P_n)$ denote the two smallest eigenvalues of the design matrix P_n (if the dimension d of $\beta^{(0)}$ equals 1, write $\lambda_2(P_n) = \lambda_1(P_n)$). We assume that there is a (non-random) $n_0 \in \mathbb{N}$ such that P_{n_0} is invertible, and there are (non-random) functions L_1, L_2 on $\mathbb{N}$ such that for all $n \geq n_0$: $\lambda_1(P_n) \geq L_1(n)$, $\lambda_2(P_n) \geq L_2(n)$, and

$$L_1(n) \geq cn^{\alpha}, \quad \text{for some } c > 0, \frac{1}{2} < \alpha \leq 1 \text{ independent of } n. \tag{7.4}$$

Based on these assumptions, we obtain three important results concerning the asymptotic existence of $\hat{\beta}_n$ and bounds on $E[||\hat{\beta}_n - \beta^{(0)}||^2]$:

1. First, notice that a solution to (7.2) need not always exist. Following Chang (1999), we therefore define the last-time that there is no solution in a neighborhood of $\beta^{(0)}$:

   $$N_\rho = \sup\left\{n \geq n_0 : \text{ there exists no } \beta \in \mathbb{R}^d \text{ with } l_n(\beta) = 0 \text{ and } \left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right| \leq \rho\right\}.$$

   For all sufficiently small $\rho > 0$, we show in Theorem 7.1 that N_ρ is finite a.s., and provide sufficient conditions such that $E[N_\rho^\eta] < \infty$, for $\eta > 0$.

2. In Theorem 7.2, we provide the upper bound

   $$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)} + \frac{n(d-1)^2}{L_1(n)L_2(n)}\right), \tag{7.5}$$

   where $\mathbf{1}_{n>N_\rho}$ denotes the indicator function of the event $\{n > N_\rho\}$.

3. In case of a canonical link function, Theorem 7.3 improves these bounds to

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)}\right). \tag{7.6}$$

This improvement clearly is also valid for general link functions provided $d = 1$. It also holds if $d = 2$ and $||x_i||$ is bounded from below by a positive constant (see Remark 7.2).

An important intermediate result in proving these bounds is Proposition 7.2, where we derive

$$E\left|\left|\left(\sum_{i=1}^n x_i x_i^T\right)^{-1} \sum_{i=1}^n x_i e_i\right|\right|^2 = O\left(\frac{\log(n)}{L(n)}\right),$$

for any function L that satisfies $\lambda_{\min}\left(\sum_{i=1}^n x_i x_i^T\right) \geq L(n) > 0$ for all sufficiently large n. This actually provides bounds on mean square convergence rates in least-squares linear regression, and forms the counterpart of Lai and Wei (1982) who prove similar bounds in an a.s. setting.

7.1.4 Applications

A useful application of Theorems 7.1 and 7.2 is the derivation of upper bounds of quadratic cost functions in β. For example, let $c(\beta)$ be a non-negative bounded function with $||c(\beta) - c(\beta^{(0)})|| \leq K||\beta - \beta^{(0)}||^2$ for all $\beta \in \mathbb{R}^d$ and some $K > 0$. Application of Theorems 7.1 and 7.2 yield the upper bound

$$\begin{aligned}
E\left[\left|\left|c(\hat{\beta}_n) - c(\beta^{(0)})\right|\right|\right] &\leq E\left[\left|\left|c(\hat{\beta}_n) - c(\beta^{(0)})\right|\right| \mathbf{1}_{n>N_\rho}\right] + E\left[\left|\left|c(\hat{\beta}_n) - c(\beta^{(0)})\right|\right| \mathbf{1}_{n\leq N_\rho}\right] \\
&\leq K \cdot E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] + \frac{E\left[N_\rho^\eta\right]}{n^\eta} \max_\beta \left|\left|c(\beta) - c(\beta^{(0)})\right|\right|^2 \\
&= O\left(\frac{\log(n)}{L_1(n)} + n^{-\eta}\right).
\end{aligned}$$

In dynamic pricing problems, such arguments are used to design decision rules and derive upper bounds on the regret, cf. Chapters 3, 4, and 5. These type of arguments can also be applied to other sequential decision problems with parametric uncertainty where the objective is to minimize the regret; for example, the multiperiod inventory control problem (Anderson and Taylor (1976), Lai and Robbins (1982)), or parametric variants of bandit problems (Goldenshluger and Zeevi (2009), Rusmevichientong and Tsitsiklis (2010)).

In his review on experimental design and control problems, Pronzato (2008, page 18, Section 9) mentions that existing consistency results for adaptive design of experiments are usually restricted to models that are linear in the parameters. The class of statistical models that we consider is much larger than only linear models; it includes all models satisfying (7.1). Our results may therefore also find application in the field of sequential design of experiments.

7.1.5 Organization of the chapter

The rest of this chapter is organized as follows. Section 7.2 contains our results concerning the last-time N_ρ and upper bounds on $E[||\hat{\beta}_n - \beta^{(0)}||^2 \mathbf{1}_{n>N_\rho}]$, for general link functions. In Section 7.3 we derive these bounds in the case of canonical link functions. Section 7.4 contains the proofs

of the assertions in Section 7.2 and 7.3. In the appendix, Section 7.5, we collect and prove several auxiliary results which are used in the proofs of the theorems of Sections 7.2 and 7.3.

Notation. For $\rho > 0$, let $B_\rho = \{\beta \in \mathbb{R}^d \mid \left|\left|\beta - \beta^{(0)}\right|\right| \leq \rho\}$ and $\partial B_\rho = \{\beta \in \mathbb{R}^d \mid \left|\left|\beta - \beta^{(0)}\right|\right| = \rho\}$. The closure of a set $S \subset \mathbb{R}^d$ is denoted by $\bar{S}$, the boundary by $\partial S = \bar{S} \backslash S$. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer that does not exceed x. The Euclidean norm of a vector y is denoted by $||y||$. The norm of a matrix A equals $||A|| = \max_{z:||z||=1} ||Az||$. The 1-norm and ∞-norm of a matrix are denoted by $||A||_1$ and $||A||_\infty$. y^T denotes the transpose of a vector or matrix y. If $f(x), g(x)$ are functions with domain in $\mathbb{R}$ and range in $(0, \infty)$, then $f(x) = O(g(x))$ means there exists a $K > 0$ such that $f(x) \leq Kg(x)$ for all $x \in \mathbb{N}$, $f(x) = \Omega(g(x))$ means $g(x) = O(f(x))$, and $f(x) = o(g(x))$ means $\lim_{x\to\infty} f(x)/g(x) = 0$.

7.2 Results for general link functions

In this section we consider the statistical model introduced in Section 7.1.1 for general link functions h, under all the assumptions listed in Section 7.1.3. The first main result is Theorem 7.1, which shows finiteness of moments of N_{ρ_0}. The second main result is Theorem 7.2, which proves asymptotic existence and strong consistency of the MQLE, and provides bounds on the mean square convergence rates.

Our results on the existence of the quasi-likelihood estimate $\hat{\beta}_n$ are based on the following fact, which is a consequence of the Leray-Schauder theorem (Leray and Schauder, 1934).

Lemma 7.1 (Ortega and Rheinboldt, 2000, 6.3.4, page 163)**.** *Let C be an open bounded set in $\mathbb{R}^n$, $F : \bar{C} \to \mathbb{R}^n$ a continuous mapping, and $(x - x_0)^T F(x) \geq 0$ for some $x_0 \in C$ and all $x \in \partial C$. Then $F(x) = 0$ has a solution in $\bar{C}$.*

This lemma yields a sufficient condition for the existence of $\hat{\beta}_n$ in the proximity of $\beta^{(0)}$ (recall the definitions $B_\rho = \{\beta \in \mathbb{R}^d \mid \left|\left|\beta - \beta^{(0)}\right|\right| \leq \rho\}$ and $\partial B_\rho = \{\beta \in \mathbb{R}^d \mid \left|\left|\beta - \beta^{(0)}\right|\right| = \rho\}$):

Corollary 7.1. *For all $\rho > 0$, if $\sup_{\beta\in\partial B_\rho}(\beta - \beta^{(0)})^T l_n(\beta) \leq 0$ then there exists a $\beta \in B_\rho$ with $l_n(\beta) = 0$.*

A first step in applying Corollary 7.1 is to provide an upper bound for $(\beta - \beta^{(0)})^T l_n(\beta)$. To this end, write $g(x) = \frac{\dot{h}(x)}{v(h(x))}$, and choose a $\rho_0 > 0$ such that $(c_2 - c_1 c_3 \rho) \geq c_2/2$ for all $0 < \rho \leq \rho_0$, where

$$c_1 = \sup_{\substack{x\in X,\\ \beta\in B_{\rho_0}}} \frac{1}{2}|\ddot{g}(x^T\beta)|\,||x||\,, \quad c_2 = \inf_{\substack{x\in X,\\ \beta,\tilde{\beta}\in B_{\rho_0}}} g(x^T\beta)\dot{h}(x^T\tilde{\beta}), \quad c_3 = \sup_{i\in\mathbb{N}} E[|e_i| \mid \mathcal{F}_{i-1}]. \tag{7.7}$$

The existence of such a ρ_0 follows from the fact that $\dot{h}(x) > 0$ and $g(x) > 0$ for all $x \in \mathbb{R}$.

Lemma 7.2. *Let $0 < \rho \leq \rho_0$, $\beta \in B_\rho$, $n \in \mathbb{N}$, and define*

$$A_n = \sum_{i=1}^n g(x_i^T\beta^{(0)})x_i e_i, \quad B_n = \sum_{i=1}^n \dot{g}(x_i^T\beta^{(0)})x_i x_i^T e_i, \quad J_n = c_1 \sum_{i=1}^n (|e_i| - E[|e_i| \mid \mathcal{F}_{i-1}])x_i x_i^T.$$

Then $(\beta-\beta^{(0)})^T l_n(\beta) \leq S_n(\beta) - (c_2/2)(\beta-\beta^{(0)})^T P_n(\beta-\beta^{(0)})$, where the martingale $S_n(\beta)$ is defined as

$$S_n(\beta) = (\beta - \beta^{(0)})^T A_n + (\beta - \beta^{(0)})^T B_n(\beta - \beta^{(0)}) + \left|\left|\beta - \beta^{(0)}\right|\right| (\beta - \beta^{(0)})^T J_n(\beta - \beta^{(0)}).$$

Following Chang (1999), define the last-time

$$N_\rho = \sup\{n \geq n_0 \mid \text{there is no } \beta \in B_\rho \text{ s.t. } l_n(\beta) = 0\}.$$

The following theorem shows that the η-th moment of N_ρ is finite, for $0 < \rho \leq \rho_0$ and sufficiently small $\eta > 0$. Recall our assumptions $\sup_{i\in\mathbb{N}} E\left[|e_i|^r\right] < \infty$, for some $r > 2$, and $\lambda_{\min}(P_n) \geq L_1(n) \geq cn^\alpha$, for some $c > 0$, $\frac{1}{2} < \alpha \leq 1$ and all $n \geq n_0$.

Theorem 7.1. *$N_\rho < \infty$ a.s., and $E[N_\rho^\eta] < \infty$ for all $0 < \rho \leq \rho_0$ and $0 < \eta < r\alpha - 1$.*

Remark 7.1. Chang (1999) also approaches existence and strong consistency of $\hat{\beta}_n$ via application of Corollary 7.1. To this end, he derives an upper bound $A_n + B_n + J_n - n^\alpha\epsilon^*$ for $(\beta - \beta^{(0)})^T l_n(\beta)$, cf. his equation (21). He proceeds to show that for all $\beta \in \partial B_\rho$ the last time that this upper bound is positive, has finite expectation (cf. his equation (22)). However, to deduce existence of $\hat{\beta}_n \in B_\rho$ from Corollary 7.1, one needs to prove (in Chang's notation)

$$E\left[\sup\{n \geq 1 \mid \exists \beta \in \partial B_\rho : A_n + B_n + J_n - n^\alpha\epsilon^* \geq 0\}\right] < \infty,$$

but Chang proves

$$\forall \beta \in \partial B_\rho : E\left[\sup\{n \geq 1 \mid A_n + B_n + J_n - n^\alpha\epsilon^* \geq 0\}\right] < \infty.$$

(Here the terms A_n, B_n, J_n and ϵ^* depend on β).

The following theorem shows asymptotic existence and strong consistency of $\hat{\beta}_n$, and provides mean square convergence rates.

Theorem 7.2. *Let $0 < \rho \leq \rho_0$. For all $n > N_\rho$ there exists a solution $\hat{\beta}_n \in B_\rho$ to $l_n(\beta) = 0$, and $\lim_{n\to\infty} \hat{\beta}_n = \beta^{(0)}$ a.s. Moreover,*

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)} + \frac{n(d-1)^2}{L_1(n)L_2(n)}\right). \tag{7.8}$$

Remark 7.2. If $d = 1$ then the term $\frac{n(d-1)^2}{L_1(n)L_2(n)}$ in (7.8) vanishes. If $d = 2$, the next to smallest eigenvalue $\lambda_2(P_n)$ of P_n is actually the largest eigenvalue of P_n. If in addition $\inf_{i\in\mathbb{N}} ||x_i|| \geq d_{\min} > 0$ a.s. for some $d_{\min} > 0$, then $\lambda_{\max}(P_n) \geq \frac{1}{2}\text{tr}(P_n) \geq \frac{d_{\min}}{2}n$, and $\frac{n(d-1)^2}{L_1(n)L_2(n)} = O\left(\frac{1}{L_1(n)}\right)$. The bound in Theorem 7.2 then reduces to

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)}\right). \tag{7.9}$$

Remark 7.3. In general, the equation $l_n(\beta) = 0$ may have multiple solutions. Procedures for selecting the "right" root are discussed in Small et al. (2000) and Heyde (1997, Section 13.3). Tzavelas (1998) shows that with probability one there exists not more than one consistent solution.

7.3 Results for canonical link functions

In this section we consider again the statistical model introduced in Section 7.1.1, under all the assumptions listed in Section 7.1.3. In addition, we restrict to canonical link functions, i.e. functions

h that satisfy $\dot{h} = v \circ h$. The quasi-likelihood equations (7.2) then simplify to

$$l_n(\beta) = \sum_{i=1}^{n} x_i(y_i - h(x_i^T\beta)) = 0. \tag{7.10}$$

This simplification enables us to improve the bounds from Theorem 7.2. In particular, the main result of this section is Theorem 7.3, which shows that the term $O\left(\frac{n(d-1)^2}{L_1(n)L_2(n)}\right)$ in (7.8) vanishes, yielding the following upper bound on the mean square convergence rates:

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n>N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)}\right).$$

In the previous section, we invoked a corollary of the Leray-Schauder Theorem to prove existence of $\hat{\beta}_n$ in a proximity of $\beta^{(0)}$. In the case of canonical link function, a similar existence result is derived from the following fact:

Lemma 7.3 (Chen et al., 1999, Lemma A(i))**.** *Let $H : \mathbb{R}^d \to \mathbb{R}^d$ be a continuously differentiable injective mapping, $x_0 \in \mathbb{R}^d$, and $\delta > 0$, $r > 0$. If $\inf_{x:||x-x_0||=\delta} ||H(x) - H(x_0)|| \geq r$ then for all $y \in \{y \in \mathbb{R}^d \mid ||y - H(x_0)|| \leq r\}$ there is an $x \in \{x \in \mathbb{R}^d \mid ||x - x_0|| \leq \delta\}$ such that $H(x) = y$.*

Chen et al. (1999) assume that H is smooth, but an inspection of their proof reveals that H being a continuously differentiable injection is sufficient.

We apply Lemma 7.3 with $H(\beta) = P_n^{-1/2} l_n(\beta)$ and $y = 0$:

Corollary 7.2. *Let $0 < \rho \leq \rho_0$, $n \geq N_\rho$, $\delta > 0$ and $r > 0$. If $\inf_{\beta\in\partial B_\delta} \left|\left|H_n(\beta) - H_n(\beta^{(0)}))\right|\right| \geq r$ and $\left|\left|H_n(\beta^{(0)})\right|\right| \leq r$ then there is a $\beta \in B_\delta$ with $P_n^{-1/2} l_n(\beta) = 0$, and thus $l_n(\beta) = 0$.*

Remark 7.4. The proof of Corollary 7.2 reveals that $l_n(\beta)$ is injective for all $n \geq n_0$, and thus $\hat{\beta}_n$ is uniquely defined for all $n \geq N_\rho$.

The following theorem improves the mean square convergence rates of Theorem 7.2 in case of canonical link functions.

Theorem 7.3. *In case of a canonical link function,*

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n\geq N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)}\right), \quad (0 < \rho \leq \rho_0). \tag{7.11}$$

Remark 7.5. Some choices of h, e.g. h the identity or the logit function, have the property that $\inf_{x\in X,\beta\in\mathbb{R}^d} \dot{h}(x^T\beta) > 0$, i.e. c_2 in equation (7.7) has a positive lower bound independent of ρ_0. Since canonical link functions have $c_1 = 0$ in equation (7.7), we then can choose $\rho_0 = \infty$ in Lemma 7.2, Theorem 7.1 and Theorem 7.3. Then $N_{\rho_0} = n_0$ and $\hat{\beta}_n$ exists a.s. for all $n \geq n_0$. Moreover, we can drop assumption (7.4) and obtain

$$E\left[\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right|^2\right] = O\left(\frac{\log(n)}{L_1(n)}\right), \quad (n \geq n_0). \tag{7.12}$$

for any positive lower bound $L_1(n)$ on $\lambda_{\min}(P_n)$. Naturally, one needs to assume $\log(n) = o(L_1(n))$ in order to conclude from (7.12) that $E\left[||\hat{\beta}_n - \beta^{(0)}||^2\right]$ converges to zero as $n \to \infty$.

7.4 Proofs

Proof of Lemma 7.2

A Taylor expansion of h and g yields

$$y_i - h(x_i^T\beta) = y_i - h(x_i^T\beta^{(0)}) + h(x_i^T\beta^{(0)}) - h(x_i^T\beta) = e_i - \dot{h}(x_i^T\tilde{\beta}_{i,\beta}^{(1)})x_i^T(\beta-\beta^{(0)}), \quad (7.13)$$

$$g(x_i^T\beta) = g(x_i^T\beta^{(0)}) + \dot{g}(x_i^T\beta^{(0)})x_i^T(\beta-\beta^{(0)}) + \frac{1}{2}(\beta-\beta^{(0)})^T\ddot{g}(x_i^T\tilde{\beta}_{i,\beta}^{(2)})x_ix_i^T(\beta-\beta^{(0)}), \quad (7.14)$$

for some $\tilde{\beta}_{i,\beta}^{(1)}$, $\tilde{\beta}_{i,\beta}^{(2)}$ on the line segment between β and $\beta^{(0)}$. As in Chang (1999, page 241), it follows that

$$\begin{aligned}
(\beta-\beta^{(0)})^T l_n(\beta) &= (\beta-\beta^{(0)})^T \sum_{i=1}^n g(x_i^T\beta)x_i(e_i - \dot{h}(x_i^T\tilde{\beta}_{i,\beta}^{(1)})x_i^T(\beta-\beta^{(0)})) \\
&= (\beta-\beta^{(0)})^T \sum_{i=1}^n g(x_i^T\beta^{(0)})x_ie_i \\
&+ (\beta-\beta^{(0)})^T \sum_{i=1}^n \dot{g}(x_i^T\beta^{(0)})x_i^T(\beta-\beta^{(0)})x_ie_i \\
&+ (\beta-\beta^{(0)})^T \sum_{i=1}^n \left[\frac{1}{2}(\beta-\beta^{(0)})^T\ddot{g}(x_i^T\tilde{\beta}_{i,\beta}^{(2)})x_ix_i^T(\beta-\beta^{(0)})\right]x_ie_i \\
&- (\beta-\beta^{(0)})^T \sum_{i=1}^n g(x_i^T\beta)x_i\dot{h}(x_i^T\tilde{\beta}_{i,\beta}^{(1)})x_i^T(\beta-\beta^{(0)}) \\
&= (\beta-\beta^{(0)})^T A_n + (\beta-\beta^{(0)})^T B_n(\beta-\beta^{(0)}) + (I) - (II),
\end{aligned}$$

where we write $(I) = (\beta-\beta^{(0)})^T \sum_{i=1}^n \left[\frac{1}{2}(\beta-\beta^{(0)})^T\ddot{g}(x_i^T\tilde{\beta}_{i,\beta}^{(2)})x_ix_i^T(\beta-\beta^{(0)})\right]x_ie_i$ and $(II) = (\beta-\beta^{(0)})^T \sum_{i=1}^n g(x_i^T\beta)x_i\dot{h}(x_i^T\tilde{\beta}_{i,\beta}^{(1)})x_i^T(\beta-\beta^{(0)})$. Since

$$\begin{aligned}
(I) =& (\beta-\beta^{(0)})^T \sum_{i=1}^n \left[\frac{1}{2}(\beta-\beta^{(0)})^T\ddot{g}(x_i^T\tilde{\beta}_{i,\beta}^{(2)})x_i\right]x_ix_i^T(\beta-\beta^{(0)})e_i \\
\le& (\beta-\beta^{(0)})^T \sum_{i=1}^n \left[\frac{1}{2}\left|\left|\beta-\beta^{(0)}\right|\right| |\ddot{g}(x_i^T\tilde{\beta}_{i,\beta}^{(2)})|\,||x_i||\right]x_ix_i^T(\beta-\beta^{(0)})|e_i| \\
\le& c_1(\beta-\beta^{(0)})^T \sum_{i=1}^n \left|\left|\beta-\beta^{(0)}\right|\right| x_ix_i^T|e_i|(\beta-\beta^{(0)}) \\
\le& c_1(\beta-\beta^{(0)})^T \sum_{i=1}^n \left|\left|\beta-\beta^{(0)}\right|\right| x_ix_i^T(|e_i| - E\left[|e_i| \mid \mathcal{F}_{i-1}\right])(\beta-\beta^{(0)}) \\
&+ c_1(\beta-\beta^{(0)})^T \sum_{i=1}^n \left|\left|\beta-\beta^{(0)}\right|\right| x_ix_i^T E\left[|e_i| \mid \mathcal{F}_{i-1}\right](\beta-\beta^{(0)}) \\
\le& \left|\left|\beta-\beta^{(0)}\right|\right| (\beta-\beta^{(0)})^T J_n(\beta-\beta^{(0)}) \\
&+ c_1c_3\left|\left|\beta-\beta^{(0)}\right|\right| (\beta-\beta^{(0)})^T \sum_{i=1}^n x_ix_i^T(\beta-\beta^{(0)})
\end{aligned}$$

and

$$(II) \geq c_2(\beta - \beta^{(0)})^T \sum_{i=1}^{n} x_i x_i^T (\beta - \beta^{(0)}),$$

by combining all relevant inequalities we obtain

$$\begin{aligned}(\beta - \beta^{(0)})^T l_n(\beta) \leq\ & (\beta - \beta^{(0)})^T A_n + (\beta - \beta^{(0)})^T B_n(\beta - \beta^{(0)}) \\ & + \left|\left|\beta - \beta^{(0)}\right|\right| (\beta - \beta^{(0)})^T J_n(\beta - \beta^{(0)}) - (c_2/2)(\beta - \beta^{(0)})^T \sum_{i=1}^{n} x_i x_i^T (\beta - \beta^{(0)}),\end{aligned}$$

using $(c_1 c_3 \left|\left|\beta - \beta^{(0)}\right|\right| - c_2) \leq (c_1 c_3 \rho - c_2) \leq -c_2/2$.

Proof of Theorem 7.1

Fix $\rho \in (0, \rho_0]$ and $0 < \eta < r\alpha - 1$. Let $S_n(\beta)$ be as in Lemma 7.2. Define the last-time

$$T = \sup\{n \geq n_0 \mid \sup_{\beta \in \partial B_\rho} S_n(\beta) - \rho^2 (c_2/2) L_1(n) > 0\}.$$

By Lemma 7.2, for all $n > T$,

$$\begin{aligned}0 \geq & \sup_{\beta \in \partial B_\rho} S_n(\beta) - \rho^2 (c_2/2) L_1(n) \geq \sup_{\beta \in \partial B_\rho} S_n(\beta) - (c_2/2)(\beta - \beta^{(0)})^T P_n (\beta - \beta^{(0)}) \\ & \geq \sup_{\beta \in \partial B_\rho} (\beta - \beta^{(0)})^T l_n(\beta),\end{aligned}$$

which by Corollary 7.1 implies $n > N_\rho$. Then $N_\rho \leq T$ a.s., and thus $E[N_\rho^\eta] \leq E[T^\eta]$ for all $\eta > 0$. The proof is complete if we show the assertions for T.

If we denote the entries of the vector A_n and the matrices B_n, J_n by $A_n[i]$, $B_n[i,j]$, $J_n[i,j]$, then

$$\begin{aligned}\sup_{\beta \in \partial B_\rho} S_n(\beta) & \leq \rho \left|\left|A_n\right|\right| + \rho^2 \left|\left|B_n\right|\right| + \rho^3 \left|\left|J_n\right|\right| \\ & \leq \rho \sum_{1 \leq i \leq d} |A_n[i]| + \rho^2 \sum_{1 \leq i,j \leq d} |B_n[i,j]| + \rho^3 \sum_{1 \leq i,j \leq d} |J_n[i,j]|,\end{aligned}$$

using the Cauchy-Schwartz inequality and the fact that $||x|| \leq ||x||_1$, $||A|| \leq \sum_{i,j} |A[i,j]|$ for vectors x and matrices A. (This can be derived from the inequality $||A|| \leq \sqrt{||A||_1 \, ||A||_\infty}$). We now define $d + 2d^2$ last-times:

$$\begin{aligned}T_{A[i]} &= \sup\{n \geq n_0 \mid \rho |A_n[i]| - \frac{1}{d + 2d^2} \rho^2 (c_2/2) L_1(n) > 0\}, \quad (1 \leq i \leq d), \\ T_{B[i,j]} &= \sup\{n \geq n_0 \mid \rho^2 |B_n[i,j]| - \frac{1}{d + 2d^2} \rho^2 (c_2/2) L_1(n) > 0\}, \quad (1 \leq i, j \leq d), \\ T_{J[i,j]} &= \sup\{n \geq n_0 \mid \rho^3 |J_n[i,j]| - \frac{1}{d + 2d^2} \rho^2 (c_2/2) L_1(n) > 0\}, \quad (1 \leq i, j \leq d).\end{aligned}$$

By application of Proposition 7.1, Section 7.5, the last-times $T_{A[i]}$ and $T_{B[i,j]}$ are a.s. finite and have finite η-th moment, for all $\eta > 0$ such that $r > \frac{\eta+1}{\alpha} > 2$. Chow and Teicher (2003, page 95, Lemma 3) states that any two nonnegative random variables X_1, X_2 satisfy

$$E\left[(X_1 + X_2)^\eta\right] \leq 2^\eta (E\left[X_1^\eta\right] + E\left[X_2^\eta\right]), \tag{7.15}$$

for all $\eta > 0$. Consequently

$$\sup_{i\in\mathbb{N}} E\left[\left||e_i| - E\left[|e_i| \mid \mathcal{F}_{i-1}\right]\right|^r\right] \leq \sup_{i\in\mathbb{N}} E\left[\left||e_i| + E\left[|e_i| \mid \mathcal{F}_{i-1}\right]\right|^r\right]$$
$$\leq \sup_{i\in\mathbb{N}} 2^r(E\left[|e_i|^r\right] + E\left[(E\left[|e_i| \mid \mathcal{F}_{i-1}\right])^r\right]) < \infty,$$

and Proposition 7.1 implies that the last-times $T_{J[i,j]}$ are also a.s. finite and have finite η-th moment, for all $\eta > 0$ such that $r > \frac{\eta+1}{\alpha} > 2$. Now set $\mathcal{T} = \sum_{1\leq i\leq d} T_{A[i]} + \sum_{1\leq i,j\leq d} T_{B[i,j]} + \sum_{1\leq i,j\leq d} T_{J[i,j]}$. If $n > \mathcal{T}$, then $\sup_{\beta\in\partial B_\rho} S_n(\beta) - \rho^2(c_2/2)L_1(n) \leq 0$, and thus $T \leq \mathcal{T}$ a.s. and $E\left[T^\eta\right] \leq E\left[\mathcal{T}^\eta\right]$. $\mathcal{T}$ is finite a.s., since all terms $T_{A[i]}$, $T_{B[i,j]}$ and $T_{J[i,j]}$ are finite a.s. Moreover, by repeated application of (7.15), for all $\eta > 0$ there is a constant C_η such that

$$E[\mathcal{T}^\eta] \leq C_\eta \left[\sum_{1\leq i\leq d} E\left[T_{A[i]}\right] + \sum_{1\leq i,j\leq d} E\left[T^\eta_{B[i,j]}\right] + \sum_{1\leq i,j\leq d} E\left[T^\eta_{J[i,j]}\right]\right].$$

It follows that $E\left[\mathcal{T}^\eta\right] < \infty$ for all $\eta > 0$ such that $r > \frac{\eta+1}{\alpha} > 2$. In particular, this implies $N_\rho < \infty$ a.s., and $E\left[N^\eta_\rho\right] < \infty$.

Proof of Theorem 7.2

The asymptotic existence and strong consistency of $\hat{\beta}_n$ follow directly from Theorem 7.1 which shows $N_\rho < \infty$ a.s. for all $0 < \rho \leq \rho_0$.

To prove the mean square convergence rates, let $0 < \rho \leq \rho_0$.

By contraposition of Corollary 7.1, if there is no solution $\beta \in B_\rho$ to $l_n(\beta) = 0$, then there exists a $\beta' \in \partial B_\rho$ such that $(\beta' - \beta^{(0)})^T l_n(\beta') > 0$, and thus $S_n(\beta') - (c_2/2)(\beta' - \beta^{(0)})^T P_n(\beta' - \beta^{(0)}) > 0$ by Lemma 7.2. In particular,

$$(\beta'-\beta^{(0)})^T(c_2/2)P_n(\beta'-\beta^{(0)}) - (\beta'-\beta^{(0)})^T\left[A_n + B_n(\beta'-\beta^{(0)}) + \left|\left|\beta'-\beta^{(0)}\right|\right| J_n(\beta'-\beta^{(0)})\right] \leq 0,$$

and, writing

$$(I) = \left|\left|(c_2/2)^{-1}P_n^{-1}\left[A_n + B_n(\beta'-\beta^{(0)}) + \rho J_n(\beta'-\beta^{(0)})\right]\right|\right|^2$$

and

$$(II) = \frac{(d-1)^2\left|\left|A_n + B_n(\beta'-\beta^{(0)}) + \rho J_n(\beta'-\beta^{(0)})\right|\right|^2}{L_1(n)L_2(n)(c_2/2)^2},$$

Lemma 7.7, Section 7.5, implies

$$\rho^2 = \left|\left|\beta'-\beta^{(0)}\right|\right|^2 \leq (I) + (II). \tag{7.16}$$

We now proceed to show

$$(I) + (II) < U_n, \tag{7.17}$$

for some U_n, independent of β' and ρ, that satisfies

$$E\left[U_n\right] = O\left(\frac{\log(n)}{L_1(n)} + \frac{n(d-1)^2}{L_1(n)L_2(n)}\right).$$

Thus, if there is no solution $\beta \in B_\rho$ of $l_n(\beta) = 0$, then $\rho^2 < U_n$. This implies that there is always a solution $\beta \in B_{U_n^{1/2}}$ to $l_n(\beta) = 0$, and thus $||\hat{\beta}_n - \beta^{(0)}||^2 \mathbf{1}_{n>N_\rho} \leq U_n$ a.s., and $E\left[||\hat{\beta}_n - \beta^{(0)}||^2 \mathbf{1}_{n>N_\rho}\right] \leq E[U_n]$.

To prove (7.17), we decompose (I) and (II) using the following fact: if M, N are $d \times d$ matrices, and $N(j)$ denotes the j-th column of N, then

$$||MN|| = \max_{||y||=1} ||MNy|| = \max_{||y||=1} \left|\left| M \sum_{j=1}^{d} y[j] N(j) \right|\right| \leq \max_{||y||=1} \sum_{j=1}^{d} ||My[j]N(j)|| \leq \sum_{j=1}^{d} ||MN(j)||.$$

As a result we get

$$\begin{aligned} \left|\left| P_n^{-1} B_n(\beta' - \beta^{(0)}) \right|\right| &\leq \left|\left| P_n^{-1} \sum_{i=1}^{n} \dot{g}(x_i^T \beta^{(0)}) x_i e_i x_i^T \right|\right| \left|\left| \beta' - \beta^{(0)} \right|\right| \\ &\leq \rho \sum_{j=1}^{d} \left|\left| P_n^{-1} \sum_{i=1}^{n} \dot{g}(x_i^T \beta^{(0)}) x_i e_i x_i[j] \right|\right| \end{aligned}$$

and

$$\begin{aligned} \left|\left| P_n^{-1} J_n(\beta' - \beta^{(0)}) \right|\right| &\leq \left|\left| P_n^{-1} \sum_{i=1}^{n} c_1 x_i (|e_i| - E[|e_i| \mid \mathcal{F}_{i-1}]) x_i^T \right|\right| \left|\left| \beta' - \beta^{(0)} \right|\right| \\ &\leq \rho \sum_{j=1}^{d} \left|\left| P_n^{-1} \sum_{i=1}^{n} c_1 x_i (|e_i| - E[|e_i| \mid \mathcal{F}_{i-1}]) x_i[j] \right|\right|. \end{aligned}$$

In a similar vein we can derive

$$\left|\left| B_n(\beta' - \beta^{(0)}) \right|\right| \leq \rho \sum_{j=1}^{d} \left|\left| \sum_{i=1}^{n} \dot{g}(x_i^T \beta^{(0)}) x_i e_i x_i[j] \right|\right|$$

and

$$\left|\left| J_n(\beta' - \beta^{(0)}) \right|\right| \leq \rho \sum_{j=1}^{d} \left|\left| \sum_{i=1}^{n} c_1 x_i (|e_i| - E[|e_i| \mid \mathcal{F}_{i-1}]) x_i[j] \right|\right|.$$

It follows that

$$\begin{aligned} (I) &\leq 2(c_2/2)^{-2} \left(\left|\left| P_n^{-1} A_n \right|\right|^2 + \left|\left| P_n^{-1} B_n(\beta' - \beta^{(0)}) \right|\right|^2 + \rho_0^2 \left|\left| P_n^{-1} J_n(\beta' - \beta^{(0)}) \right|\right|^2 \right) \\ &\leq U_n(1) + U_n(2) + U_n(3), \end{aligned}$$

where we write

$$\begin{aligned} U_n(1) &= 2(c_2/2)^{-2} \left|\left| P_n^{-1} A_n \right|\right|^2, \\ U_n(2) &= 2(c_2/2)^{-2} \rho_0^2 2 \left(\sum_{j=1}^{d} \left|\left| P_n^{-1} \sum_{i=1}^{n} \dot{g}(x_i^T \beta^{(0)}) x_i e_i x_i[j] \right|\right|^2 \right), \\ U_n(3) &= 2(c_2/2)^{-2} \rho_0^4 2 \left(\sum_{j=1}^{d} \left|\left| P_n^{-1} \sum_{i=1}^{n} c_1 x_i (|e_i| - E[|e_i| \mid \mathcal{F}_{i-1}]) x_i[j] \right|\right|^2 \right), \end{aligned}$$

and

$$(II) \leq U_n(4) + U_n(5) + U_n(6),$$

where we write

$$U_n(4) = \frac{2(d-1)^2 \left|\left|A_n\right|\right|^2}{L_1(n)L_2(n)(c_2/2)^2},$$

$$U_n(5) = \frac{2(d-1)^2}{L_1(n)L_2(n)(c_2/2)^2} \left(\rho_0 \sum_{j=1}^{d} \left|\left| \sum_{i=1}^{n} \dot{g}(x_i^T\beta^{(0)}) x_i e_i x_i[j] \right|\right| \right)^2,$$

$$U_n(6) = \frac{2(d-1)^2}{L_1(n)L_2(n)(c_2/2)^2} \rho_0^2 \left(\rho_0 \sum_{j=1}^{d} \left|\left| \sum_{i=1}^{n} c_1 x_i (|e_i| - E\left[|e_i| \mid \mathcal{F}_{i-1}\right]) x_i[j] \right|\right| \right)^2.$$

The desired upper bound U_n for $(I) + (II)$ equals $U_n = \sum_{j=1}^{6} U_n(j)$. For $U_n(1)$, $U_n(2)$, $U_n(3)$, apply Proposition 7.2 in Section 7.5 on the martingale difference sequences $(g(x_i^T\beta^{(0)})e_i)_{i\in\mathbb{N}}$, $(\dot{g}(x_i^T\beta^{(0)})x_i[j]e_i)_{i\in\mathbb{N}}$ and $(c_1(|e_i| - E\left[|e_i| \mid \mathcal{F}_{i-1}\right] x_i[j])_{i\in\mathbb{N}}$, respectively. This implies the existence of a constant $K_1 > 0$ such that $E[U_n(1) + U_n(2) + U_n(3)] \leq \frac{K_1 \log(n)}{L_1(n)}$. For $U_n(4)$, $U_n(5)$, $U_n(6)$, the assumption

$$\sup_{i\in\mathbb{N}} E\left[e_i^2 \mid \mathcal{F}_{i-1}\right] \leq \sigma^2 < \infty \text{ a.s.}$$

implies the existence of a constant $K_2 > 0$ such that $E\left[U_n(4) + U_n(5) + U_n(6)\right] \leq \frac{K_2 n(d-1)^2}{L_1(n)L_2(n)}$.

Proof of Corollary 7.2

It is sufficient to show that $H(\beta)$ is injective. Suppose $P_n^{-1/2} l_n(\beta) = P_n^{-1/2} l_n(\beta')$ for some β, β'. Since $n \geq n_0$ this implies $l_n(\beta) = l_n(\beta')$. By a first order Taylor expansion, there are $\tilde{\beta}_i$, $1 \leq i \leq n$, on the line segment between β and β' such that $l_n(\beta) - l_n(\beta') = \sum_{i=1}^{n} x_i x_i^T \dot{h}(x_i^T\tilde{\beta}_i)(\beta - \beta') = 0$. Since $\inf_{x\in X, \beta\in B_\rho} \dot{h}(x^T\beta) > 0$, Lemma 7.8 in Section 7.5 implies that the matrix $\sum_{i=1}^{n} x_i x_i^T \dot{h}(x_i^T\tilde{\beta}_i)$ is invertible, and thus $\beta = \beta'$.

Proof of Theorem 7.3

Let $0 < \rho \leq \rho_0$ and $n \geq N_\rho$. A Taylor expansion of $l_n(\beta)$ yields

$$l_n(\beta) - l_n(\beta^{(0)}) = \sum_{i=1}^{n} x_i (h(x_i^T\beta^{(0)}) - h(x_i^T\beta)) = \sum_{i=1}^{n} x_i x_i^T \dot{h}(x_i^T\beta_{in})(\beta^{(0)} - \beta),$$

for some β_{in}, $1 \leq i \leq n$, on the line segment between $\beta^{(0)}$ and β.

Write $T_n(\beta) = \sum_{i=1}^{n} x_i x_i^T \dot{h}(x_i^T\beta_{in})$, and choose $k_2 > \left(\inf_{\beta\in B_\rho, x\in X} \dot{h}(x^T\beta)\right)^{-1}$. Then for all $\beta \in B_\rho$,

$$\begin{aligned} \lambda_{\min}\left(k_2 T_n(\beta) - P_n\right) &= \lambda_{\min}\left(\sum_{i=1}^{n} x_i x_i^T (k_2 \dot{h}(x_i^T\beta_{in}) - 1) \right) \\ &\geq \left(\inf_{\beta\in B_{\rho_0}, x\in X} (k_2 \dot{h}(x^T\beta) - 1) \right) \lambda_{\min}(P_n), \end{aligned}$$

by Lemma 7.8. This implies

$$y^T k_2 T_n(\beta) y \geq y^T P_n y \quad \text{and } y^T k_2^{-1} T_n(\beta)^{-1} y \leq y^T P_n^{-1} y \quad \text{for all } y \in \mathbb{R}^d,$$

cf. Bhatia (2007, page 11, Exercise 1.2.12).

Define $H_n(\beta) = P_n^{-1/2} l_n(\beta)$, $r_n = \left|\left|H_n(\beta^{(0)})\right|\right|$, and $\delta_n = \frac{r_n}{k_2^{-1}\sqrt{L_1(n)}}$. If $\delta_n > \rho$ then it follows immediately that $||\hat{\beta}_n - \beta^{(0)}|| \leq \rho < \frac{\left|\left|H_n(\beta^{(0)})\right|\right|}{k_2^{-1}\sqrt{L_1(n)}}$. Suppose $\delta_n \leq \rho$. Then for all $\beta \in \partial B_{\delta_n}$,

$$\begin{aligned}
\left|\left|H_n(\beta) - H_n(\beta^{(0)})\right|\right|^2 &= \left|\left|P_n^{-1/2}(l_n(\beta) - l_n(\beta^{(0)}))\right|\right|^2 \\
&= (\beta^{(0)} - \beta)^T T_n(\beta) P_n^{-1} T_n(\beta)(\beta^{(0)} - \beta) \\
&\geq (\beta^{(0)} - \beta)^T T_n(\beta) k_2^{-1} T_n(\beta)^{-1} T_n(\beta)(\beta^{(0)} - \beta) \\
&\geq (\beta^{(0)} - \beta)^T P_n k_2^{-2} (\beta^{(0)} - \beta) \\
&\geq k_2^{-2} \left|\left|\beta^{(0)} - \beta\right|\right|^2 \lambda_{\min}(P_n) \\
&\geq k_2^{-2} \delta_n^2 L_1(n),
\end{aligned}$$

and thus $\inf_{\beta \in \partial B_{\delta_n}} \left|\left|H_n(\beta) - H_n(\beta^{(0)})\right|\right| \geq k_2^{-1}\sqrt{L_1(n)}\delta_n = r_n$ and $\left|\left|H(\beta^{(0)})\right|\right| \leq r_n$. By Corollary 7.2 we conclude that $\left|\left|\hat{\beta}_n - \beta^{(0)}\right|\right| \leq \frac{\left|\left|H_n(\beta^{(0)})\right|\right|}{k_2^{-1}\sqrt{L_1(n)}}$ a.s.

Now $E\left[\left|\left|H_n(\beta^{(0)})\right|\right|^2\right] = E\left[\left(\sum_{i=1}^n x_i e_i\right)^T P_n^{-1} \left(\sum_{i=1}^n x_i e_i\right)\right] = E[Q_n]$, where Q_n is as in the proof of Proposition 7.2. There we show $E[Q_n] \leq K \log(n)$, for some $K > 0$ and all $n \geq n_0$, and thus we have $E\left[\left|\left|\beta - \beta^{(0)}\right|\right|^2 \mathbf{1}_{n \geq N_\rho}\right] = O\left(\frac{\log(n)}{L_1(n)}\right)$.

7.5 Appendix: auxiliary results

In this appendix, we prove and collect several probabilistic results which are used in the preceding sections. Proposition 7.1 is fundamental to Theorem 7.1, where we provide sufficient conditions such that the η-th moment of the last-time N_ρ is finite, for $\eta > 0$. The proof of the proposition makes use of two auxiliary lemma's. Lemma 7.4 is a maximum inequality for tail probabilities of martingales; for sums of i.i.d. random variables this statement can be found e.g. in Loève (1977a, Section 18.1C, page 260), and a martingale version was already hinted at in Loève (1977b, Section 32.1, page 51). Lemma 7.5 contains a so-called Baum-Katz-Nagaev type theorem proven by Stoica (2007). There exists a long tradition of these type of results for sums of independent random variables, see e.g. Spataru (2009) and the references therein. Stoica (2007) makes an extension to martingales. In Proposition 7.2 we provide L^2 bounds for least-squares linear regression estimates, similar to the a.s. bounds derived by Lai and Wei (1982). The bounds for the quality of maximum quasi-likelihood estimates, Theorem 7.2 in Section 7.2 and Theorem 7.3 in Section 7.3, are proven by relating them to these bounds from Proposition 7.2. Lemma 7.6 is an auxiliary result used in the proof of Proposition 7.2. Finally, Lemma 7.7 is used in the proof of Theorem 7.2, and Lemma 7.8 in the proof of Theorem 7.3.

Lemma 7.4. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. Write $S_n = \sum_{i=1}^n X_i$, and suppose $\sup_{i\in\mathbb{N}} E[X_i^2 \mid \mathcal{F}_{i-1}] \leq \sigma^2 < \infty$ a.s., for some $\sigma > 0$. Then for all $n \in \mathbb{N}$ and*

$\epsilon > 0$,

$$P\left(\max_{1\le k\le n}|S_k| \ge \epsilon\right) \le 2P\left(|S_n| \ge \epsilon - \sqrt{2\sigma^2 n}\right). \tag{7.18}$$

Proof. We use similar techniques as de la Peña et al. (2009, Theorem 2.21, p.16), where (7.18) is proven for independent random variables $(X_i)_{i\in\mathbb{N}}$. Define the events $A_1 = \{S_1 \ge \epsilon\}$ and $A_k = \{S_k \ge \epsilon, S_1 < \epsilon, \ldots, S_{k-1} < \epsilon\}$, $2 \le k \le n$. Then $A_k (1 \le k \le n)$ are mutually disjoint, and $\{\max_{1\le k\le n} S_k \ge \epsilon\} = \bigcup_{k=1}^n A_k$.

$$\begin{aligned}
P\left(\max_{1\le k\le n} S_k \ge \epsilon\right) &\le P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + P\left(\max_{1\le k\le n} S_k \ge \epsilon, S_n < \epsilon - \sqrt{2\sigma^2 n}\right) \\
&\le P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + \sum_{k=1}^n P\left(A_k, S_n < \epsilon - \sqrt{2\sigma^2 n}\right) \\
&\le P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + \sum_{k=1}^n P\left(A_k, S_n - S_k < -\sqrt{2\sigma^2 n}\right) \\
&\overset{(1)}{=} P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + \sum_{k=1}^n E\left[\mathbf{1}_{A_k} E\left[\mathbf{1}_{S_n - S_k < -\sqrt{2\sigma^2 n}} \mid \mathcal{F}_k\right]\right] \\
&\overset{(2)}{\le} P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + \sum_{k=1}^n \frac{1}{2} P\left(A_k\right) \\
&= P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + P\left(\max_{1\le k\le n} S_k \ge \epsilon\right),
\end{aligned}$$

where (1) uses $A_k \in \mathcal{F}_k$, and (2) uses $E[\mathbf{1}_{S_n - S_k < -\sqrt{2\sigma^2 n}} \mid \mathcal{F}_k] = P(S_k - S_m > \sqrt{2\sigma^2 n} \mid \mathcal{F}_k) \le E[(S_n - S_k)^2 \mid \mathcal{F}_k]/(2\sigma^2 n) \le 1/2$ a.s. This proves $P(\max_{1\le k\le n} S_k \ge \epsilon) \le 2P(S_n \ge \epsilon - \sqrt{2\sigma^2 n})$. Replacing S_k by $-S_k$ gives $P(\max_{1\le k\le n} -S_k \ge \epsilon) \le 2P(-S_n \ge \epsilon - \sqrt{2\sigma^2 n})$. If $\epsilon - \sqrt{2\sigma^2 n} \le 0$ then (7.18) is trivial; if $\epsilon > \sqrt{2\sigma^2 n}$ then

$$\begin{aligned}
P\left(\max_{1\le k\le n}|S_k| \ge \epsilon\right) &\le P\left(\max_{1\le k\le n} S_k \ge \epsilon\right) + P\left(\max_{1\le k\le n} -S_k \ge \epsilon\right) \\
&\le 2P\left(S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) + 2P\left(-S_n \ge \epsilon - \sqrt{2\sigma^2 n}\right) \\
&= 2P\left(|S_n| \ge \epsilon - \sqrt{2\sigma^2 n}\right).
\end{aligned}$$

□

Lemma 7.5 (Stoica, 2007). *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. Write $S_n = \sum_{i=1}^n X_i$ and suppose $\sup_{i\in\mathbb{N}} E[X_i^2 \mid \mathcal{F}_{i-1}] \le \sigma^2 < \infty$ a.s. for some $\sigma > 0$. Let $c > 0$, $\frac{1}{2} < \alpha \le 1$, $\eta > 2\alpha - 1$, $r > \frac{\eta+1}{\alpha}$. If $\sup_{i\in\mathbb{N}} E\left[|X_i|^r\right] < \infty$, then*

$$\sum_{k\ge 1} k^{\eta-1} P\left(|S_k| \ge ck^\alpha\right) < \infty.$$

Proposition 7.1. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. Write $S_n = \sum_{i=1}^n X_i$ and suppose $\sup_{i\in\mathbb{N}} E[X_i^2 \mid \mathcal{F}_{i-1}] \le \sigma^2 < \infty$ a.s. for some $\sigma > 0$. Let $c > 0$, $\frac{1}{2} < \alpha \le 1$, $\eta > 2\alpha - 1$, $r > \frac{\eta+1}{\alpha}$, and define the random variable $T = \sup\{n \in \mathbb{N} \mid |S_n| \ge cn^\alpha\}$, where T takes values in $\mathbb{N} \cup \{\infty\}$. If $\sup_{i\in\mathbb{N}} E\left[|X_i|^r\right] < \infty$, then*

$$T < \infty \text{ a.s.}, \quad \text{and } E\left[T^\eta\right] < \infty.$$

Proof. There exists an $n' \in \mathbb{N}$ such that for all $n > n'$, $c(n/2)^\alpha - \sqrt{2\sigma^2 n} \geq c(n/2)^\alpha/2$. For all $n > n'$,

$$\begin{aligned}
P(T > n) &= P(\exists k > n : |S_k| \geq ck^\alpha) \\
&\leq \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(\exists 2^{j-1} \leq k < 2^j : |S_k| \geq ck^\alpha\right) \\
&\leq \sum_{j \geq \lfloor \log_2(n) \rfloor} P\Big(\sup_{1 \leq k \leq 2^j} |S_k| \geq c(2^{j-1})^\alpha\Big) \\
&\overset{(1)}{\leq} 2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c(2^{j-1})^\alpha - \sqrt{2\sigma^2 2^j}\right) \\
&\overset{(2)}{\leq} 2 \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c(2^{j-1})^\alpha/2\right).
\end{aligned}$$

where (1) follows from Lemma 7.4 and (2) from the definition of n'.

For $t \in \mathbb{R}_+$ write $S_t = S_{\lfloor t \rfloor}$. Then

$$\sum_{j \geq \log_2(n)} P\left(|S_{2^j}| \geq c(2^{j-1})^\alpha/2\right) = \int_{j \geq \log_2(n)} P\left(|S_{2^j}| \geq c(2^{j-1})^\alpha/2\right) dj \tag{7.19}$$

$$= \int_{k \geq n} P\left(|S_k| \geq c(k/2)^\alpha/2\right) \frac{1}{k \log(2)} dk = \sum_{k \geq n} P\left(|S_k| \geq c(k/2)^\alpha/2\right) \frac{1}{k \log(2)}, \tag{7.20}$$

using a variable substitution $k = 2^j$.

By Chebyshev's inequality,

$$P(T > n) \leq 2 \sum_{k \geq n} P\left(|S_k| \geq c(k/2)^\alpha/2\right) \frac{1}{k \log(2)} \leq 2 \sum_{k \geq n} \sigma^2 k (c(k/2)^\alpha/2)^{-2} \frac{1}{k \log(2)},$$

which implies $P(T = \infty) \leq \liminf_{n \to \infty} P(T > n) = 0$. This proves $T < \infty$ a.s.

Since

$$\begin{aligned}
E[T^\eta] &\leq \eta \left[1 + \sum_{n \geq 1} n^{\eta - 1} P(T > n)\right] \\
&\leq \eta \left[1 + n' \cdot (n')^{\eta - 1} + \sum_{n > n'} n^{\eta - 1} P(T > n)\right] \\
&\leq M \sum_{n > n'} n^{\eta - 1} \sum_{j \geq \lfloor \log_2(n) \rfloor} P\left(|S_{2^j}| \geq c(2^{j-1})^\alpha/2\right),
\end{aligned}$$

for some constant $M > 0$, it follows by (7.19), (7.20) that $E[T^\eta] < \infty$ if

$$\sum_{n \geq 1} n^{\eta - 1} \sum_{k \geq n} P\left(|S_k| \geq c(k/2)^\alpha/2\right) k^{-1} < \infty.$$

By interchanging the sums, it suffices to show

$$\sum_{k \geq 1} k^{\eta - 1} P\left(|S_k| \geq 2^{-1-\alpha} c k^\alpha\right) < \infty.$$

This last statement follows from Lemma 7.5. □

Let $(e_i)_{i\in\mathbb{N}}$ be a martingale difference sequence w.r.t. a filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$, such that $\sup_{i\in\mathbb{N}} E[e_i^2 \mid \mathcal{F}_{i-1}] = \sigma^2 < \infty$ a.s., for some $\sigma > 0$. Let $(x_i)_{i\in\mathbb{N}}$ be a sequence of vectors in $\mathbb{R}^d$. Assume that $(x_i)_{i\in\mathbb{N}}$ are predictable w.r.t. the filtration (i.e. $x_i \in \mathcal{F}_{i-1}$ for all $i \in \mathbb{N}$), and $\sup_{i\in\mathbb{N}} ||x||_i \leq M < \infty$ for some (non-random) $M > 0$. Write $P_n = \sum_{i=1}^n x_i x_i^T$. Let $L : \mathbb{N} \to \mathbb{R}_+$ be a (non-random) function and $n_0 \geq 2$ a (non-random) integer such that $\lambda_{\min}(P_n) \geq L(n)$ for all $n \geq n_0$, and $\lim_{n\to\infty} L(n) = \infty$.

Proposition 7.2. *There is a constant $K > 0$ such that for all $n \geq n_0$,*

$$E\left|\left|\left(\sum_{i=1}^n x_i x_i^T\right)^{-1} \sum_{i=1}^n x_i e_i\right|\right|^2 \leq K\frac{\log(n)}{L(n)}.$$

The proof of Proposition 7.2 uses the following result:

Lemma 7.6. *Let $(y_n)_{n\in\mathbb{N}}$ be a nondecreasing sequence with $y_1 \geq e$. Write $R_n = \frac{1}{\log(y_n)} \sum_{i=1}^n \frac{y_i - y_{i-1}}{y_i}$, where we put $y_0 = 0$. Then $R_n \leq 2$ for all $n \in \mathbb{N}$.*

Proof. Induction on n. $R_1 = \frac{1}{\log(y_1)} \leq 1 \leq 2$. Let $n \geq 2$ and define $g(y) = \frac{1}{\log(y)} \frac{y - y_{n-1}}{y} + \frac{\log(y_{n-1})}{\log(y)} R_{n-1}$. If $R_{n-1} \leq 1$, then $R_n = g(y_n) \leq \frac{1}{\log(y_n)} + 1 \leq 2$. Now suppose $R_{n-1} > 1$. Since $z \mapsto (1 + \log(z))/z$ is decreasing in z on $z \geq 1$, and since $y_{n-1} \geq 1$, we have $(1 + \log(y))/y \leq (1 + \log(y_{n-1}))/y_{n-1}$ for all $y \geq y_{n-1}$. Together with $R_{n-1} > 1$ this implies

$$\frac{\partial g(y)}{\partial y} = \frac{1}{y(\log(y))^2}\left[-1 + \frac{y_{n-1}}{y}(1 + \log(y)) - \log(y_{n-1})R_{n-1}\right] < 0, \quad \text{for all } y \geq y_{n-1}.$$

This proves $R_n = g(y_n) \leq \max_{y\geq y_{n-1}} g(y) = g(y_{n-1}) = R_{n-1} \leq 2$. □

Proof of Proposition 7.2. Write $q_n = \sum_{i=1}^n x_i e_i$ and $Q_n = q_n P_n^{-1} q_n$. For $n \geq n_0$, P_n is invertible, and

$$\left|\left|P_n^{-1} q_n\right|\right|^2 \leq \left|\left|P_n^{-1/2}\right|\right|^2 \cdot \left|\left|P_n^{-1/2} q_n\right|\right|^2 \leq \lambda_{\min}(P_n)^{-1} q_n P_n^{-1} q_n \leq L(n)^{-1} Q_n \text{ a.s.},$$

where we used $\left|\left|P_n^{-1/2}\right|\right| = \lambda_{\max}(P_n^{-1/2}) = \lambda_{\min}(P_n)^{-1/2}$. We show $E[Q_n] \leq K \log(n)$, for a constant K to be defined further below, and all $n \geq n_0$.

Write $V_n = P_n^{-1}$. Since $P_n = P_{n-1} + x_n x_n^T$, it follows from the Sherman-Morrison formula (Bartlett, 1951) that $V_n = V_{n-1} - \frac{V_{n-1} x_n x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n}$, and thus

$$x_n^T V_n = x_n^T V_{n-1} - \frac{(x_n^T V_{n-1} x_n) x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n} = x_n^T V_{n-1}/(1 + x_n^T V_{n-1} x_n).$$

As in Lai and Wei (1982), Q_n satisfies

$$\begin{aligned}
Q_n &= \left(\sum_{i=1}^{n} x_i^T e_i\right) V_n \left(\sum_{i=1}^{n} x_i e_i\right) \\
&= \left(\sum_{i=1}^{n-1} x_i^T e_i\right) V_n \left(\sum_{i=1}^{n-1} x_i e_i\right) + x_n^T V_n x_n e_n^2 + 2 x_n^T V_n \left(\sum_{i=1}^{n-1} x_i e_i\right) e_n \\
&= Q_{n-1} + \left(\sum_{i=1}^{n-1} x_i^T e_i\right) \left(-\frac{V_{n-1} x_n x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n}\right) \left(\sum_{i=1}^{n-1} x_i e_i\right) \\
&+ x_n^T V_n x_n e_n^2 + 2 \frac{x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n} \left(\sum_{i=1}^{n-1} x_i e_i\right) e_n \\
&= Q_{n-1} - \frac{(x_n^T V_{n-1} \sum_{i=1}^{n-1} x_i e_i)^2}{1 + x_n^T V_{n-1} x_n} + x_n^T V_n x_n e_n^2 + 2 \frac{x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n} \left(\sum_{i=1}^{n-1} x_i e_i\right) e_n.
\end{aligned}$$

Observe that

$$E\left[\frac{x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n} \left(\sum_{i=1}^{n-1} x_i e_i\right) e_n\right] = E\left[\frac{x_n^T V_{n-1}}{1 + x_n^T V_{n-1} x_n} \left(\sum_{i=1}^{n-1} x_i e_i\right) E\left[e_n \mid \mathcal{F}_{n-1}\right]\right] = 0$$

and

$$E\left[x_n^T V_n x_n e_n^2\right] = E\left[x_n^T V_n x_n E\left[e_n^2 \mid \mathcal{F}_{n-1}\right]\right] \leq E\left[x_n^T V_n x_n\right] \sigma^2.$$

By telescoping the sum we obtain

$$E[Q_n] \leq E[Q_{\min\{n,n_1\}}] + \sigma^2 \sum_{i=n_1+1}^{n} E[x_i^T V_i x_i],$$

where we define $n_1 \in \mathbb{N}$ to be the smallest $n \geq n_0$ such that $L(n) > e^{1/d}$ for all $n \geq n_1$. We have

$$\begin{aligned}
\det(P_{n-1}) &= \det(P_n - x_n x_n^T) \\
&= \det(P_n) \det(I - P_n^{-1} x_n x_n^T) \\
&= \det(P_n)(1 - x_n^T V_n x_n), \quad (n \geq n_1).
\end{aligned} \tag{7.21}$$

Here the last equality follows from Sylvester's determinant theorem $\det(I + AB) = \det(I + BA)$, for matrices A, B of appropriate size. We thus have $x_n^T V_n x_n = \frac{\det(P_n) - \det(P_{n-1})}{\det(P_n)}$. Define the sequence $(y_n)_{n\in\mathbb{N}}$ by $y_n = \det(P_{n+n_1})$. Then $(y_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence with $y_1 \geq \det(P_{n_1+1}) \geq \lambda_{\min}(P_{n_1+1})^d \geq e$. Lemma 7.6 implies

$$\sum_{i=n_1+1}^{n} x_i^T V_i x_i = \sum_{i=n_1+1}^{n} \frac{y_{i-n_1} - y_{i-1-n_1}}{y_{i-n_1}} = \sum_{i=1}^{n-n_1} \frac{y_i - y_{i-1}}{y_i} \leq 2\log(y_{n-n_1}) = 2\log(\det(P_n)), \text{ a.s.}$$

Now $\log(\det(P_n)) \leq d\log(\lambda_{\max}(P_n)) \leq d\log(\mathrm{tr}(P_n)) \leq d\log(n \sup_{i\in\mathbb{N}} ||x_i||^2) \leq d\log(nM^2)$. Furthermore, for all $n_0 \leq n \leq n_1$ we have

$$\begin{aligned}
E[Q_n] &\leq E\left[||q_n||^2 \lambda_{\max}(P_n^{-1})\right] \leq E\left[\left|\left|\sum_{i=1}^{n} x_i \epsilon_i\right|\right|^2 L(n_0)^{-1}\right] \leq L(n_0)^{-1} E\left[2\sum_{i=1}^{n} \epsilon_i^2 \sup_{i\in\mathbb{N}} ||x_i||^2\right] \\
&\leq 2L(n_0)^{-1} M^2 n_1 \sigma^2,
\end{aligned}$$

and thus for all $n \geq n_0$,

$$\begin{aligned} E\left[Q_n\right] &\leq E\left[Q_{\min\{n,n_1\}}\right] + \sigma^2 \sum_{i=n_1+1}^{n} E\left[x_i^T V_i x_i\right] \\ &\leq 2L(n_0)^{-1} M^2 n_1 \sigma^2 + d\log(n) + d\log(M^2) \\ &\leq K\log(n), \end{aligned}$$

where $K = d + [2L(n_0)^{-1}M^2 n_1\sigma^2 + d\log(M^2)]/\log(n_0)$. □

Lemma 7.7. *Let A be a positive definite $d \times d$ matrix, and b, $x \in \mathbb{R}^d$. If $x^T Ax + x^T b \leq 0$ then $||x||^2 \leq \left|\left|A^{-1}b\right|\right|^2 + (d-1)^2 \frac{||b||^2}{\lambda_1 \lambda_2}$, where $0 < \lambda_1 \leq \lambda_2$ are the two smallest eigenvalues of A.*

Proof. Let $0 < \lambda_1 \leq \ldots \leq \lambda_d$ be the eigenvalues of A, and $v_1, \ldots, v_d$ the corresponding eigenvectors. We can assume that these form an orthonormal basis, such that each $x \in \mathbb{R}^d$ can be written as $\sum_{i=1}^d \alpha_i v_i$, for coordinates $(\alpha_1, \ldots, \alpha_d)$, and $b = \sum_{i=1}^d \beta_i v_i$ for some $(\beta_1, \ldots, \beta_d)$. Write

$$S = \left\{ (\alpha_1, \ldots, \alpha_d) \mid \sum_{i=1}^{d} \alpha_i(\lambda_i \alpha_i + \beta_i) \leq 0 \right\}.$$

The orthonormality of $(v_i)_{1 \leq i \leq d}$ implies that S equals $\left\{x \in \mathbb{R}^d \mid x^T Ax + x^T b \leq 0\right\}$.

Let $\alpha = (\alpha_1, \ldots, \alpha_d) \in S$ and write $R = \{i \mid \alpha_i(\lambda_i\alpha_i + \beta_i) \leq 0, 1 \leq i \leq d\}$. For all $i \in R$, standard properties of quadratic equations imply $\alpha_i^2 \leq \lambda_i^{-2}\beta_i^2$ and $\alpha_i(\lambda_i\alpha_i + \beta_i) \geq \frac{-\beta_i^2}{4\lambda_i}$. For all $i \in S \backslash R$,

$$\alpha_i(\lambda_i\alpha_i + \beta_i) \leq \sum_{i \in S\backslash R} \alpha_i(\lambda_i\alpha_i + \beta_i) \leq -\sum_{i \in S} \alpha_i(\lambda_i\alpha_i + \beta_i) \leq c,$$

where we define $c = \sum_{i \in S} \frac{\beta_i^2}{4\lambda_i}$. By the quadratic formula, $\alpha_i(\lambda_i\alpha_i + \beta_i) - c \leq 0$ implies

$$\frac{-\beta_i - \sqrt{\beta_i^2 + 4\lambda_i c}}{2\lambda_i} \leq \alpha_i \leq \frac{-\beta_i + \sqrt{\beta_i^2 + 4\lambda_i c}}{2\lambda_i}.$$

(Note that $\lambda_i > 0$ and $c > 0$ implies that the square root is well-defined). It follows that

$$\alpha_i^2 \leq 2\frac{\beta_i^2 + \beta_i^2 + 4\lambda_i c}{4\lambda_i^2} = \frac{\beta_i^2}{\lambda_i^2} + 2c/\lambda_i, \quad (i \in S\backslash R),$$

and thus

$$\begin{aligned} ||x||^2 = \sum_{i=1}^{d} \alpha_i^2 &\leq \sum_{i \in R} \lambda_i^{-2}\beta_i^2 + \sum_{i \in S\backslash R} \left(\frac{\beta_i^2}{\lambda_i^2} + \frac{2}{\lambda_i} \sum_{j \in R} \frac{\beta_i^2}{4\lambda_j} \right) \\ &\leq \sum_{i=1}^{d} \lambda_i^{-2}\beta_i^2 + \frac{1}{2} \left(\sum_{i \in S\backslash R} \frac{1}{\lambda_i} \right) \left(\sum_{j \in R} \frac{1}{\lambda_j} \right) \left(\sum_{i=1}^{n} \beta_i^2 \right) \\ &\leq \left|\left|A^{-1}b\right|\right|^2 + (d-1)^2 \frac{1}{\lambda_1} \frac{1}{\lambda_2} ||b||^2, \end{aligned}$$

where we used $\left|\left|A^{-1}b\right|\right|^2 = \sum_{j=1}^d \beta_j^2 \lambda_j^{-2}$ and $\left(\sum_{i \in S\backslash R} 1\right)\left(\sum_{j \in R} 1\right) \leq 2(d-1)^2$. □

Lemma 7.8. *Let $(x_i)_{i\in\mathbb{N}}$ be a sequence of vectors in $\mathbb{R}^d$, and $(w_i)_{i\in\mathbb{N}}$ a sequence of scalars with $0 < \inf_{i\in\mathbb{N}} w_i$. Then for all $n \in \mathbb{N}$,*

$$\lambda_{min}\left(\sum_{i=1}^{n} x_i x_i^T w_i\right) \geq \lambda_{min}\left(\sum_{i=1}^{n} x_i x_i^T\right)(\inf_{i\in\mathbb{N}} w_i).$$

Proof. For all $z \in \mathbb{R}^d$,

$$z^T\left(\sum_{i=1}^{n} x_i x_i^T w_i\right) z \geq (\inf_{i\in\mathbb{N}} w_i) z^T\left(\sum_{i=1}^{n} x_i x_i^T\right) z.$$

Let $\tilde{v}$ be a normalized eigenvector corresponding to $\lambda_{\min}\left(\sum_{i=1}^{n} x_i x_i^T w_i\right)$. Then

$$\begin{aligned}\lambda_{\min}\left(\sum_{i=1}^{n} x_i x_i^T\right) &= \min_{||v||=1} v^T\left(\sum_{i=1}^{n} x_i x_i^T\right) v \leq \tilde{v}^T\left(\sum_{i=1}^{n} x_i x_i^T\right)\tilde{v} \leq \tilde{v}^T\left(\sum_{i=1}^{n} x_i x_i^T w_i\right)\tilde{v}(\inf_{i\in\mathbb{N}} w_i)^{-1} \\ &= \lambda_{\min}\left(\sum_{i=1}^{n} x_i x_i^T w_i\right)(\inf_{i\in\mathbb{N}} w_i)^{-1}.\end{aligned}$$

□

Future directions

In this section we identify several open questions and relevant avenues for future research, which could lead to interesting extensions of the results obtained in this thesis.

We show in Chapter 3 that certainty equivalent pricing in a single product setting, with infinite inventory and linear demand function, is not strongly consistent. More precisely, we show that with positive probability the price converges to the highest admissible price, which is different from the optimal price. Simulations however suggest that a much stronger result holds: the price sequence converges with probability one, but the limit price equals the optimal price with probability zero. This means that with certainty equivalent pricing the price never converges to the optimal price. This has, however, not yet been formally proven.

In Chapter 4, we study dynamic pricing and learning for multiple products with infinite inventory, and in Chapter 5, we consider a single product with finite inventory. A natural extension is to combine both settings, and to consider multiple products with finite inventories. Some additional modeling is then needed to describe what happens if some of the products are out-of-stock: does this increase the demand for substitute products? And how does this depend on the selling prices? An interesting question is whether a certainty equivalent policy has a good performance, as in Chapter 5, or whether exogenous price dispersion is necessary.

The excellent performance of certainty equivalent pricing in the finite capacity setting of Chapter 5 can be explained by an endogenous learning property: using a policy that is optimal with respect to a parameter estimate close to the true parameter induces price dispersion, which leads to fast learning of the parameters. This property does not only occur in dynamic pricing problems, but possibly also in many other Markov decision problems. One could formulate a general framework of repeated Markov decision problems with incomplete information, and study how the presence or absence of an endogenous learning property influences the performance of a certainty equivalent policy. Another interesting question is whether the derived $O(\log^2(T))$ bound on the regret can be improved by any pricing strategy.

The demand function of Chapter 6 consists of an unknown, time-varying part and a known, price-dependent part. An extension that would be of great practical interest is to assume that the price-dependent part is also unknown to the firm and has to be learned from data. One step further, one could assume that this term is time-varying as well. It is not clear if a certainty equivalent pricing policy performs well in such a setting, or that active price experimentation is necessary. Two other interesting extensions related to pricing in time-varying markets are models with finite inventory (see, for instance, the recent work by Besbes and Saure, 2012), and non-parametric approaches (for example, similar to Yu and Mannor, 2011).

Selling prices of competitors often play a major role in the pricing strategy of a firm, and it therefore seems natural to study pricing and learning policies in a competitive environment. Neglecting competitive aspects may have negative consequences, as shown in Cooper et al. (2012). The area of repeated games with incomplete information may provide a natural framework to study

dynamic pricing and learning in a competitive environment. However, even without incomplete information, the long-term dynamics of repeated games can be very complicated. This possibly explains why the literature on dynamic pricing and learning with competition is still relatively scarce.

Recent literature on pricing and assortment problems reveals that it may be beneficial to consider pricing decisions and assortment planning simultaneously instead of separately. The same holds for pricing and inventory problems: in many situations, it is beneficial to determine a joint pricing and inventory replenishment strategy. These observations imply that it may be rewarding to study data-driven policies for these combined decision problems under uncertainty. For joint pricing and assortment problems, a first step in this direction is taken by Talebian et al. (2012).

A further suggestion for future research is related to the quality of the upper bounds on mean square convergence rates of maximum quasi-likelihood estimators, which are studied in Chapter 7. In Theorem 7.2, we provide such upper bounds for general link functions. Application of these bounds in Chapter 4 leads to Regret$(T) = O(T^{2/3})$ for the pricing policy studied there. It is an open question whether this growth rate on the regret can be improved upon, and whether the bounds in Theorem 7.2 are tight. Likewise, the growth rate Regret$(T) = O(\sqrt{T \log(T)})$ of the policy discussed in Section 4.4.2 is based on Theorem 7.3, where upper bounds on the mean square convergence rates for maximum quasi-likelihood estimators are obtained in case of canonical link functions. These bounds involve a $\log(t)$-term, which causes the factor $\sqrt{\log(T)}$ in the regret bounds. It is an open question whether this $\log(t)$-term is only a result of the used proof techniques, or that this term really is present in the asymptotic behavior of the regret. In the former case, removing this $\sqrt{\log(T)}$-term in the regret bound would imply that, for canonical link functions, the pricing policy of Chapter 4 has Regret$(T) = O(\sqrt{T})$. In single-product settings such policies are asymptotically optimal, in the sense that there is no policy with Regret$(T) = o(\sqrt{T})$.

The MLE-cycle policy by Broder and Rusmevichientong (2012) is an example of a pricing policy that achieves optimal rate Regret$(T) = O(\sqrt{T})$. However, in this policy the unknown parameters are estimated using only a small subset of the available sales data. The fact that this subset is non-random and determined in advance greatly simplifies the mathematical analysis. In contrast, the pricing policies of Chapter 3 and 4 do always use all available sales data; this comes with the price of a small additional term in the regret bounds. It would be somewhat remarkable and counterintuitive if a policy that uses all available data does always perform worse than a policy that neglects a large part of the data. However, den Boer (2012b) shows that adding data does not necessarily improve the expected quality of parameter estimates. It is interesting to study conditions that guarantee improved parameter estimates when data is added, and to apply these results to different pricing policies.

Nederlandse samenvatting

Dynamisch prijzen is voor veel commerciële bedrijven een onmisbaar onderdeel van het prijsbeleid. Het kernidee van dynamisch prijzen is dat verkoopprijzen op een slimme manier continu aangepast kunnen worden aan veranderende omstandigheden. Dit omvat zowel externe factoren (bijvoorbeeld variërende marktomstandigheden of prijswijzigingen bij concurrenten) als interne factoren (bijvoorbeeld wisselende voorraadniveaus van de aangeboden producten). Dynamisch prijzen is met name effectief als verkoopprijzen eenvoudig en kosteloos zijn aan te passen, zoals bijvoorbeeld bij internetverkoop of als gebruik wordt gemaakt van digitale prijsetiketten.

In zulke digitale verkoopomgevingen is vaak veel historische verkoopdata beschikbaar, met waardevolle informatie over klantengedrag en marktomstandigheden. Zulke gegevens hebben een enorm potentieel om prijsbeslissingen van bedrijven te verbeteren, en een belangrijke vraag is dan ook hoe deze gegevensstroom effectief gebruikt kan worden om optimale verkoopprijzen te genereren. Idealiter zou een bedrijf steeds beter willen "leren" hoe klanten reageren op verschillende verkoopprijzen naarmate er meer gegevens beschikbaar komen, om vervolgens haar prijsbeleid daar op aan te passen. Er is dan een continue wederzijdse beïnvloeding tussen de prijsbeslissingen van het bedrijf en de verkoopdata die daardoor voortgebracht wordt. In dit proefschrift bestuderen we de vraag hoe deze "combinatie van dynamisch prijzen en leren" het beste gerealiseerd kan worden.

Een belangrijke eerste stap om bruikbare informatie te abstraheren uit de stroom van verkoopgegevens is het formuleren van een vraagmodel: een wiskundige beschrijving van de relatie tussen verkoopprijzen en de vraag naar producten. Zo'n model bevat doorgaans een aantal onbekende parameters, en het model wordt pas bruikbaar in de praktijk als een schatting van de waarden van deze parameters beschikbaar is. Er bestaan diverse statistische technieken om op basis van de beschikbare verkoopgegevens zulke schattingen te bepalen. Telkens als er nieuwe verkoopgegevens bijkomen kunnen deze schattingen worden bijgewerkt, zodat deze op steeds meer gegevens zijn gebaseerd. In het ideale geval leidt dit ertoe dat, naarmate er meer verkoopgegevens beschikbaar komen, het vraagmodel een steeds betere beschrijving van de werkelijkheid geeft, en het bedrijf zo steeds beter "leert" hoe de vraag naar haar producten afhangt van de verkoopprijzen.

Dit leren is geen vanzelfsprekendheid. Sommige prijsstrategieën die in de praktijk veel gebruikt worden staan leren juist in de weg. Eén van deze strategieën is het zogenaamde "certainty equivalent" (CE) prijzen, waarbij, op elk moment dat de prijs wordt aangepast, de prijs wordt gekozen die optimaal is als men aanneemt dat de dan beschikbare statistische schattingen correct zijn. Met andere woorden, men kiest altijd voor de verkoopprijs die op het moment van beslissen het beste lijkt. Dit lijkt een logische en intuïtieve manier van prijzen bepalen: waarom zou men afwijken van wat de beste beslissing lijkt te zijn? Het blijkt echter dat deze manier van prijzen in sommige situaties tot grote verliezen kan leiden. Een illustratief voorbeeld hiervan wordt beschreven in hoofdstuk 3.

Een verklaring voor dit enigszins tegenintuïtieve en verrassende resultaat ligt in het feit dat het leren van parameterwaarden en het bepalen van optimale prijzen niet twee volkomen onafhankelijke processen zijn, maar juist sterk van elkaar afhangen. De kwaliteit van prijsbeslissingen wordt rechtstreeks beïnvloed door de kwaliteit van de schatters in het vraagmodel, en deze worden op hun beurt sterk beïnvloed door de mate van variatie of spreiding in de verkoopprijzen. Over het algemeen geldt: hoe meer spreiding in de prijzen, hoe betere schattingen. Spreiding in prijzen betekent echter ook dat er afgeweken wordt van de optimale prijs, en dit brengt kosten met zich mee. Een goede prijsstrategie zorgt dus voor een optimale balans tussen enerzijds optimalisatie van de verwachte opbrengsten op korte termijn, en anderzijds de optimalisatie van het leerproces. Bij CE prijzen wordt geen rekening gehouden met deze balans, maar ligt de nadruk volledig op het maximaliseren van de directe opbrengsten. Dit gaat ten koste van de kwaliteit van het leerproces, wat resulteert in slechtere prijs-beslissingen op de lange termijn.

In hoofdstuk 3 stellen we een prijsstrategie voor die wél een optimale balans bereikt: "controlled variance pricing" (CVP). Het idee is dat de prijzen zo dicht mogelijk bij de CE prijzen worden gekozen, maar met een extra conditie die een bepaalde hoeveelheid prijsvariatie garandeert. Door deze mate van prijsvariate zorgvuldig te fine-tunen kan deze prijsstrategie een goede balans vinden tussen de twee doelen van winstmaximalisatie en optimalisatie van het leerproces. We leiden structurele resultaten af over de kwaliteit van deze prijsstrategie (dat wil zeggen: resultaten die onafhankelijk zijn van specifieke getallenvoorbeelden). Hieruit blijkt dat CVP in structurele zin vrijwel niet meer te verbeteren is. Daarnaast is CVP eenvoudig implementeerbaar en geschikt voor een zeer grote klasse van vraagmodellen, waardoor deze prijsstrategie in veel praktische toepassingen van nut kan zijn.

De vraag naar een product hangt doorgaans niet alleen af van de eigen prijs, maar ook van prijzen van vergelijkbare producten die aangeboden worden. Als een bedrijf meerdere producten aanbiedt hangen de optimale verkoopprijzen niet alleen af van de prijselasticiteit van de afzonderlijke producten, maar ook van de verschillende substitutie-effecten. Dit vraagt om een gemeenschappelijke prijsstrategie voor het gehele aanbod van producten.

In hoofdstuk 4 introduceren we een prijsstrategie die in staat is om uit de stroom van binnenkomende verkoopdata al deze substitutie-effecten te leren. De prijsstrategie is gebaseerd op dezelfde principes als CVP: het optimaal balanceren van winstmaximalisatie op de korte termijn en het leren van alle onbekende parameters op de lange termijn. Net als bij de situatie met één product vindt het leren alleen plaats als er voldoende prijsvariatie is. Door deze prijsvariatie in meerdere dimensies op een slimme manier te meten kunnen we een prijsstrategie formuleren die in staat is om alle onbekende prijselasticiteiten en substitutie-effecten te leren. We leiden structurele resultaten af over de kwaliteit van deze prijsstrategie. Hieruit blijkt ondermeer dat, voor een bepaalde klasse van vraagmodellen, dynamisch prijzen en leren met meerdere producten niet structureel moeilijker is dan met één product.

In de situatie beschreven in hoofdstuk 3 en 4 worden de verkoopprijzen niet beïnvloed door voorraadniveaus van de aangeboden producten. Er zijn echter ook producten waarvan de optimale prijs sterk afhangt van de resterende voorraad. Dit is in het bijzonder het geval bij producten die in beperkte hoeveelheid gedurende een begrensde periode worden aangeboden, zoals vliegtickets, hotelkamerreserveringen en concertkaartjes. Voor dit type producten geldt dat alle voorraad die gedurende de verkoopperiode niet wordt verkocht geen geld oplevert. Als gevolg hiervan is het optimaal om de verkoopprijs continu aan te passen, afhankelijk van de resterende voorraad en de duur van de resterende verkoopperiode. Als er nog veel producten op voorraad zijn die nog maar korte tijd verkocht kunnen worden, dan is het voordelig om de prijs te laten zakken; als er nog slechts enkele producten beschikbaar zijn die nog vrij lange tijd verkocht kunnen worden, dan is het optimaal om de prijs te verhogen. Dit continu aanpassen van de verkoopprijs, afhankelijk van de resterende voorraad en de duur van de resterende verkoopperiode,

wordt al in veel branches toegepast.

Hoofdstuk 5 behandelt dynamisch prijzen en leren voor de situatie dat een beperkte voorraad gedurende een begrensde periode aangeboden wordt. Net als in hoofdstuk 3 en 4 is er een onderliggend vraagmodel dat de relatie tussen vraag en prijs beschrijft. Onbekende parameters in het vraagmodel worden geschat op grond van binnenkomende verkoopdata, gebruikmakend van statistische technieken. We bestuderen de vraag wat een goede prijsstrategie is en hoe de kwaliteit ervan gekarakteriseerd kan worden.

Een verrassend resultaat uit hoofdstuk 5 is dat de eenvoudige CE strategie in de situatie met eindige voorraad en eindige verkoopperiode uitstekend werkt. De firma hoeft geen rekening te houden met het optimaliseren van het leerproces, maar kan op elk moment eenvoudigweg de prijs kiezen die optimaal lijkt op grond van de beschikbare parameterschattingen. De onbekende parameters worden "geleerd" naarmate er steeds meer data beschikbaar komt, en vergeleken met de situatie beschreven in hoofdstuk 3 en 4 gebeurt dit leren ook nog eens zeer snel.

De verklaring voor deze uitkomsten ligt in de grote mate van variatie in verkoopprijzen die de CE prijsstrategie genereert. De optimale verkoopprijs van het aangeboden product wordt bepaald door de hoeveelheid resterende voorraad en de duur van de resterende verkoopperiode; omdat deze twee constant wijzigen, varieert de optimale verkoopprijs ook voortdurend. Daardoor treedt er als vanzelf veel prijsvariatie op, waardoor de onbekende parameters van het vraagmodel zeer snel geleerd worden. Dit snelle leren leidt er toe dat de CE prijzen al snel van goede kwaliteit zijn. Ter vergelijking: in de situatie beschreven in hoofdstuk 3 en 4 moet de firma zelf actief zorgen voor voldoende prijsvariatie om uiteindelijk de onbekende parameters te leren. Leren is in die situatie daardoor in structurele zin veel moeilijker.

Het leren van klantengedrag en marktkarakteristieken uit binnenkomende verkoopdata veronderstelt een zekere stabiliteit in de markt. Fluctuaties in de marktomstandigheden hebben doorgaans een negatief effect op de kwaliteit van het leerproces. In de praktijk zijn veranderingen in de markt vaak eerder regel dan uitzondering: technologische innovaties, marketing campagnes, prijsveranderingen bij concurrenten en vele andere gebeurtenissen kunnen veranderingen in de relatie tussen vraag en prijs veroorzaken. Deze veranderingen verhinderen het daadwerkelijk leren van de vraag-prijs relatie. Dit roept de vraag op of de binnenkomende verkoopdata dan nog wel bruikbaar is voor het bepalen van verkoopprijzen, en zo ja, hoe dat dan het beste kan gebeuren.

Hoofdstuk 6 behandelt dynamisch prijzen en leren in een omgeving met veranderende marktomstandigheden. We introduceren twee technieken om fluctuerende marktcijfers te schatten uit historische verkoopdata, en karakteriseren de kwaliteit van deze schattingsmethodes. Beide technieken kennen aan elk datapunt een gewicht toe dat bepaalt hoeveel invloed dit punt heeft op de schatter; deze gewichten worden steeds kleiner naarmate de corresponderende data langer geleden is gegenereerd. Door deze gewichten zorgvuldig te kiezen kan de verkoper een schatter definiëren die zowel snel reageert op veranderingen in de markt, als ook gebaseerd is op zoveel mogelijk relevante data.

We bestuderen een eenvoudige CE prijsstrategie, die op elk beslismoment op basis van de beschikbare schattingen de optimale verkoopprijs kiest. We karakteriseren de kwaliteit van deze prijsstrategie, en laten zien hoe vervolgens de optimale gewichten in de schatter gekozen moeten worden. Dit resulteert in een methodologie die een bedrijf in staat stelt dynamisch te prijzen en te leren in een veranderende markt, en zo gefundeerde prijsbeslissingen te nemen en actief te reageren op fluctuaties in de markt.

In hoofdstuk 7 onderzoeken we de relatie tussen de kwaliteit van statistische schatters en de mate van spreiding in de design variabelen. De resultaten uit dit hoofdstuk worden veelvuldig

toegepast in de hoofdstukken 3 - 6 om de kwaliteit van prijsstrategieën te karakteriseren.

De resultaten in dit proefschrift zijn niet alleen relevant voor dynamisch prijzen, maar in feite voor elk beslissingsprobleem waarbij er meerdere beslissingen na elkaar genomen moeten worden en waarbij de kwaliteit van beslissingen niet van tevoren bekend is maar slechts "al doende" geleerd kan worden. Belangrijke vragen bij zulke beslissingsproblemen zijn: is het genoeg om altijd de beslissing te nemen die, op het moment van beslissen, het beste lijkt? Of moet daar soms bewust van afgeweken worden, met het doel meer te leren over hoe het systeem werkt, ook als dit kosten met zich meebrengt? Hoe snel vindt dit leren plaats? En wat gebeurt er als het gedrag van het systeem aan veranderingen onderhevig is? Dit proefschrift behandelt deze vragen in de context van dynamisch prijzen en leren.

Bibliography

E. Adida and G. Perakis. A robust optimization approach to dynamic pricing and inventory control with no backorders. *Mathematical Programming, Series B*, 107(1):97–129, 2006.

P. Aghion, P. Bolton, C. Harris, and B. Jullien. Optimal learning by experimentation. *Review of Economic Studies*, 58(4):621–654, 1991.

H. S. Ahn, M. Gumus, and P. Kaminsky. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Operations Research*, 55(6):1039–1057, 2007.

T. W. Anderson and J. B. Taylor. Some experimental results on the statistical properties of least squares estimates in control problems. *Econometrica*, 44(6):1289–1302, 1976.

M. Aoki. On a dual control approach to the pricing policies of a trading specialist. In R. Conti and A. Ruberti, editors, *5th Conference on Optimization Techniques Part II*, volume 4 of *Lecture Notes in Computer Science*, chapter 24, pages 272–282. Springer, Berlin, Heidelberg, 1973.

M. Aoki. On some price adjustment schemes. *Annals of Economic and Social Measurement*, 3(1): 95–115, 1974.

V. F. Araman and R. Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57(5):1169–1188, 2009.

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

Y. Aviv and A. Pazgal. A partially observed Markov decision process for dynamic pricing. *Management Science*, 51(9):1400–1416, 2005a.

Y. Aviv and A. Pazgal. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, 10(3):339–359, 2008.

Y. Aviv and A. Pazgal. Pricing of short life-cycle products through active learning. Working paper, 2005b.

J. Baetge, G. Bolenz, W. Ballwieser, R. Hömberg, and P. Wullers. Dynamic price policies in monopolistic competition. In J. Rose and C. Bilciu, editors, *Modern Trends in Cybernetics and Systems. Proceedings of the Third International Congress of Cybernetics and Systems, Bucharest, Romania, August 25-29, 1975*, volume 1, pages 401–417. Springer Verlag, 1977.

R. J. Balvers and T. F. Cosimano. Actively learning about demand and the dynamics of price adjustment. *The Economic Journal*, 100(402):882–898, 1990.

R. Baptista. The diffusion of process innovations: a selective review. *International Journal of the Economics of Business*, 6(1):107–129, 1999.

D. P. Baron. Price uncertainty, utility, and industry equilibrium in pure competition. *International Economic Review*, 11(3):463–480, 1970.

D. P. Baron. Demand uncertainty in imperfect competition. *International Economic Review*, 12(2): 196–208, 1971.

M. S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 22(1):107–111, 1951.

F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5): 215–227, 1969.

W. J. Baumol and R. E. Quandt. Rules of thumb and optimally imperfect decisions. *The American Economic Review*, 54(2):23–46, 1964.

R. Benini. Sull'uso delle formole empiriche a nell'economia applicata. *Giornale degli economisti, 2nd series*, 35:1053–1063, 1907.

D. Bergemann and K. H. Schlag. Pricing without priors. *Journal of the European Economic Association*, 6(2-3):560–569, 2008a.

D. Bergemann and K. H. Schlag. Robust monopoly pricing. Cowles Foundation Discussion Paper 1527R, 2008b.

D. P. Bertsekas. *Constrained optimization and Lagrange Multiplier methods*. Computer Science and Applied Mathematics. Academic Press, New York, 1982.

D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, 1999.

D. Bertsimas and G. Perakis. Dynamic pricing: a learning approach. In *Mathematical and Computational Models for Congestion Charging*, pages 45–79. Springer, New York, 2006.

O. Besbes and C. Maglaras. Dynamic pricing with financial milestones: feedback-form policies. *Management Science*, to appear, 2012.

O. Besbes and D. Saure. Dynamic pricing strategies in the presence of demand shocks. Working paper, 2012.

O. Besbes and A. Zeevi. Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

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

O. Besbes and A. Zeevi. Blind network revenue management. *Operations Research*, to appear, 2012.

R. Bhatia. *Matrix Analysis*. Springer Verlag, New York, 1997.

R. Bhatia. *Positive Definite Matrices*. Princeton University Press, Princeton, 2007.

F. Billström and S. Thore. Simulation experiments with dynamic price strategies in monopoly theory. In H. O. A. Wold, editor, *Econometric model building. Essays on the Causal Chain Approach*, pages 297–321. North Holland Publishing Co., Amsterdam, 1964.

F. Billström, S. Thore, L. O. Friberg, H. Laadi, O. Johansson, H. O. A. Wold, and B. Hansen. Some experiments with dynamization of monopoly models. 16th Meeting of the Econometric Society in Uppsala, 1954.

A. Bisi and M. Dada. Dynamic learning, pricing, and ordering by a censored newsvendor. *Naval Research Logistics*, 54(4):448–461, 2007.

G. Bitran and R. Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–230, 2003.

G. Bitran, R. Caldentey, and S. Mondschein. Coordinating clearance markdown sales of seasonal products in retail chains. *Operations Research*, 46(5):609–624, 1998.

G. R. Bitran and S. V. Mondschein. Periodic pricing of seasonal products in retailing. *Management Science*, 43(1):64–79, 1997.

G. R. Bitran and H. K. Wadhwa. A methodology for demand learning with an application to the optimal pricing of seasonal products. Working paper, WP#3898-96, 1996.

A. V. den Boer. Dynamic pricing with multiple products and partially specified demand distribution. Submitted for publication, 2011.

A. V. den Boer. Tracking the market: Dynamic pricing and learning in a changing environment. Working paper, 2012a.

A. V. den Boer. Does adding data always improve linear regression estimates? Submitted for publication, 2012b.

A. V. den Boer and B. Zwart. Simultaneously learning and optimizing using controlled variance pricing. Submitted for publication, 2010.

A. V. den Boer and B. Zwart. Mean square convergence rates for maximum quasi-likelihood estimators. Working paper, 2011a.

A. V. den Boer and B. Zwart. Dynamic pricing and learning with finite inventories. Submitted for publication, 2011b.

A. Borovkov. *Mathematical Statistics*. Gordon and Breach Science Publishers, Amsterdam, 1998.

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

A. Brown and A. Deaton. Surveys in applied economics: models of consumer behaviour. *The Economic Journal*, 82(328):1145–1236, 1972.

A. N. Burnetas and C. E. Smith. Adaptive ordering and pricing for perishable products. *Operations Research*, 48(3):436–443, 2000.

G. P. Cachon and R. Swinney. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science*, 55(3):497–511, 2009.

A. X. Carvalho and M. L. Puterman. Learning and pricing in an internet environment with binomial demand. *Journal of Revenue and Pricing Management*, 3(4):320–336, 2005a.

A. X. Carvalho and M. L. Puterman. Dynamic optimization and learning: How should a manager set prices when the demand function is unknown? Technical report, Instituto de Pesquisa Economica Aplicada - IPEA, Discussion Papers 1117, 2005b.

N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, New York, 2006.

L. Chan, Z. Shen, and D. Simchi-Levi. Coordination of pricing and inventory decisions: a survey and classification. In D. Simchi-Levi, D. Wu, and Z. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, pages 335–392. Springer, New York, 2004.

Y. I. Chang. Strong consistency of maximum quasi-likelihood estimate in generalized linear models via a last time. *Statistics & Probability Letters*, 45(3):237–246, 1999.

K. Chen and I. Hu. On consistency of Bayes estimates in a certainty equivalence adaptive system. *IEEE Transactions on Automatic Control*, 43(7):943–947, 1998.

K. Chen, I. Hu, and Z. Ying. Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *The Annals of Statistics*, 27(4):1155–1163, 1999.

X. Chen and D. Simchi-Levi. Joint pricing and inventory management. In Ö. Özer and R. Phillips, editors, *The Oxford Handbook of Pricing Management*, chapter 30. Oxford University Press, London, 2012.

Y. M. Chen and D. C. Jain. Dynamic monopoly pricing under a Poisson-type uncertain demand. *The Journal of Business*, 65(4):593–614, 1992.

V. L. R. Chinthalapati, N. Yadati, and R. Karumanchi. Learning dynamic prices in MultiSeller electronic retail markets with price sensitive customers, stochastic demands, and inventory replenishments. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 36(1):92–106, 2006.

C. Y. Chong and D. Cheng. Multistage pricing under uncertain demand. In S. V. Berg, editor, *Annals of Economic and Social Measurement*, volume 4, pages 311–323. NBER, 1975.

Y. S. Chow and H. Teicher. *Probability theory: independence, interchangeability, martingales*. Springer Verlag, New York, third edition, 2003.

C. F. Christ. Early progress in estimating quantitative economic relationships in America. *The American Economic Review*, 75(6):39–52, 1985.

B. D. Chung, J. Li, T. Yao, C. Kwon, and T. L. Friesz. Demand learning and dynamic pricing under competition in a state-space framework. *Engineering Management, IEEE Transactions on*, 59(2): 240–249, 2012.

D. G. Clarke and R. J. Dolan. A simulation analysis of alternative pricing strategies for dynamic environments. *The Journal of Business*, 57(1):S179–S200, 1984.

F. Clarke, M. N. Darrough, and J. Heineke. Optimal pricing policy in the presence of experience effects. *The Journal of Business*, 55(4):517–530, 1982.

R. W. Clower. Some theory of an ignorant monopolist. *The Economic Journal*, 69(276):705–716, 1959.

W. L. Cooper, T. Homem de Mello, and A. J. Kleywegt. Learning and pricing with models that do not explicitly incorporate competition. Working paper, 2012.

E. Cope. Bayesian strategies for dynamic pricing in e-commerce. *Naval Research Logistics*, 54(3): 265–281, 2007.

A. A. Cournot. *Researches into the Mathematical Principles of the Theory of Wealth*. Translated in 1897 by N.T. Bacon, with a bibliography of Mathematical Economics by Irving Fisher. The Macmillan Company, New York, 1838.

J. Creedy. On the King-Davenant "law" of demand. *Scottish Journal of Political Economy*, 33(3): 193–212, 1986.

C. Davenant. *An essay upon the probable methods of making a people gainers in the balance of trade*. James Knapton, London, 1699.

V. H. de la Peña, T. L. Lai, and Q. M. Shao. *Self-Normalized Processes: Limit Theory and Statistical Applications*. Springer Series in Probability and its Applications. Springer, New York, first edition, 2009.

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.

W. Dodds. An application of the Bass model in long-term new product forecasting. *Journal of Marketing Research*, 10(3):308–311, 1973.

R. J. Dolan and A. P. Jeuland. Experience curves and dynamic demand models: implications for optimal pricing strategies. *Journal of Marketing*, 45(1):52–62, 1981.

J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis: Differentiation*. Series: Cambridge Studies in Advanced Mathematics (No. 86). Cambridge University Press, Cambridge, 2004.

D. Easley and N. M. Kiefer. Controlling a stochastic process with unknown parameters. *Econometrica*, 56(5):1045–1064, 1988.

J. Eliashberg and R. Steinberg. Marketing-production joint decision-making. In J. Eliashberg and J. D. Lilien, editors, *Management Science in Marketing, Handbooks in Operations Research and Management Science*, volume 5, pages 827–880. North Holland Publishing Co., Amsterdam, 1993.

W. Elmaghraby and P. Keskinocak. Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions. *Management Science*, 49(10):1287–1309, 2003.

W. Elmaghraby, A. Gülcü, and P. Keskinocak. Designing optimal pre-announced markdowns in the presence of rational customers with multi-unit demands. *Manufacturing & Service Operations Management*, 10(1):126–148, 2008.

S. S. Eren and C. Maglaras. Monopoly pricing with limited demand information. *Journal of Revenue and Pricing Management*, 9:23–48, 2010.

G. C. Evans. The dynamics of monopoly. *The American Mathematical Monthly*, 31(2):77–83, 1924.

G. C. Evans. The mathematical theory of economics. *The American Mathematical Monthly*, 32(3): 104–110, 1925.

L. Fahrmeir and H. Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13(1):342–368, 1985.

R. W. Farebrother. Early explorations in econometrics. In T. C. Mills and K. Patterson, editors, *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*. Palgrave MacMillan, Basingstoke, 2006.

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

Y. Feng and G. Gallego. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, 41(8):1371–1391, 1995.

Y. Feng and G. Gallego. Perishable asset revenue management with Markovian time dependent demand intensities. *Management Science*, 46(7):941–956, 2000.

Y. Feng and B. Xiao. Optimal policies of yield management with multiple predetermined prices. *Operations Research*, 48(2):332–343, 2000a.

Y. Feng and B. Xiao. A continuous-time yield management model with multiple prices and reversible price changes. *Management Science*, 46(5):644–657, 2000b.

G. Fibich, A. Gavious, and O. Lowengart. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research*, 51(5): 721–734, 2003.

I. Fisher. Cournot and mathematical economics. *The Quarterly Journal of Economics*, 12(2):119–138, 1898.

R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, second edition, 2000.

G. Gallego and G. van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41, 1997.

G. Gallego and G. J. van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

W. Gaul and A. D. Azizi. A demand learning data based approach to optimize revenues of a retail chain. In H. Locarek-Junge and C. Weihs, editors, *Classification as a Tool for Research*, Studies in Classification, Data Analysis, and Knowledge Organization, chapter 75, pages 683–691. Springer, Berlin, Heidelberg, 2010.

J. Gill. *Generalized linear models: a unified approach*. Sage Publications, Thousand Oaks, CA, 2001.

C. Gini. Prezzi e consumi. *Giornale degli Economisti, 3rd series*, 40:99–114, 235–249, 1910.

J. C. Gittins. *Multi-armed bandit allocation indices*. Wiley-Interscience series in Systems and Optimization. Wiley, New York, 1989.

V. P. Godambe and C. C. Heyde. Quasi-likelihood and optimal estimation. *International Statistical Review*, 55(3):231–244, 1987.

B. I. Godoy, G. C. Goodwin, J. C. Aguero, and A. J. Rojas. An algorithm for estimating time-varying commodity price models. In *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference*, pages 1563–1568. IEEE, 2009.

A. Goldenshluger and A. Zeevi. Woodroofe's one-armed bandit problem revisited. *The Annals of Applied Probability*, 19(4):1603–1633, 2009.

E. A. Greenleaf. The impact of reference price effects on the profitability of price promotions. *Marketing Science*, 14(1):82–104, 1995.

S. J. Grossman, R. E. Kihlstrom, and L. J. Mirman. A Bayesian approach to the production of information and learning by doing. *The Review of Economic Studies*, 44(3):533–547, 1977.

W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31(2):221–239, 1989.

P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Probability and Mathematical Statistics, a Series of Monographs and Textbook. Academic Press, New York, 1980.

D. M. Hanssens, L. J. Parsons, and R. L. Schultz. *Market response models: econometric and time series analysis*. International series in quantitative marketing. Kluwer Academic Publishers, Boston, second edition, 2001.

J. M. Harrison, N. B. Keskin, and A. Zeevi. Bayesian dynamic pricing policies: learning and earning under a binary prior distribution. *Management Science*, to appear, 2011a.

J. M. Harrison, N. B. Keskin, and A. Zeevi. Dynamic pricing with an unknown linear demand model: asymptotically optimal semi-myopic policies. Working paper, 2011b.

E. R. Hawkins. Methods of estimating demand. *Journal of Marketing*, 21(4):428–438, 1957.

P. Heidhues and B. Köszegi. The impact of consumer loss aversion on pricing. CEPR Discussion Paper No. 4849, 2005.

C. C. Heyde. *Quasi-likelihood and its application*. Springer Series in Statistics. Springer Verlag, New York, 1997.

C. C. Heyde and R. Morton. Quasi-likelihood and generalizing the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(2):317–327, 1996.

J. R. Hicks. Annual survey of economic theory: the theory of monopoly. *Econometrica*, 3(1):1–20, 1935.

S. L. Horner, C. F. Roos, V. von Szeliski, A. T. Court, and S. M. DuBrul. *The Dynamics of Automobile Demand*. General Motors Corporation, New York, 1939.

D. Horsky and L. S. Simon. Advertising and the diffusion of new products. *Marketing Science*, 2 (1):1–17, 1983.

S. Kalish. Monopolist pricing with dynamic demand and production cost. *Marketing Science*, 2 (2):135–159, 1983.

K. Kalyanam, R. Lal, and G. Wolfram. Future store technologies and their impact on grocery retailing. In M. Krafft and M. K. Mantrala, editors, *Retailing in the 21st Century*, chapter 7, pages 95–112. Springer, Berlin, Heidelberg, 2006.

B. Kamrad, S. S. Lele, A. Siddique, and R. J. Thomas. Innovation diffusion uncertainty, advertising and pricing policies. *European Journal of Operational Research*, 164(3):829–850, 2005.

S. Karlin and C. R. Carr. Prices and optimal inventory policies. In K. Arrow, S. Karlin, and H. Scarf, editors, *Studies in Applied Probability and Management Science*, pages 159–172. Stanford University Press, Stanford, 1962.

G. Keller and S. Rady. Optimal experimentation in a changing environment. *The Review of Economic Studies*, 66(3):475–507, 1999.

N. M. Kiefer and Y. Nyarko. Optimal control of an unknown linear process with learning. *International Economic Review*, 30(3):571–586, 1989.

W. M. Kincaid and D. A. Darling. An inventory pricing problem. *Journal of Mathematical Analysis and Applications*, 7:183–208, 1963.

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

A. J. Kleywegt. An optimal control problem of dynamic pricing. Working paper, 2001.

P. K. Kopalle, A. G. Rao, and J. L. Assunção. Asymmetric reference price effects and dynamic pricing policies. *Marketing Science*, 15(1):60–85, 1996.

E. Kutschinski, T. Uthmann, and D. Polani. Learning competitive pricing strategies by multi-agent reinforcement learning. *Journal of Economic Dynamics and Control*, 27(11-12):2207–2218, 2003.

T. L. Lai. Limit theorems for delayed sums. *The Annals of Probability*, 2(3):432–440, 1974.

T. L. Lai and H. Robbins. Iterated least squares in multiperiod control. *Advances in Applied Mathematics*, 3:50–73, 1982.

T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

T. L. Lai and C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 10(1):154–166, 1982.

M. A. Lariviere. A note on probability distributions with increasing generalized failure rates. *Operations Research*, 54(3):602–604, 2006.

M. A. Lariviere and E. L. Porteus. Manufacturer-retailer contracting under an unknown demand distribution. Working paper, 1995.

E. P. Lazear. Retail pricing and clearance sales. *The American Economic Review*, 76(1):14–32, 1986.

T. Le Guen. Data-driven pricing. Master's thesis, Sloan School of Management, MIT, 2008.

P. S. H. Leeflang, T. H. A. Bijmolt, J. van Doorn, D. M. Hanssens, H. J. van Heerde, P. C. Verhoef, and J. E. Wieringa. Creating lift versus building the base: current trends in marketing dynamics. *International Journal of Research in Marketing*, 26(1):13–20, 2009.

R. A. Lehfeldt. The elasticity of demand for wheat. *The Economic Journal*, 24(94):212–217, 1914.

H. E. Leland. Theory of the firm facing uncertain demand. *The American Economic Review*, 62(3): 278–291, 1972.

J. Leray and J. Schauder. Topologie et equations fonctionelles. *Annales Scientifiques de l'École Normale Supérieure*, 51:45–78, 1934.

Y. Levin, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management Science*, 55(1):32–46, 2009.

T. Levina, Y. Levin, J. McGill, and M. Nediak. Dynamic pricing with online learning and strategic consumers: an application of the aggregating algorithm. *Operations Research*, 57(2):327–341, 2009.

A. E. B. Lim and J. G. Shanthikumar. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research*, 55(2):198–214, 2007.

A. E. B. Lim, J. G. Shanthikumar, and T. Watewai. Robust multi-product pricing. Working paper, 2008.

K. Y. Lin. Dynamic pricing with real-time demand learning. *European Journal of Operational Research*, 174(1):522–538, 2006.

Q. Liu and G. J. van Ryzin. Strategic capacity rationing to induce early purchases. *Management Science*, 54(6):1115–1131, 2008.

R. Lobel and G. Perakis. Dynamic pricing through sampling based optimization. Working paper, 2011.

M. S. Lobo and S. Boyd. Pricing and learning with uncertain demand. Working paper, 2003.

M. Loève. *Probability Theory I*. Springer Verlag, New York, Berlin, Heidelberg, 4th edition edition, 1977a.

M. Loève. *Probability Theory II*. Springer Verlag, New York, Berlin, Heidelberg, 4th edition edition, 1977b.

V. Mahajan, E. Muller, and F. M. Bass. New product diffusion models in marketing: a review and directions for research. *Journal of Marketing*, 54(1):1–26, 1990.

R. Manning. Market research by a monopolist: a Bayesian analysis. *Economica*, 46(183):301–306, 1979.

T. Mazumdar, S. P. Raj, and I. Sinha. Reference price research: review and propositions. *Journal of Marketing*, 69(4):84–102, 2005.

P. McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11(1):59–67, 1983.

P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman & Hall, London, 1983.

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

N. Meade and T. Islam. Modelling and forecasting the diffusion of innovation - a 25-year review. *International Journal of Forecasting*, 22(3):519–545, 2006.

E. S. Mills. Uncertainty and price theory. *The Quarterly Journal of Economics*, 73(1):116–130, 1959.

L. J. Mirman, L. Samuelson, and A. Urbano. Monopoly experimentation. *International Economic Review*, 34(3):549–563, 1993.

H. L. Moore. *Ecomomic Cycles; Their Law and Cause*. The Macmillan Company, New York, 1914.

H. L. Moore. *Forecasting the Yield and the Price of Cotton*. The Macmillan Company, New York, 1917.

P. B. Mullen, C. K. Monson, K. D. Seppi, and S. C. Warnick. Particle swarm optimization in dynamic pricing. In *2006 IEEE International Conference on Evolutionary Computation*, pages 1232–1239. IEEE, 2006.

K. Nassiri-Toussi and W. Ren. On the convergence of least squares estimates in white noise. *IEEE Transactions on Automatic Control*, 39(2):364–368, 1994.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, 135(3):370–384, 1972.

A. J. Nevins. Some effects of uncertainty: simulation of a model of price. *The Quarterly Journal of Economics*, 80(1):73–87, 1966.

D. Nguyen. The monopolistic firm, random demand, and Bayesian learning. *Operations Research*, 32(5):1038–1051, 1984.

D. Nguyen. Marketing decisions under uncertainty. volume 6 of *International Series in Quantitative Marketing*, chapter 2, pages 23–57. Springer, Boston, MA, 1997.

J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30 of *SIAM's Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2000.

M. R. Osborne. Fisher's method of scoring. *International Statistical Review / Revue Internationale de Statistique*, 60(1):99–117, 1992.

Ö. Özer and R. Phillips, editors. *The Oxford Handbook of Pricing Management*. Oxford University Press, London, 2012.

N. C. Petruzzi and M. Dada. Dynamic pricing and inventory control with learning. *Naval Research Logistics*, 49(3):303–325, 2002.

R. Phillips. *Pricing and Revenue Optimization*. Stanford University Press, Stanford, CA, 2005.

I. Popescu and Y. Wu. Dynamic pricing strategies with reference effects. *Operations Research*, 55 (3):413–429, 2007.

W. B. Powell. The knowledge gradient for optimal learning. In J. J. Cochran, editor, *Encyclopedia for Operations Research and Management Science*. Wiley, New York, 2010.

E. C. Prescott. The multi-period control problem under uncertainty. *Econometrica*, 40(6):1043–1058, 1972.

L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44(2): 303–325, 2008.

M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, first edition, 1994.

K. Raman and R. Chatterjee. Optimal monopolist pricing under demand uncertainty in dynamic markets. *Management Science*, 41(1):144–162, 1995.

S. Ramezani, P. A. N. Bosman, and H. La Poutre. Adaptive strategies for dynamic pricing agents. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pages 323–328. IEEE, 2011.

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

B. Robinson and C. Lakhani. Dynamic price models for new-product planning. *Management Science*, 21(10):1113–1122, 1975.

C. F. Roos. A mathematical theory of competition. *American Journal of Mathematics*, 47(3):163–175, 1925.

C. F. Roos. A dynamic theory of economics. *Journal of Political Economy*, 35(5):632–656, 1927a.

C. F. Roos. Dynamical economics. *Proceedings of the National Academy of Sciences of the United States of America*, 13(3):145–150, 1927b.

C. F. Roos. *Dynamic Economics: Theoretical and Statistical Studies of Demand, Production, and Prices*. The Principia Press, Bloomington, 1934.

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

P. Rusmevichientong and J. N. Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

A. Rustichini and A. Wolinsky. Learning about variable demand in the long run. *Journal of Economic Dynamics and Control*, 19(5-7):1283–1292, 1995.

A. Sandmo. On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1):65–73, 1971.

H. Schultz. The statistical law of demand as illustrated by the demand for sugar. *Journal of Political Economy*, 33(6):481–504, 1925.

H. Schultz. *The theory and measurement of demand*. University of Chicago Press, Chicago, 1938.

A. Sen and A. X. Zhang. Style goods pricing with demand learning. *European Journal of Operational Research*, 196(3):1058–1075, 2009.

M. A. Simaan and T. Takayama. Optimum monopolist control in a dynamic market. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(12):799–807, 1976.

C. G. Small, J. Wang, and Z. Yang. Eliminating multiple root problems in estimation. *Statistical Science*, 15(4):313–332, 2000.

A. Smithies. The maximization of profits over time with changing cost and demand functions. *Econometrica*, 7(4):312–318, 1939.

A. Spataru. Improved convergence rates for tail probabilities. *Bulletin of the Transilvania University of Brasov - Series III: Mathematics, Informatics, Physics*, 2(51):137–142, 2009.

G. J. Stigler. The early history of empirical studies of consumer behavior. *Journal of Political Economy*, 62(2):95–113, 1954.

G. J. Stigler. Henry L. Moore and statistical economics. *Econometrica*, 30(1):1–21, 1962.

G. Stoica. Baum-Katz-Nagaev type results for martingales. *Journal of Mathematical Analysis and Applications*, 336(2):1489–1492, 2007.

W. F. Stout. A martingale analogue of Kolmogorov's law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 15(4):279–290, 1970.

X. Su. Optimal pricing with speculators and strategic consumers. *Management Science*, 56(1): 25–40, 2010.

S. Subrahmanyan and R. Shoemaker. Developing optimal pricing and inventory policies for retailers who face uncertain demand. *Journal of Retailing*, 72(1):7–30, 1996.

M. Talebian, N. Boland, and M. Savelsbergh. Assortment and pricing with demand learning. Working paper, 2012.

K. T. Talluri and G. J. van Ryzin. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, 2004.

P. Tehrani, Y. Zhai, and Q. Zhao. Dynamic pricing under finite space demand uncertainty: a multi-armed bandit with dependent arms. Working paper, 2012.

A. Thiele. Single-product pricing via robust optimization. Working paper, 2006.

A. Thiele. Multi-product pricing via robust optimisation. *Journal of Revenue and Pricing Management*, 8:67–80, 2009.

R. G. Thompson, M. D. George, P. L. Brown, and M. S. Proctor. Optimal production, investment, and output price controls for a monopoly firm of the Evans' type. *Econometrica*, 39(1):119–129, 1971.

W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

S. Thore. Price strategies of an "ignorant" monopolist. In H. O. A. Wold, editor, *Econometric model building. Essays on the Causal Chain Approach*. North Holland Publishing Co., Amsterdam, 1964.

J. Tinbergen. Bestimmung und deutung von angebotskurven: ein beispiel. *Zeitschrift fur Nationalokonomie*, 1(5):669–679, 1930.

G. Tintner. Monopoly over time. *Econometrica*, 5(2):160–170, 1937.

D. Trefler. The ignorant monopolist: optimal learning with endogenous information. *International Economic Review*, 34(3):565–581, 1993.

G. Tzavelas. A note on the uniqueness of the quasi-likelihood estimator. *Statistics & Probability Letters*, 38(2):125–130, 1998.

I. Venezia. Optimal investments in market research. *European Journal of Operational Research*, 18 (2):198–207, 1984.

J. Vermorel and M. Mohri. Multi-armed bandit algorithms and empirical evaluation. In *Proceedings of the European Conference of Machine Learning, Porto, Portugal (October 3-7)*, volume 3720 of *Springer Lecture Notes in Computer Science*, pages 437–448, 2005.

V. G. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, COLT '90, pages 371–386, San Francisco, 1990. Morgan Kaufmann Publishers.

Z. Wang, S. Deng, and Y. Ye. Close the gaps: a learning-while-doing algorithm for a class of single-product revenue management problems. Working paper, 2011.

L. R. Weatherford and S. E. Kimes. A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19(3):401–415, 2003.

R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61(3):439–447, 1974.

A. R. Wildt and R. S. Winer. Modeling and estimation in changing market environments. *The Journal of Business*, 56(3):365–388, 1983.

U. Witt. How can complex economical behavior be investigated? The example of the ignorant monopolist revisited. *Behavioral Science*, 31(3):173–188, 1986.

P. G. Wright. *The tariff on animal and vegetable oils*. The Macmillan Company, New York, 1928.

C. H. Xia and P. Dube. Dynamic pricing in e-services under demand uncertainty. *Production and Operations Management*, 16(6):701–712, 2007.

C. A. Yano and S. M. Gilbert. Coordinated pricing and production/procurement decisions: a review managing business interfaces. In A. K. Chakravarty and J. Eliashberg, editors, *Managing Business Interfaces*, volume 16 of *International Series in Quantitative Marketing*, chapter 3, pages 65–103. Springer, New York, 2005.

C. Yin, H. Zhang, and L. Zhao. Rate of strong consistency of maximum quasi-likelihood estimator in multivariate generalized linear models. *Communications in Statistics - Theory and Methods*, 37 (19):3115–3123, 2008.

J. Y. Yu and S. Mannor. Unimodal bandits. In L. Getoor and T. Scheffer, editors, *Proceedings of the 28th International Conference on Machine Learning, Bellevue, Washington, USA*, pages 41–48, 2011.

E. Zabel. Monopoly and uncertainty. *The Review of Economic Studies*, 37(2):205–219, 1970.

S. Zhang and Y. Liao. On some problems of weak consistency of quasi-maximum likelihood estimates in generalized linear models. *Science in China Series A: Mathematics*, 51(7):1287–1296, 2008.

W. Zhao and Y. S. Zheng. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Science*, 46(3):375–388, 2000.