# Pricing and learning with uncertain demand

Miguel Sousa Lobo[1] Stephen Boyd[2]

November, 2003

[1] mlobo@duke.edu
[2] boyd@stanford.edu

**Abstract**

Practical policies for the monopolistic pricing problem with uncertain demand are discussed (for discrete time, continuous prices and demand, in a linear and Gaussian setting). With this model, the introduction of price variations is rationally justified, to allow for a better estimate of the elasticity of demand, and increased profits due to better pricing. An approximation of the dynamic programming solution is introduced, exploiting convex optimization methods for computational tractability. Numerical experiments are described.

- First draft April 1999 (talk at INFORMS Cincinnati, May 1999), second draft June 2003 (talk at INFORMS Revenue Management Conference, Columbia University, June 2003).

# 1 Introduction

This paper addresses the problem of monopolistic pricing over multiple time periods. It is assumed that the firm is a price setter, and can set prices independently for each period. The demand follows an imprecisely known curve, which is first assumed constant from period to period.

The overall goal of pricing is to maximize the discounted profit, over a finite sequence of pricing periods. The selection of a profitable price requires knowledge of the demand curve. In our setting, this knowledge can only be obtained by observation of the demand at different prices, in different time periods. Hence, pricing at each period has two goals: 1) to maximize profit, and 2) to obtain information about the demand curve in order to increase future profits. These goals are usually in conflict, requiring a tradeoff between the two.

A theoretical implication of this work is that price variations observed in a market can, in part, be explained by rational learning behavior by firms. Our main concern, however, is with practical application, and with computing a high-quality approximate solution to the optimal pricing policy.

The optimal pricing policy is given by the solution to a stochastic dynamic program. Computing the exact solution of this dynamic program is, however, a numerically intractable problem for more than very small number of periods. Instead, we propose a convex approximation of the true cost function, along the lines of Lobo and Boyd [LB99]. An approximate solution to the optimal pricing problem can then be obtained by solving a convex

optimization program. The practicability of this approach relies on new methods for convex programming, in particular the very efficient interior-point methods for semidefinite programming developed in recent years.

The model we consider, although applicable as is to a number of practical problems, is a simple one. This allows us to develop the basic framework and to investigate the key ideas, but also gives rise to a number of limitations. We assume a linear (or linearized) demand function, and all random variables to have a Gaussian distribution. The accuracy of this model and the tightness of the approximate solution we propose depend on some degree of locality in the solution. That is, the optimal solution should not involve large price variations. This will be the case if the uncertainty about the demand is not very large. Finally, this approach requires a prior distribution on the demand parameters that, in practice, may be difficult to provide. Nevertheless, our model displays all the relevant properties of a problem where the goals of estimation and of optimization must be balanced, and provides a number of valuable insights.

This problem is also of imminent practical concern. The lack of knowledge about model parameters is mentioned by revenue management practicioners as a major obstacle to the adoption of theoretical models (including dynamic pricing models which are not dynamic in this sense). We hope to contribute some insight into the trade-off between collecting information and expoiting it, in the context of pricing problems.

This framework can be extended for a wider range of problems with

practical application. One step in this direction is given by the extension that allows for a demand that changes stochastically over time, with the consequence that information 'ages' over time. Another important step for practical application is to consider similar approximation procedures for other probability distributions, and for nonlinear demand curves. The first step in this direction is to consider a multiplicative demand function with log-normal distributions.

Within this framework, it is a straightforward extension to deal with multiple products, where the substitution or complementarity effects are also uncertain and need to be estimated (*i.e.*, the elasticity matrix). Another possible extension is to consider a competitive environment which results in a joint optimization problem in the prices of each firm. This framework can only handle a situation in which the products sold by each firm are not perfect substitutes. The substitution effect can then be modeled as linear in the price difference between the competing sellers.

There are many connections between the problem addressed here and the dual-control literature. The approximation we describe is inspired by Lobo and Boyd [LB99]. Much of the early work on dual-control was done by Fel'dbaum [Fel65]. Bar-Shalom [BS81] has a body a of work dealing with its practical application and with the properties of different policies.

The literature on dynamic pricing and the economics of uncertainty is substantial. Alchian [Alc50] discusses optimal learning by experimentation with a broad model formulation. A number of authors have looked at the

infinite-horizon problem with a fixed, deterministic demand function. Rothschild [Rot74], with a two-armed bandit model, approaches the problem motivated by the fundamental question of "how a perfectly competitive firm discovers what the market price is," and how this discovery process may explain price variability observed in the market.

Aghion et al. [ABHJ91] using a very general mathematical framework, derive mostly results for deterministic payoff function, and a limited characterization of experimentation strategies. Easley and Kiefer [EK88], operating in a stochastic framework with infinite horizon, focus on convergence of policies and limit beliefs which, it is shown, are not necessarily certainty. Kiefer [Kie89] performs the numerical computation of the value function of the dynamic program, for a model with finite parameter set. Rustichini and Wolinsky [RW95] develop a model with discrete action space, with a unit demand with reservation price. Keller and Rady [KR99], work in continuous space and continuous time with differential equations, and with a randomly changing reward function. Depending on the model parameters, they identify two experimentation regimes, extreme and moderate. Balvers and Cosimano [BC90] work with a model with continuous parameters similar to the one discussed here, and also include a discussion of myopic and certainty equivalent price. Mirman et al. [MSU93] investigate monopoly price and quantity experimentation for a 2-period problem. They derive conditions for when is it optimal for the firm to experiment. Kiefer and Nyarko 89 [KN89], investigate properties of optimal control policies for a

linear model with infinite horizon. Wieland [Wie00] investigates the value of optimal experimentaion in a framework that addresses regression in general.

Revenue management has been a very active area in recent years (see Gallego and van Ryzin [GvR93, GvR97]), and with some recent attention to dynamic pricing with active learning (see Aviv and Pazcal [AP02]). There is also a considerable literature on demand models and elasticity estimation (for a review in the context of the recent revenue management literature, see Caldentey and Bitran [CB02]).

Our approach is made feasible by relatively recent advances in nonlinear convex optimization, specifically interior-point methods. While interior-point methods have been discussed for at least thirty years, the current development was launched in 1984 by Karmarkar [Kar84] with an algorithm for linear programming that was more efficient than the simplex method. More recently, Nesterov and Nemirovsky observed that interior-point methods for linear programming can be extended to handle a very wide variety of *nonlinear* convex optimization problems [NN94]. Current algorithms can solve problems with hundreds of variables and constraints in times measured in seconds or at most a few minutes, on a personal computer. Far larger problems can be handled if problem structure, such as sparsity, is exploited. A large body of literature now exists on interior-point methods for linear programming, and a number of books have been written (see, for instance, [BV03] for extensive references). Some specific types of non-

linear convex optimization problems have recently been the focus of much research, both in terms of algorithms and applications. These include semidefinite programming (SDP) [VB96] and second-order cone programming (SOCP) [LVBL98]. The numerical examples in this paper were solved using the optimization software SP, by Vandenberge and Boyd [VB94]. Other software packages that handle SDP problems are now available, such as SDPPACK by Alizadeh *et al.* [AHN+97], and SEDUMI by Sturm [Stu98]. These recent advances make feasible in practice policies based on online optimization, such as the one developed here.

## 2 Demand model

We assume that the demand $q_t$ for time period $t$, is a function of the price $p_t$ established for that period and of a random independent perturbation. We consider a linear model, with an additive perturbation:

$$q_t(p_t) = g - h\, p_t + e_t,$$

where

- $q_t$ is the demand for period $t$,
- $p_t$ is the price for period $t$,
- $g, h$ are the demand function coefficients (intersect and elasticity),
- $e_t$ is the random perturbation.

The perturbations are assumed Gaussian i.i.d.,

$$e_t \sim \mathcal{N}(0, \sigma_e^2).$$

Although this model is chosen mainly for mathematical convenience, it is not far removed from applicability. We discuss extensions and modifications of this model in later sections.

The n-product problem is obtained with $p_t, q_t, g \in \mathbf{R}^n$, and $h \in \mathbf{R}^{n\times n}$. The treatment and results are analogous to the one-product case. For notational simplicity, we restrict ourselves in this paper to the one-product problem, except for a numerical example in §13.

## 3 Demand function coefficients

The uncertain knowledge of the demand function coefficients is handled in a Bayesian framework. The coefficients are assumed to be random variables with a known *a priori* Gaussian distribution. The distribution for the vector formed by the demand function coefficients is then

$$\begin{bmatrix} g \\ h \end{bmatrix} = \mathcal{N}\Big( \begin{bmatrix} \hat{g}_0 \\ \hat{h}_0 \end{bmatrix}, \Pi_0^{-1} \Big).$$

The *a priori information matrix* $\Pi_0$ is the inverse of the covariance matrix of the vector of coefficients.

## 4 Problem objective

The problem objective is to maximize the expected discounted profit

$$\mathbf{E}\Big(\sum_{t=1}^{T} \delta^{t-1}\, R_t(p_t)\Big) \;=\; \mathbf{E}\Big(\sum_{t=1}^{T} \delta^{t-1}\, (p_t - c)\, q_t(p_t)\Big),$$

where $c$ is the (variable) costs, and $\delta$ is the discount factor. Note that this formulation implies the assumption that there is always sufficient stock to meet demand, and that the stock holding costs are negligible.

The prices must be selected at each period using only information available from the demand observed in previous periods, and the *a priori* knowledge of the distributions. Hence, the maximization is over the set of *feasible policies*, where a feasible policy is one in which the prices are functions of the form

$$p_t = \varphi(\hat{g}_0, \hat{h}_0, \Pi_0, \sigma_e^2, q_1, \ldots, q_{t-1}),$$

that is, $p_t$ is a random variable measurable $\sigma(q_1, \ldots, q_{t-1})$. We will also consider *randomized feasible policies*, where $p_t$ is measurable $\sigma(q_1, \ldots, q_{t-1}, w)$, with $w$ some independent random variable introduced to allow for the randomization of the $p_t$.

## 5 Tails

Note that this model makes two unusual assumptions. The first is that, since the normal distribution does not have bounded support, the slope of the demand function may be positive. This might lead to the conclusion that the optimal price is infinitely large. If there is a positive probability of demand increasing with price, we might conclude that an infinitely large price results in an infinitely large expected profit.

The second assumption of note is that linearity extends into negative demand. That is, if the price is too high, the seller may be forced to buy

items back at a loss. In terms of the solution, this ensures that very large prices have very large negative expected profit (assuming $\hat{h} > 0$). The optimal solution is then determined by the distribution in the region of interest of 'reasonable' parameter values, rather than by the tail of the distribution.

The simulations described here were run with a truncated tail, to disallow non-negative slopes (draws in the Monte Carlo simulation with non-negative slope are thrown out, otherwise the active learning algorithm will usually quickly find out that this is the case, and the price diverges to arbitrarily large values). This can be interpreted as a form of model mismatch, where the 'true' model is the jointly normal distribution with truncated tails used in the simulation. The tail truncation is ignored in the model used to derive our approximation, which can also be thought of as another approximation step.

## 6 A posteriori distribution

The *a posteriori* distribution of the coefficients after $t$ periods is given by

$$\begin{bmatrix} g \\ h \end{bmatrix} = \mathcal{N}\Big( \begin{bmatrix} \hat{g}_t \\ \hat{h}_t \end{bmatrix}, \Pi_t^{-1} \Big),$$

with

$$\begin{aligned} \Pi_t &= \Pi_0 + \sigma_e^{-2} \sum_{k=1}^{t} \begin{bmatrix} 1 & -p_k \\ -p_k & p_k^2 \end{bmatrix}, \\ \begin{bmatrix} \hat{g}_t \\ \hat{h}_t \end{bmatrix} &= \Pi_t^{-1} \left( \Pi_0 \begin{bmatrix} \hat{g}_0 \\ \hat{h}_0 \end{bmatrix} + \sigma_e^{-2} \sum_{k=1}^{t} \begin{bmatrix} 1 \\ -p_k \end{bmatrix} q_k \right). \end{aligned}$$

Note that if price remains constant over all periods, the information matrix can be ill-conditioned. We will later see that this translates into expected loss of profit. Equivalent recursive formulas for the *a posteriori* distribution are

$$\begin{aligned}
\Pi_t &= \Pi_{t-1} + \sigma_e^{-2} \begin{bmatrix} 1 & -p_t \\ -p_t & p_t^2 \end{bmatrix}, \\
\begin{bmatrix} \hat{g}_t \\ \hat{h}_t \end{bmatrix} &= \begin{bmatrix} \hat{g}_{t-1} \\ \hat{h}_{t-1} \end{bmatrix} + \sigma_e^{-2} \Pi_t^{-1} \begin{bmatrix} 1 \\ -p_k \end{bmatrix} \left( q_t - (\hat{g}_{t-1} - \hat{h}_{t-1} p_t) \right),
\end{aligned}$$

which are easily obtained from, say, the Kalman filter equations. XXX ref? Note that since all distributions are Gaussian, $\hat{g}_t$, $\hat{h}_t$ and $\Pi_t$ provide a complete description of the distributions. They can be interpreted as a system state, with the stochastic state transition being controlled by the prices.

# 7 Static policy

If the prices are decided *a priori*, that is if $p_t$ is not a function of $q_1, \ldots, q_{t-1}$, the expected period $t$ profit is

$$\mathbf{E}\left(R_t(p_t)\right) = -\hat{g}_0 c + (\hat{g}_0 + c\hat{h}_0) p_t - \hat{h}_0 p_t^2.$$

The price that maximizes this is

$$p_1^{\mathrm{m}} = \frac{\hat{g}_0 + c\hat{h}_0}{2\hat{h}_0},$$

(for the justification of the sub- and superscript see the myopic policy next). Replacing this in the expression for $R$, we get

$$R_t(p_1^{\mathrm{m}}) = -h\left(\frac{\hat{g}_0 + c\hat{h}_0}{2\hat{h}_0}\right)^2 + (g - ch)\frac{\hat{g}_0 + c\hat{h}_0}{2\hat{h}_0} - cg - ce_t,$$

the expected value of which is

$$\mathbf{E}\left(R_t(p_0^{\mathrm{m}})\right) = \frac{1}{4\hat{h}_0}\left(\hat{g}_0 + c\hat{h}_0\right)^2 - c\hat{g}_0.$$

We call this a static (or naive, or dumb) policy in the sense that it does not make use of new information about the demand function that becomes available through the observation of $q_t$, so that the prices for all time periods are determined *a priori*, before the first period. Note that in the full information case the static policy is optimal (*i.e.*, if $\hat{g}_0 = g$ and $\hat{h}_0 = h$, which can also be stated as $\Pi_0^{-1} = 0$).

## 8 Myopic policy

The expected period $t$ profit, given the information available up to $t-1$, is

$$\mathbf{E}\left(R_t(p_t) \,|\, t-1\right) = -\hat{g}_{t-1}c + (\hat{g}_{t-1} + c\hat{h}_{t-1})p_t - \hat{h}_{t-1}p_t^2.$$

The price that maximizes this is

$$p_t^{\mathrm{m}} = \frac{\hat{g}_{t-1} + c\hat{h}_{t-1}}{2\hat{h}_{t-1}}$$

where the superscript stands for *myopic*. In fact, among all feasible policies, this is the one that, at each period, maximizes the immediate expected profit. No attention is paid to the effect of prices on the *a posteriori* distribution, nor

to the consequent effects on the expected profits in future periods. Another way to describe this is to note that, while learning occurs, there is no design for learning. That is, no effort is made to select prices that are informative about the demand function coefficients. Using $p_t^{\mathrm{m}}$ in the expression for $R$, we get

$$R_t(p_t^{\mathrm{m}}) = -h\left(\frac{\hat{g}_{t-1} + c\hat{h}_{t-1}}{2\hat{h}_{t-1}}\right)^2 + (g - ch)\frac{\hat{g}_{t-1} + c\hat{h}_{t-1}}{2\hat{h}_{t-1}} - cg - ce_t,$$

the expected value of which is

$$\mathbf{E}\,(\,R_t(p_t^{\mathrm{m}})\,|\,t-1) = \frac{1}{4\hat{h}_{t-1}}\left(\hat{g}_{t-1} + c\hat{h}_{t-1}\right)^2 - c\hat{g}_{t-1}.$$

In the full information case the myopic policy is optimal, and equal to the static policy.

# 9 Myopic policy with dithering

Consider now a simple modification of the myopic policy, which consists in adding a random perturbation to the price. The motivation behind the use of such a randomized policy is that price variations will "excite" the learning process. More specifically, they will make the information matrix well-conditioned. In fact, as we will see later, it is desirable that the information be large in the sense of having a large minimum eigenvalue.

Consider a simple numerical example. The number of products is $n = 10$, and the number of pricing periods is $T = 20$. The expected revenue for a range of dithering amplitudes was estimated by Monte Carlo simulation. Figure 1 shows the results, with the $\log_{10}$ of the dithering amplitude on the

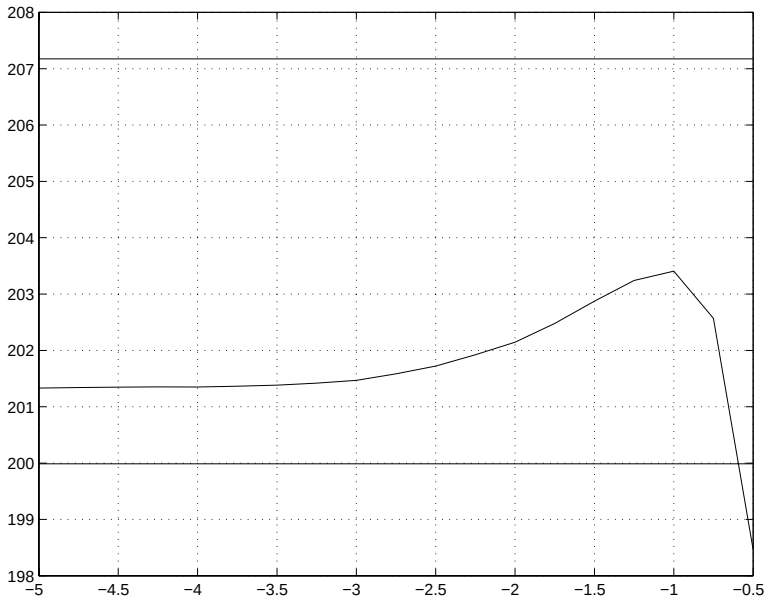


**Figure 1:** Myopic policy with dithering, profit as a function of the log of the dithering level.

horizontal axis. The vertical axis shows the expected profit (a large number of trials was run and error bars are omitted on account of their very small height).

The upper horizontal line represents the profit that could be expected if full information about the demand was available (that is, if $g$ and $h$ were known). The bottom horizontal line represents the expected profit for the static policy. The curve shows the expected profit for the myopic policy with dithering. Note that the leftmost end of the curve has close to no dithering at all, and therefore shows the results for the myopic policy without dithering.

As the dithering level is increased, the learning is improved, leading to higher profits. Note the apparently counter-intuitive result: profits are increased by adding a random, independent term to the prices. If too much dithering is added, a very accurate knowledge of demand can be obtained, but at the cost of very unprofitable prices. The critical problem lies in determining what the best level of dithering is, which may be difficult to do *a priori*.

The best dithering level, and the increase in profit that can be achieved by the introduction of dithering, depend in nontrivial ways on the problem parameters (the prior distributions and the demand noise level). In any case, the general form of the curve is as shown. The conclusion then, is that the introduction of price variations, in this framework, is rationally justified (assuming, of course, that the information processing ability exists in order to learn from such variations).

Mirman et al. [MSU93] derive conditions for when it is justified to experiment in a 2-period setting (in our next example of Figure 2, this corresponds to determining when is the maximum not achieved at the myopic price of 0.5).

We next turn to the problem of computing more efficient, nonrandom perturbations.

## 10 A two period example

Consider an example with one product and two time periods. The optimal price for the second (and last) period is, of course, the myopic price. With

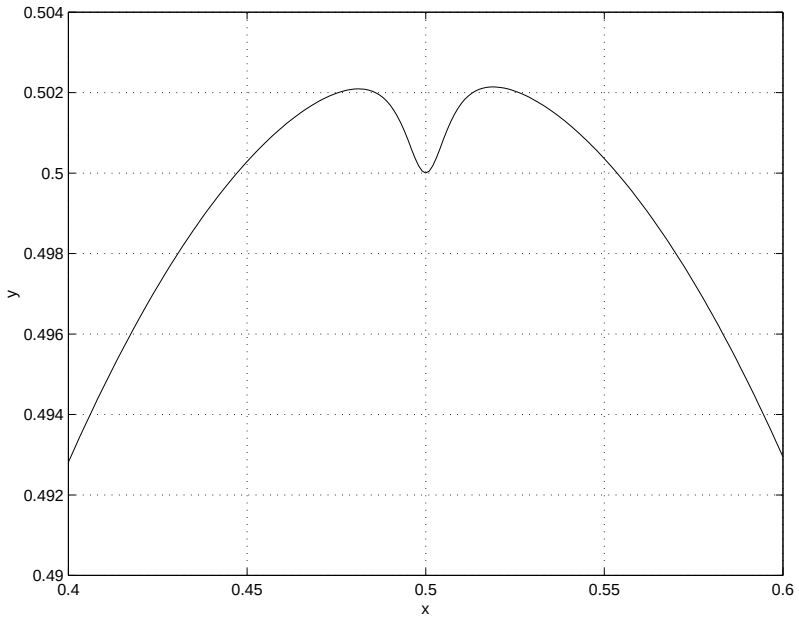


**Figure 2:** Two period example, expected profit as function of first-period price.

the second-period price selected in this manner, Figure 2 plots the total expected revenue as a function of the first-period price.

In this example, the myopic first-period price is 0.5. It is clearly seen that the myopic price is not optimal for the first period. In fact, a small deviation in either direction generates information about the demand slope, without a significant reduction in the first period profit. This extra information pays off, on average, allowing for more accurate pricing in the second-period, and increased profit.

The shape of the curve depends on the available *a priori* information. In this example, we assume that accurate information exists about the value of the demand at the myopic price level, but that the slope of the demand

function around this point is more uncertain. This is a likely scenario if the information about the demand function has been obtained from a number of previous pricing periods where the price was kept (nearly) constant. The curve in Figure 2 was obtained by Monte Carlo simulation (error bars are omitted on account of their very small height).

## 11 Dynamic program

The exact solution to the optimal pricing problem is given by the following dynamic program, of which we omit the derivation:

$$V_t\,(p_1,\ldots,p_{t-1}) = \inf_{p_t} \mathbf{E}(\ (p_t - c)\,(g - hp_t + e_t) + \delta\, V_{t+1}\,(p_1,\ldots,p_t) \mid t-1\,),$$

for $t = 1,\ldots,T$, and $V_{T+1} = 0$. The expectation conditioned on $t-1$ denotes conditioning on $q_1,\ldots,q_{t-1}$ (recall that $p_t$ is restricted to be measurable $\sigma(q_1,\ldots,q_{t-1})$). Note that, in general, $V_t$ is a very complicated function of $p_1,\ldots,p_{t-1}$. In fact, the conditional expectation depends on the *a posteriori* distribution of $g$ and $h$, which depends on all the previous prices. The optimal objective is $V_1$, and the optimal prices are given by

$$p_t^* = \operatorname*{arginf}_{p_t} \mathbf{E}(\ (p_t - c)\,(g - hp_t + e_t) + \delta\, V_{t+1}\,(p_1,\ldots,p_t) \mid t-1\,).$$

The general procedure for the derivation of the dynamic program for a similar problem is described in more detail in Lobo and Boyd [LB99], which also discusses in more detail how a convex approximation of the type presented next relates to the dynamic program.

While this problem can be solved by traditional stochastic programming methods, for more than very few time periods (or for multiple products,) such an approach quiclkly becomes numerically intractable. Several approximate dynamic programming procedures are available. We will however derive a specific approximation for this problem. This is in part motivated by the desire better to understand the structure of the problem. Further, fast and effective methods will be required if we wish to be able to eventually solve large practical problem,

## 12 Convex approximation

We now propose a computationally tractable convex approximation of the dynamic program. Without loss of generality, we simplify notation by assuming the current time period to be $t = 1$, and the horizon to be $T$. That is, we only seek to approximate $p_1^*$. Note that, for Gaussian distributions, $\hat{g}_1$, $\hat{h}_1$, and $\Pi_1$ are sufficient statistics for the demand function coefficients. They can be interpreted as the state of a Markov system, so that at time period $t = 2$ we do not need to know $p_1$ nor $q_1$ to determine the optimal price $p_2^*$. Therefore, the problem at time $t = 2$ can be solved by changing the time index to $t' = t - 1$ (so that the current time is $t' = 1$), by reducing the horizon to $T' = T - 1$, and by using as the *a priori* distribution $\hat{g}'_0 = \hat{g}_1$, $\hat{h}'_0 = \hat{h}_1$, and $\Pi'_0 = \Pi_1$.

We consider an alternative, more convenient formulation of the problem. Instead of maximizing the expected profit, we minimize the expected profit

loss relative to the full information case. That is, we wish to minimize

$$\mathbf{E}\Big(\sum_{t=1}^{T}\delta^{t-1}\left(R_t(p^{\mathrm{f}})-R_t(p_t)\right)\Big),$$

where the full-information optimal price is

$$p^{\mathrm{f}}=\frac{g+ch}{2h},$$

and, as before, the minimization is over the set of feasible policies. Note that $\mathbf{E}\left(R_t\left(p^{\mathrm{f}}\right)\right)$ is a constant. Using the expected profit under full information in this manner as a reference point provides for simpler and more elegant formulas.

Now, divide the profit loss into two terms:

$$R_t(p^{\mathrm{f}})-R_t(p_t)=\underbrace{\left(R_t(p^{\mathrm{f}})-R_t(p_t^{\mathrm{m}})\right)}_{1}+\underbrace{(R_t(p_t^{\mathrm{m}})-R_t(p_t))}_{2},$$

where $p_t^{\mathrm{m}}$ is the myopic price, as described in §8. We consider each of these terms in turn.

The first term is the difference in period profit, between full information and partial information under myopic pricing. It is a lower bound on the period profit loss due to incomplete information. With the appropriate substitutions, we find this term to be:

$$\begin{aligned}
&R_t(p^{\mathrm{f}})-R_t(p_t^{\mathrm{m}})\ =\\
&=\ \left(\frac{1}{4h}(g+ch)^2-cg-ce_t\right)-\left(-h\left(\frac{\hat{g}_{t-1}+c\hat{h}_{t-1}}{2\hat{h}_{t-1}}\right)^2+(g-ch)\frac{\hat{g}_{t-1}+c\hat{h}_{t-1}}{2\hat{h}_{t-1}}-cg-ce_t\right)\\
&=\ \frac{h}{4}\left(\frac{g+ch}{h}-\frac{\hat{g}_{t-1}+c\hat{h}_{t-1}}{\hat{h}_{t-1}}\right)^2\\
&=\ \frac{h}{4}\left(\frac{g}{h}-\frac{\hat{g}_{t-1}}{\hat{h}_{t-1}}\right)^2.
\end{aligned}$$

Expanding $(g/h - \hat{g}_{t-1}/\hat{h}_{t-1})^2$ in Taylor series to second order around $\hat{g}_{t-1}$ and $\hat{h}_{t-1}$, we get:

$$R_t(p^{\mathrm{f}}) - R_t(p_t^{\mathrm{m}}) \;\approx\; \frac{h}{4}\left(\frac{1}{\hat{h}_{t-1}^2}(g-\hat{g}_{t-1})^2 + \frac{\hat{g}_{t-1}^2}{\hat{h}_{t-1}^4}(h-\hat{h}_{t-1})^2 - 2\frac{\hat{g}_{t-1}}{\hat{h}_{t-1}^3}(g-\hat{g}_{t-1})(h-\hat{h}_{t-1})\right).$$

And it is easily seen that the conditional expectation of this expression, given the information available up to period $t-1$, is equal to:

$$\mathbf{E}\left(R_t(p^{\mathrm{f}}) - R_t(p_t^{\mathrm{m}})\middle| t-1\right) \;\approx\; \frac{1}{4\hat{h}_{t-1}}\begin{bmatrix} 1 & -\hat{g}_{t-1}/\hat{h}_{t-1} \end{bmatrix} \Pi_{t-1}^{-1} \begin{bmatrix} 1 \\ -\hat{g}_{t-1}/\hat{h}_{t-1} \end{bmatrix}.$$

We may call this term the *cost of ignorance*. Note that a large minimum eigenvalue of $\Pi_{t-1}$ guarantees this term to be small. While the expression may seem complicated, once we linearize $\Pi_{t-1}$ in the prices, it becomes convex. The information matrix $\Pi_t$ is approximated by a matrix $P_t$ linear in the prices, which is given by

$$P_t = P_{t-1} + \sigma_e^{-2}\begin{bmatrix} 1 & -p_t \\ -p_t & 2p_t^{\mathrm{r}}p_t - (p_t^{\mathrm{r}})^2 \end{bmatrix}$$

for $t = 1, \ldots, T$, and $P_0 = \Pi_0$. The $p_t^{\mathrm{r}}$ are a sequence of reference prices, around which the linearization is performed. If only relatively small price deviations are introduced for the purpose of improving the learning process, the approximation can be expected to be accurate.

With this procedure, the cost of ignorance is approximated by a convex matrix-fractional expression. Naturally, the question arises as to why should the information matrix be linearized, and not, say, the whole expression. Or why not use a second order expansion of the whole expression?

In practice, the approximation given here is observed to be accurate over a much larger set. This can be understood by noting that its structure is more closely related to the original function than is the case for the alternative approximations, but more analysis work needs to be done in this direction.

Consider now the second term. Writing $p_t = p_t^{\mathrm{m}} + p_t^{\mathrm{d}}$, its conditional expectation is

$$\begin{aligned}
\mathbf{E}\left(R_t(p_t^{\mathrm{m}}) - R_t(p_t)\middle| t-1\right) &= \mathbf{E}\left(R_t(p_t^{\mathrm{m}}) - R_t(p_t^{\mathrm{m}} + p_t^{\mathrm{d}})\middle| t-1\right) \\
&= -(\hat{g}_{t-1} + c\hat{h}_{t-1})p_t^{\mathrm{d}} + \hat{h}_{t-1}\left(2p_t^{\mathrm{m}}p_t^{\mathrm{d}} + (p_t^{\mathrm{d}})^2\right) \\
&= -(\hat{g}_{t-1} + c\hat{h}_{t-1})p_t^{\mathrm{d}} + \hat{h}_{t-1}\left(\frac{\hat{g}_{t-1} + c\hat{h}_{t-1}}{\hat{h}_{t-1}}p_t^{\mathrm{d}} + (p_t^{\mathrm{d}})^2\right) \\
&= \hat{h}_{t-1}\left(p_t^{\mathrm{d}}\right)^2 \\
&= \hat{h}_{t-1}\left(p_t - p_t^{\mathrm{m}}\right)^2.
\end{aligned}$$

This is convex quadratic in the price and can be called the *cost of experimentation*, since it penalizes the deviations from the myopic price which are introduced to improve learning.

The total expected profit loss is approximated by:

$$\mathbf{E}\left(R_t(p^{\mathrm{f}}) - R_t(p_t)\middle| t-1\right) \approx \underbrace{\frac{1}{4\hat{h}_{t-1}}\begin{bmatrix} 1 & -\hat{g}_{t-1}/\hat{h}_{t-1}\end{bmatrix} P_{t-1}^{-1}\begin{bmatrix} 1 \\ -\hat{g}_{t-1}/\hat{h}_{t-1}\end{bmatrix}}_{1} + \underbrace{\hat{h}_{t-1}\left(p_t - p_t^{\mathrm{m}}\right)^2}_{2},$$

where $P_t$ is as above (linear in $p_t$). The two terms in the expected profit loss are: 1) the cost of ignorance, and 2) the cost of experimentation. The approximation can be expected to be tight for $\Pi_{t-1}$ large, and for the $p_t$ close to the reference price sequence $p_t^{\mathrm{r}}$ and close to the myopic prices $p_t^{\mathrm{m}}$.

Now, by the tower property of conditional expectation,

$$\mathbf{E}\left(R_t(p^{\mathrm{f}}) - R_t(p_t)\right) = \mathbf{E}\left(\mathbf{E}\left(R_t(p^{\mathrm{f}}) - R_t(p_t)\middle| t-1\right)\right).$$

Note that $\hat{g}_{t-1}$, $\hat{h}_{t-1}$, and $p_t^{\mathrm{m}}$ are random variables whose distributions depend non-trivially on the prices $p_1, \ldots, p_{t-1}$. We now make a key approximation that permits the crucial simplification of the dynamic program: $\hat{g}_{t-1} = \hat{g}_0$ and $\hat{h}_{t-1} = \hat{h}_0$, (and, therefore, $p_t^{\mathrm{m}} = p_1^{\mathrm{m}}$). The informal justification for this approximation is as follows: we assume that changes in the information matrix are more important in determining the expected loss of profit in future periods than are the eventual changes in the estimates of the demand coefficients. Or, similarly: we assume that changes in the coefficient estimates will not significantly affect the future value of the information collected now. These arguments are supported by numerical experiments described later in this paper, and to some extent by analysis, and are further discussed in Lobo and Boyd [LB99].

The overal objective is then

$$\begin{aligned}
&\mathbf{E}\Big(\sum_{t=1}^{T} \delta^{t-1}\left(R_t(p^{\mathrm{f}}) - R_t(p_t)\right)\Big) \;\approx \\
&\;\approx\; \sum_{t=1}^{T} \delta^{t-1}\left(\frac{1}{4\hat{h}_0}\begin{bmatrix} 1 & -\hat{g}_0/\hat{h}_0 \end{bmatrix} P_{t-1}^{-1} \begin{bmatrix} 1 \\ -\hat{g}_0/\hat{h}_0 \end{bmatrix} + \hat{h}_0\,(p_t - p_1^{\mathrm{m}})^2\right),
\end{aligned}$$

which is to be minimized over the set of feasible policies. We may write this as a dynamic program, as in §11. However, given the last approximation, we are now able to drop the conditional expectations and group all the inf

operators:

$$p_1^{\mathrm{a}} = \operatorname*{arginf}_{p_1} \inf_{p_2} \cdots \inf_{p_T} \sum_{t=1}^{T} \delta^{t-1} \left( \frac{1}{4\hat{h}_0} \begin{bmatrix} 1 & -\hat{g}_0/\hat{h}_0 \end{bmatrix} P_{t-1}^{-1} \begin{bmatrix} 1 \\ -\hat{g}_0/\hat{h}_0 \end{bmatrix} + \hat{h}_0 \, (p_t - p_1^{\mathrm{m}})^2 \right).$$

Using Schur complements, we see that the optimal price for the approximation can be obtained from the solution of the semidefinite and quadratic program:

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T} \delta^{t-1} \, (\alpha_t + \beta_t) \\ \text{subject to} & \\ & \begin{bmatrix} 4\,\hat{h}_0\,\alpha_t & 1 & -\hat{g}_0/\hat{h}_0 \\ 1 & \Pi_0^{1,1} + \sigma_e^{-2}(t-1) & \Pi_0^{1,2} - \sigma_e^{-2} \sum_{k=1}^{t-1} p_k \\ -\hat{g}_0/\hat{h}_0 & \Pi_0^{2,1} - \sigma_e^{-2} \sum_{k=1}^{t-1} p_k & \Pi_0^{2,2} + \sigma_e^{-2} \sum_{k=1}^{t-1} \left( 2 p_k^{\mathrm{r}} p_k - (p_k^{\mathrm{r}})^2 \right) \end{bmatrix} \succeq 0, \quad t = 1, \ldots, T \\ & \hat{h}_0 \, (p_t - p_1^{\mathrm{m}})^2 \leq \beta_t, \quad t = 1, \ldots, T. \end{array}$$

The auxiliary variables $\alpha_1, \ldots, \alpha_T \in \mathbf{R}$ and $\beta_1, \ldots, \beta_T \in \mathbf{R}$ upper bound the matrix-fractional and quadratic terms, respectively. This convex program can be solved globally and efficiently.

In practice, most of the approximation error arises from the linearization of the information matrix. This can be remediated by a simple iterative procedure, as follows. After solving the program, the optimal sequence obtained is used as a new reference sequence. The linearization is then be repeated, and the program resolved. This is repeated until the optimal/reference sequence converges. In our numerical experiments, convergence always occured after a small number of iterations. The minimum found, of course, is not guaranteed to be global.

A final question concerns the selection of the initial reference sequence $p_t^{\mathrm{r}}$. The most obvious pick is the myopic price $p_1^{\mathrm{m}}$. This turns out to be a bad choice in most circumstances, since it corresponds to an unstable equilibrium point in the iterative relinearization procedure. In the examples that follow, we used the myopic price with the addition of a random perturbation. This is similar to the prices generated by a myopic policy with dithering, and was quite effective in our numerical experiments. In fact, another interpretation is that we are adjusting a dithering policy, to make it as efficient as possible, and to obtain the right balance between informativeness and immediate revenue.

## 13 Simulation results

We consider the pricing of one product, over $T = 10$ periods. As in the previous examples, the *a priori* distribution was selected such that accurate information exists about the value of the demand at the myopic price level, and such that the slope of the demand function around this point is more uncertain.

Figure 3, compares the expected profit under different policies, obtained by Monte Carlo simulation, with the corresponding error range. Bar $S$ is the result for the static policy, and bar $M$ for the myopic policy with learning. Bars $D_1$ to $D_7$ are the results for the dithering policy, with different dithering levels (starting at 0.01, with factor increments of $10^{1/4}$). Note that while dithering can significantly improve results, an excessive amount of dithering

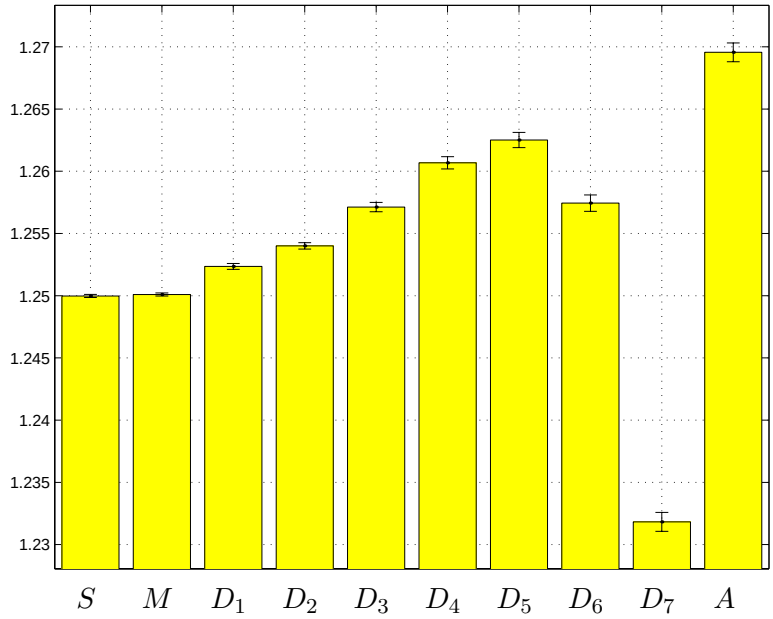


**Figure 3:** Comparison of different policies by Monte Carlo simulation, expected revenue.

quickly becomes very costly. Bar $A$ is the result for the approximation of the dynamic program. (The expected profit under full informattion is approximately 1.285).

The amount of gain to be had from using a policy more sophisticated than the myopic policy depends in a nontrivial way on the problem data: the *a priori* distribution of the demand function coefficients, the variance $\sigma_e^2$ of the random demand perturbations, the horizon $T$, and the discount factor $\delta$.

Figure 4 shows an example of the optimal price sequence obtained by the approximation procedure described in §12. The horizon was $T = 20$, and the vertical axis is the deviation from the myopic price. Note that, except

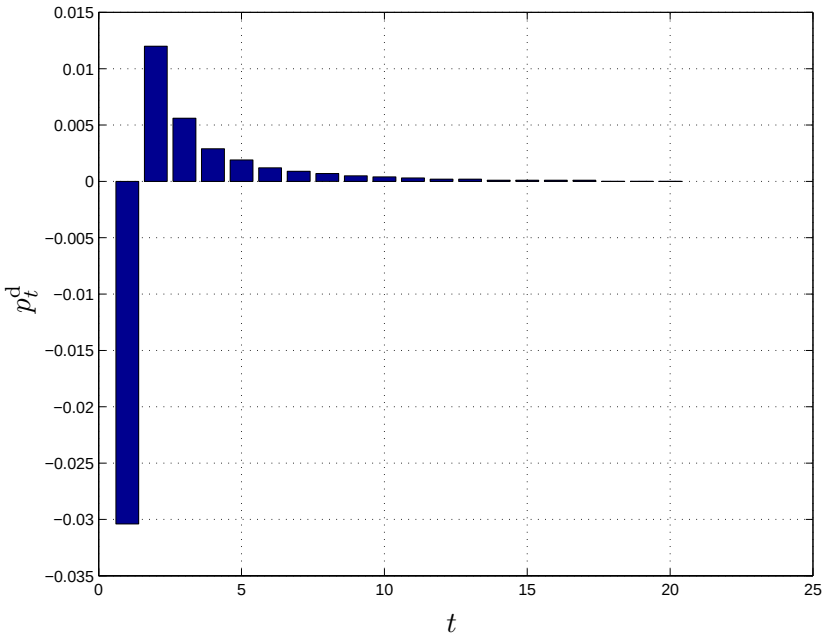


**Figure 4:** Optimal reference price sequence at $t = 1$ (deviation from myopic price).

for the first price, these are not the actual prices set in each period, but only the reference sequence obtained by the optimization procedure in the first period. The prices in future periods (as well as the reference sequence) will be adjusted to account for the information obtained from the demand observed in each period.

Figure 5, shows the first-period price deviation from the myopic price as a function of $\sigma_e^2$ (the variance of the "demand noise"). This curve has a quite intuitive explanation. For a large noise, the optimal deviation is zero because learning is then very difficult. Any significant learning could only be achieved at the cost of very high deviations, and the consequent immediate loss of profit would not be warranted. Below a certain noise level, it starts

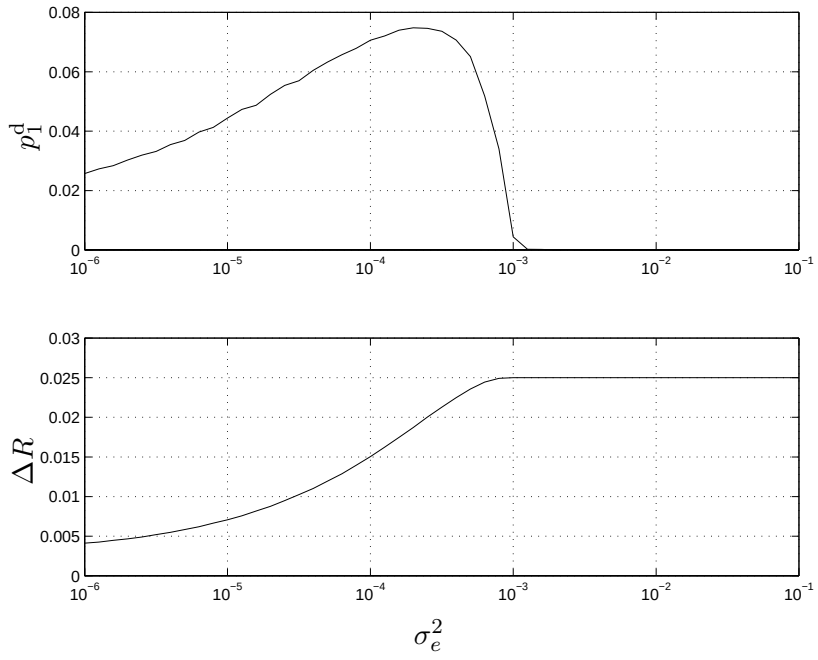


**Figure 5:** Effect of demand noise variance $\sigma_e^2$ (results from convex approximation). Top: Optimal first-period price, deviation from myopic price. Bottom: Expected profit loss, relative to pricing under full information.

to pay off to introduce price deviations, to improve the knowledge of the demand and increase future profits. For medium noise, large deviations in the price are needed to obtain significant information, while with small noise, smaller deviations suffice. Of course, this analysis assumes other factors held constant.

## 14 Extensions and conclusion

Several extensions merit consideration:

- *Competition.* The model can be extended to an oligopoly setting, with optimization problem solved jointly for $n$ firms. Such a demand model can account for a linear substitution effect, but cannot handle perfect substitutes.

- *The $n$-product problem.* As mentioned before, the $n$-product problem is a trivial extension, with the model accounting for substitution and complementarity effects. The price elasticity of demand is now an $n$-by-$n$ matrix, which can be assumed positive definite. If this matrix is not diagonal dominant (*i.e.*, if there are significant complementary and substitution effects), the expected percentual revenue loss due to innacurate estimation of the elasticity can be shown to be much greater than for the 1-product case. That is, the scope for gain by using policies with active learning greatly increases when a larger number of products is considered jointly.

- *Time-varying demand function.* The model can be extended by letting the demand function coefficients change in a random walk-like fashion (or any other linear stochastic dynamics). Change in slope can be interpreted as a change in preferences, and a change in the intersect as a change in market size. (Competition is another source for such "external shocks"; questions of strategic interaction, however, are not

considered in this framework.) The resulting expression for the information matrix $\Pi_t$ can be linearized as before, and the same policies considered. As expected, a stochastic drift reduces the future value of information obtained now. Hence, the effect of a randomly changing demand on the optimal policy is similar to the effect of reducing the horizon considered in the static demand case (*i.e.*, the number of pricing periods).

- *Non-linear demand models.* A commonly used model is that of a multiplicative demand function, where the exponents are the estimated parameters. If we consider multiplicative log-normal perturbations, we can work with the logarithms of price and demand to obtain the analogous policies for such models. A significant literature on aggregate demand models exists that can be drawn upon.

- *Marketing and quality variables.* Marketing and quality variables can be added to the model (assuming, as for the prices, that they are observable).

Price variations observed in a market can, in part, be explained by rational learning behavior by firms. We have explored some properties of the optimal learning behavior with a simple model. A goal of this work has been to develop approximate solutions to the optimal pricing policy that can be computed in reasonable time by exploiting convex optimization methods. The availability of efficient optimization algorithms for large-scale non-linear convex problems makes practicable a class of policies based on

online optimization. The framework used to develop these approximations can be extended for a wider range of problems with practical application.

## References


[ABHJ91] Philippe Aghion, Patrick Bolton, Christopher Harris, and Bruno Jullien. Optimal learning by experimentation. *Review of Economic Studies*, 58(4):621–54, July 1991.

[AHN+97] F. Alizadeh, J. P. Haeberly, M. V. Nayakkankuppam, M. L. Overton, and S. Schmieta. SDPPACK *User's Guide, Version 0.9 Beta.* NYU, June 1997.

[Alc50] Armen Alchian. Uncertainty, evolution and economic theory. *Journal of Political Economy*, 58:211–221, 1950.

[AP02] Y. Aviv and A. Pazcal. Pricing of short life-cycle products through active learning. *Under revision for Management Science*, 2002.

[BC90] Ronald J. Balvers and Thomas F. Cosimano. Actively learning about demand and the dynamics of price adjustment. *The Economic Journal*, 100(402):882–898, September 1990.

[BS81] Y. Bar-Shalom. Stochastic dynamic programming: Caution and probing. *IEEE Trans. Aut. Control*, AC-26(5):1184–1195, 1981.

[BV03] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.

[CB02] R. A. Caldentey and G. Bitran. An overview of models for revenue management. *Manufacturing and Service Operations Management*, 5:203–229, 2002.

[EK88] D. Easley and N. M. Kiefer. Controlling a stochastic-process with unknown-parameters. *Econometrica*, 56:1045–1064, 1988.

[Fel65] A. A. Fel'dbaum. *Optimal Control Systems*. Academic Press, New York, 1965.

[GvR93] G. Gallego and G van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40:999–1020, 1993.

[GvR97] G. Gallego and G van Ryzin. A multiple product dynamic pricing problem with applications to network yield management. *Operations Research*, 45:24–41, 1997.

[Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[Kie89] N. M. Kiefer. A value function arising in the economics of information. *J. Econ. Dyn. Control*, 13:201–223, 1989.

[KN89] N. M. Kiefer and Y. Nyarko. Optimal-control of an unknown linear process with learning. *Int. Econ. Rev.*, 30:571–586, 1989.

[KR99] Godfrey Keller and Sven Rady. Optimal experimentation in a changing environment. *Review of Economic Studies*, 66(3):475–507, July 1999.

[LB99] M. S. Lobo and S. Boyd. Policies for simultaneous estimation and optimization. In *Proc. American Control Conf.*, volume 2, pages 958–964, June 1999.

[LVBL98] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Applications*, 284(1-3):193–228, November 1998.

[MSU93] Leonard J. Mirman, Larry Samuelson, and Amparo Urbano. Monopoly experimentation. *International Economic Review*, 34(3):549–63, August 1993.

[NN94] Yu. Nesterov and A. Nemirovsky. *Interior-point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[Rot74] Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9:185–202, 1974.

[RW95] Aldo Rustichini and Asher Wolinsky. Learning about variable demand in the long run. *Journal of Economic Dynamics and Control*, 19:1283–92, 1995.

[Stu98] J. Sturm. SeDuMi *Version 1.02*. McMaster University, 1998.

[VB94] L. Vandenberghe and S. Boyd. SP*: Software for Semidefinite Programming. User's Guide, Beta Version.* Stanford University, October 1994. Available at `www.stanford.edu/~boyd`.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[Wie00] Volker Wieland. Learning by doing and the value of optimal experimentation. *Journal of Economic Dynamics and Control*, 24(4):501–34, April 2000.