Pricing and learning with uncertain demand

Miguel Sousa Lobo
Stephen Boyd

June 2003

Duke University Fuqua School of Business
Stanford University Information Systems Lab

Pricing and learning with uncertain demand

Outline:

1. model, problem statement
2. dumb policy (open-loop), myopic policy (closed-loop, passive learning)
3. dithering, dynamic programming, approximation
4. numerical examples
5. extensions, conclusions

Introduction

- the selection of a profitable price requires accurate knowledge of the demand function
- knowledge (beyond prior) can only be obtained by observation of the demand at different prices, in different time periods
- pricing at each period has two goals:
  - to maximize immediate profit
  - to obtain information about the demand fct. (increase future profits)

  these goals are usually in conflict, requiring a tradeoff between the two
- simple model, maximize discounted profit, over a finite sequence of pricing periods, computational issues, qualitative results

Previous, related work

- economics of uncertainty, learning by doing, value of experimentation, *etc*
  - Alchian 50, Rothschild 74, Mirman et al. 93, *etc.*

  "Since price variability is such a pervasive phenomenon, it seems unsatisfactory to regard it as simply an artifact of disequilibrium." Rothschild 74
- bandit problems
- control
  - predictive control, adaptive control, large literature
  - dual-control, Fel'dbaum 60, 65, Bar-Shalom 69, 89
  - persistency of excitation, Genceli and Nikolaou 96, Cooley and Lee 97, *etc.*
  - quadratic control, optimal adaptive robust control, Lobo and Boyd 99
- convex optimization (SDP, Vandenberghe and Boyd 96, SOCP, Lobo et al. 98, *etc.*)
- approximate dynamic programming
- clinical trials, dose-discovery (Berry et al.)
- demand models

Model, a priori distribution

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

p_t	is the price for period t,
q_t	is the demand for period t,
g, h	are the demand function coefficients (intersect and elasticity),
e_t	is the random perturbation, Gaussian i.i.d.: $e_t \sim \mathcal{N}(0, \sigma_e^2)$,

coefficients have known *a priori* Gaussian distribution:

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

the *a priori information matrix* Π_0 is the inverse of the covariance matrix of the vector of coefficients.

Problem objective

maximize the expected discounted profit

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

over the set of *feasible policies*, $i.e.$, prices are functions of the form

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

p_t is a random variable measurable $\sigma(q_1, \ldots, q_{t-1}, \omega)$

A posteriori distribution

the distribution of the coefficients conditional on $q_1, \ldots, q_t$ is

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

with the recursive formulas for the distribution parameters:

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

if price remains constant, the information matrix can be ill-conditioned

Dumb policy

Prices are decided *a priori*, p_t is not a function of $q_1, \dots, q_{t-1}$.
The expected period t profit is:

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

Maximizing price and expected profit:

$$p_1^{\mathrm{m}} = \frac{\hat{g}_0 + c\,\hat{h}_0}{2\hat{h}_0}, \qquad \mathbf{E}\left(R_t(p_0^{\mathrm{m}})\right) = \frac{1}{4\hat{h}_0}\left(\hat{g}_0 + c\,\hat{h}_0\right)^2 - c\,\hat{g}_0.$$

Does not make use of new information about the demand function that becomes available through the observation of q_t.
In the full information case the dumb policy is optimal.

Myopic policy

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

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

Maximizing price and conditional expected profit:

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

Maximizes the immediate expected profit over all feasible policies.
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.
While learning occurs, there is no design for learning.
In the full information case the myopic policy is also optimal.

The paradox of noise: myopic price with "dithering"

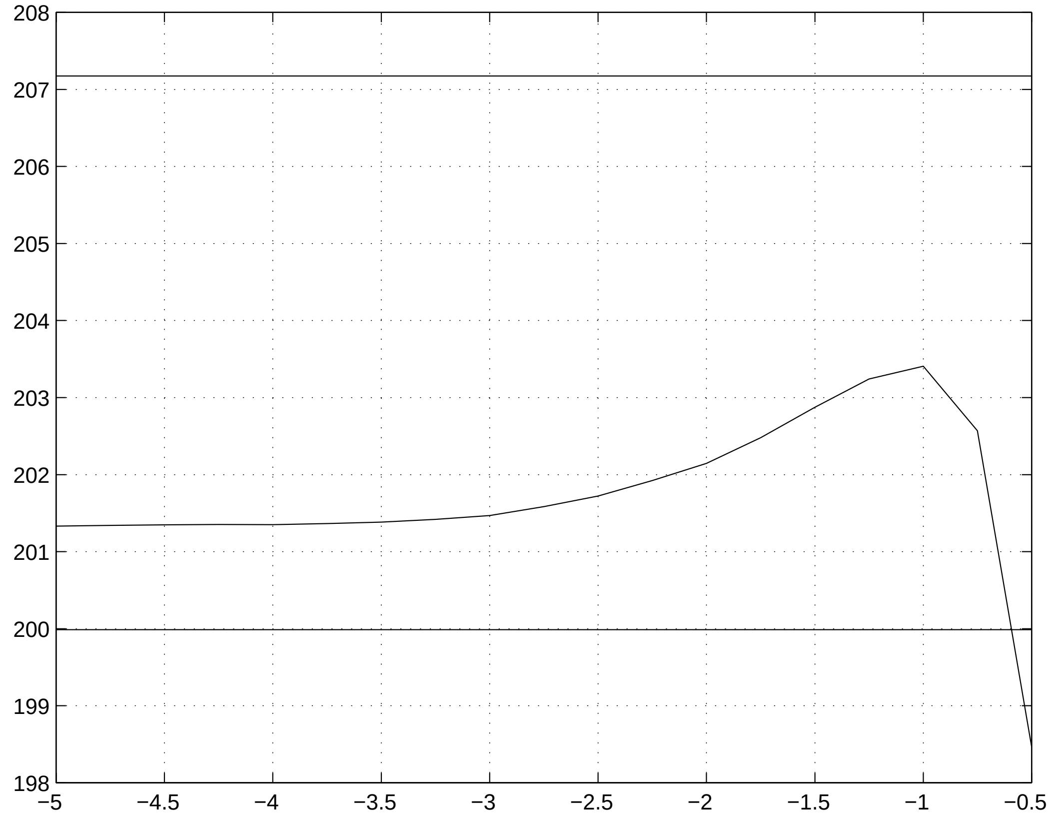


Figure 1: Profit as a fct. of log of dithering level.
(Botton line is dumb policy, top is full information.)

A two period example

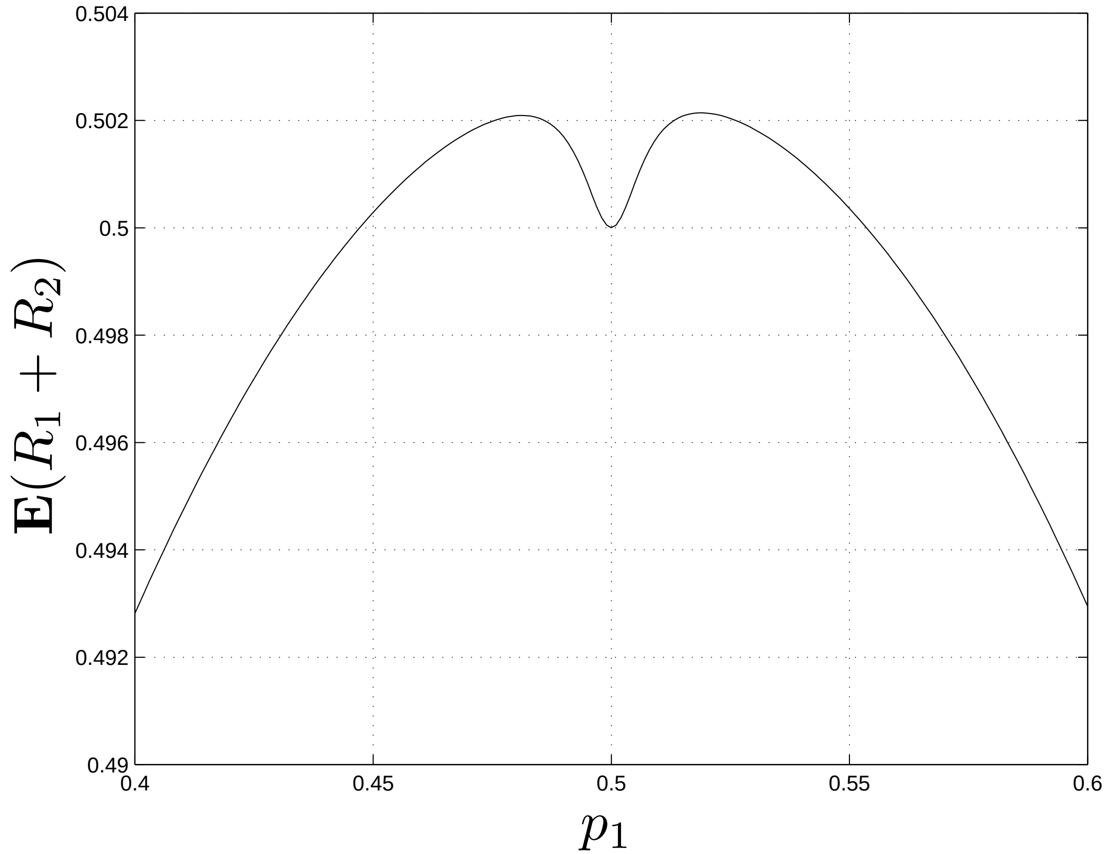


Figure 2: Expected profit as fct. of first-period price.

Dynamic program

The exact solution to the pricing problem is given by the dynamic program:

$$
\begin{aligned}
& V_t\,(p_1, \ldots, p_{t-1}) = \\
& \quad = \quad \sup_{p_t} \mathbf{E}\Big(\,(p_t - c)\,(g - hp_t + e_t) + \delta\, V_{t+1}\,(p_1, \ldots, p_t) \mid t-1\,\Big),
\end{aligned}
$$

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 $\sigma(q_1, \ldots, q_{t-1})$.)

In general, V_t is a very complicated function of $p_1, \ldots, p_{t-1}$, on which the conditional distribution of g and h depends.

Convex approximation

Minimize the expected profit loss relative to the full information case:

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

where the full-information optimal price is:

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

Divide the profit loss into two terms (p_t^{m} is the myopic price):

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

First term: cost of ignorance

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

Taylor series to 2^{nd} order, conditional expectation:

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

Information matrix Π_t is approximated by matrix P_t, linear in the prices:

$$P_t = P_{t-1} + \sigma_e^{-2}\left[\begin{array}{cc} 1 & -p_t \\ -p_t & 2p_t^{\mathrm{r}}p_t - (p_t^{\mathrm{r}})^2 \end{array}\right]$$

for $t = 1, \ldots, T$, and $P_0 = \Pi_0$. The p_t^{r} are a sequence of reference prices.

Second term: cost of experimentation

$$\mathbf{E}\left(R_t(p_t^{\mathrm{m}}) - R_t(p_t) \middle| t-1\right) = \hat{h}_{t-1}\left(p_t - p_t^{\mathrm{m}}\right)^2 .$$

1st + 2nd, total conditional expected profit loss:

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

$$\approx \underbrace{\frac{1}{4\hat{h}_{t-1}} \begin{bmatrix} 1 & -\hat{g}_{t-1}/\hat{h}_{t-1} \end{bmatrix} P_{t-1}^{-1} \begin{bmatrix} 1 \\ -\hat{g}_{t-1}/\hat{h}_{t-1} \end{bmatrix}}_{1} + \underbrace{\hat{h}_{t-1}\left(p_t - p_t^{\mathrm{m}}\right)^2}_{2} .$$

The approximation can be expected to be tight for Π_{t-1} large — if p_t is close to the reference price sequence p_t^{r} and close to the myopic prices p_t^{m}.

Stochastic approximation

$\hat{g}_{t-1}$, $\hat{h}_{t-1}$, and p_t^{m} are random variables whose distributions depend non-trivially on the prices $p_1, \ldots, p_{t-1}$.

Key approximation: $\hat{g}_{t-1} = \hat{g}_0$ and $\hat{h}_{t-1} = \hat{h}_0$, (and, therefore, $p_t^{\mathrm{m}} = p_1^{\mathrm{m}}$.)

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.

Use tower property of conditional expectation, dynamic prog. simplifies to:

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

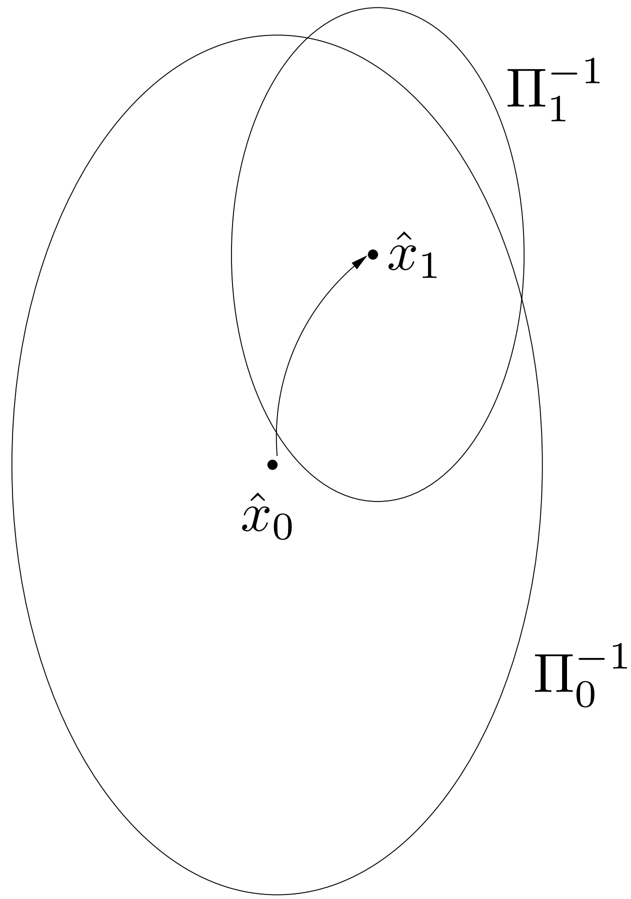


Figure 3: Changes in the conditional distribution.
The change in volume is deterministic (given the price).
The change in center is random.

Convex program

using Schur complements, equivalent semidefinite and quadratic program:

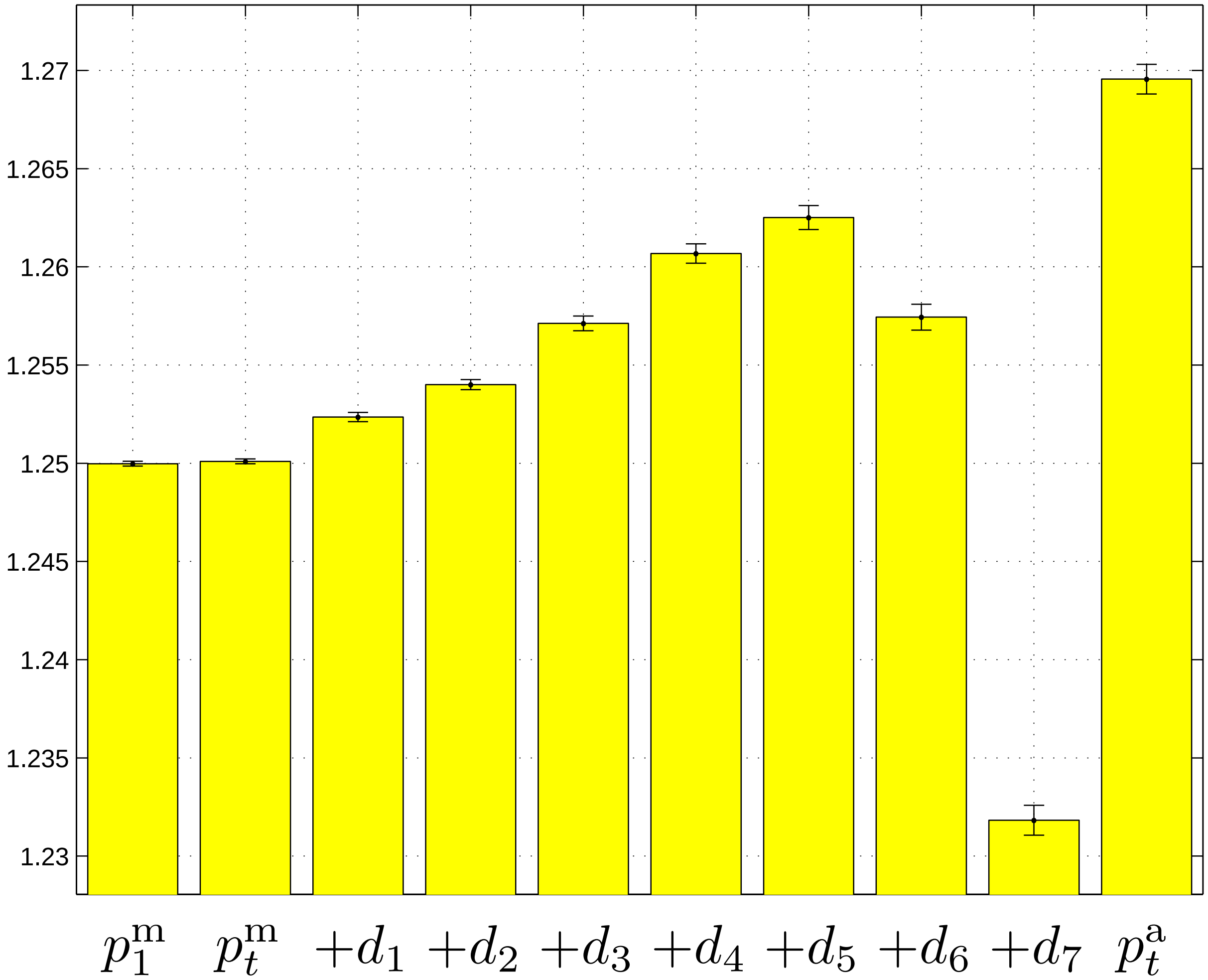
$$\begin{array}{ll}\text{minimize} & \sum_{t=1}^{T} \delta^{t-1}\left(\alpha_t + \beta_t\right) \\ \text{subject to} & \end{array}$$

$$\begin{bmatrix} 4\,\hat{h}_0\,\alpha_t & 1 & -\hat{g}_0/\hat{h}_0 \\ 1 & \Pi_0^{1,1} + \sigma_e^{-2}(t-1) & \Pi_0^{1,2} - \sigma_e^{-2}\sum_{k=1}^{t-1} p_k \\ -\hat{g}_0/\hat{h}_0 & \Pi_0^{2,1} - \sigma_e^{-2}\sum_{k=1}^{t-1} p_k & \Pi_0^{2,2} + \sigma_e^{-2}\sum_{k=1}^{t-1}\left(2p_k^{\mathrm{r}} p_k - (p_k^{\mathrm{r}})^2\right) \end{bmatrix} \succeq 0, \quad t = 1, \ldots, T$$

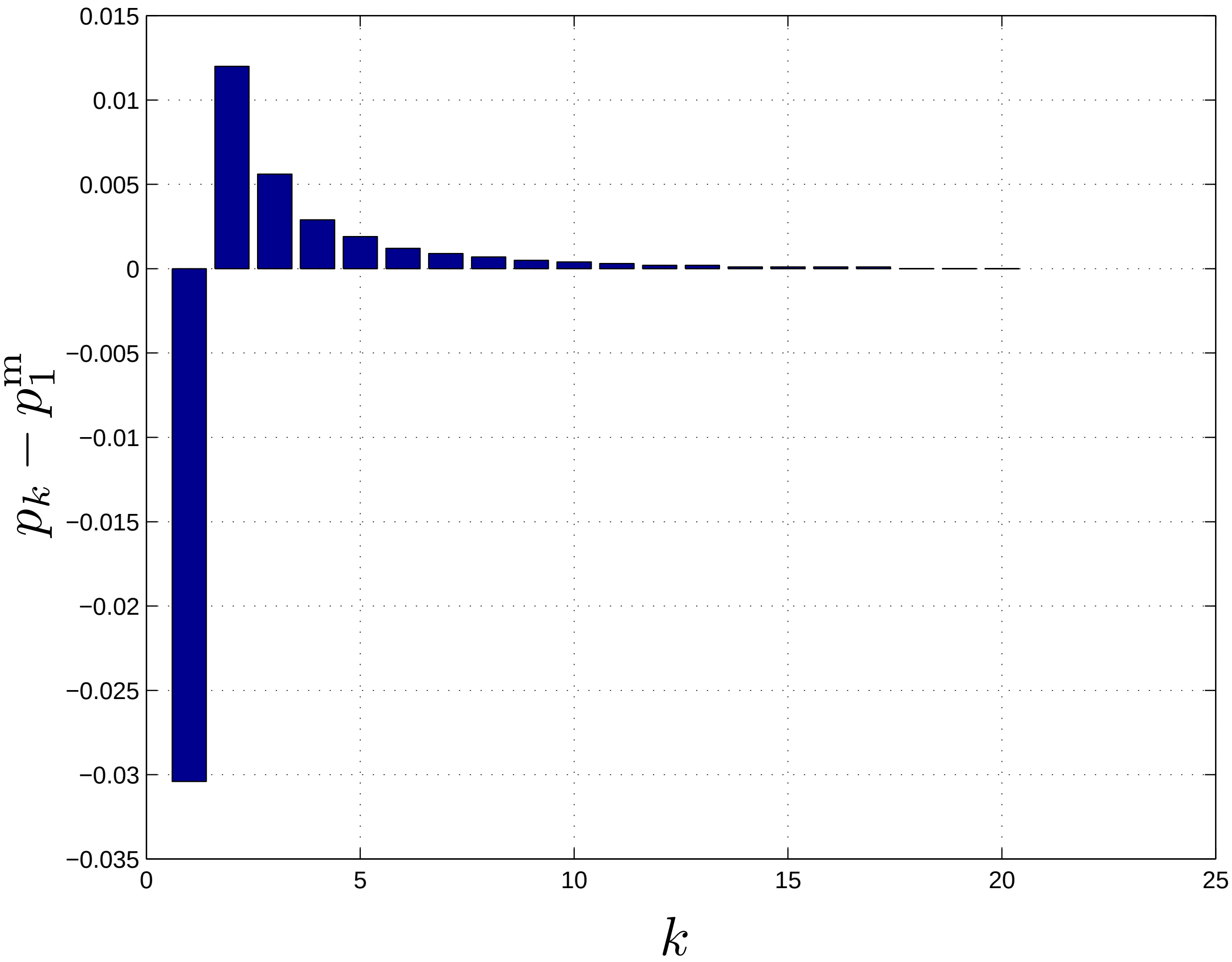
$$\hat{h}_0\left(p_t - p_1^{\mathrm{m}}\right)^2 \leq \beta_t, \quad t = 1, \ldots, T.$$

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

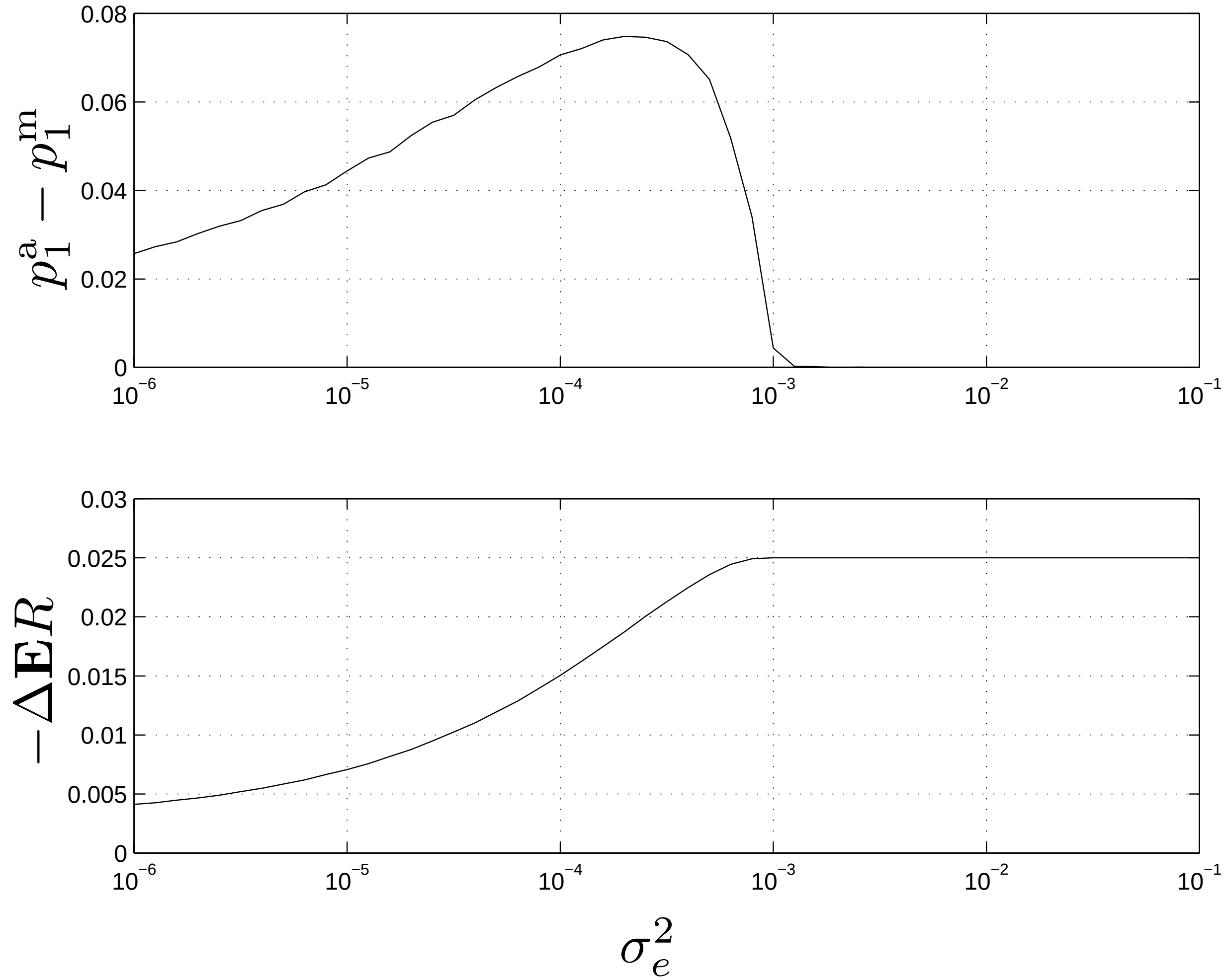
Policy comparison by Monte Carlo



Optimal reference price sequence at $t = 1$



Effect of demand noise variance



Extensions

- n products (information matrix size $O(n^4)$, introduce structure?)
- time-varying demand function
- other demand models, marketing variables
- competition (DP + game)

Conclusions

- active price exploration or discovery, practical importance
- price variations are a rational response to uncertainty about demand, but depend on problem parameters in non-trivial ways
- faster computers, better optimization algorithms: can use policies that optimize on-line
- early results, simple class of problems, many extensions
- dynamic program can be effectively approximated