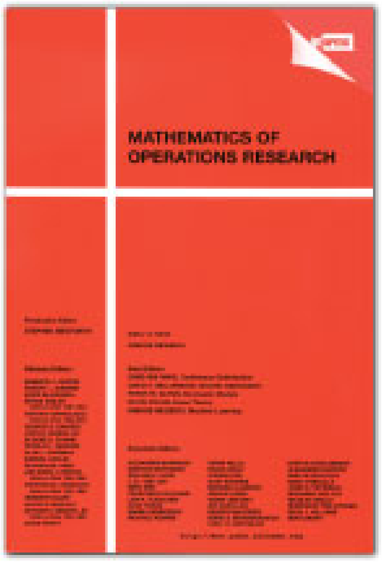

Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution

Arnoud V. den Boer

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 39, No. 3, August 2014, pp. 863–888
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.2013.0636

Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution

Arnoud V. den Boer
University of Twente, 7500 AE Enschede, The Netherlands, a.v.denboer@utwente.nl

We study a dynamic pricing problem with multiple products and infinite inventories. The demand for these products depends on the selling prices and on parameters unknown to the seller. Their value can be learned from accumulating sales data using statistical estimation techniques. The quality of the parameter estimates is influenced by the amount of price dispersion; however, a large amount of variation in the selling prices can be costly since it means that suboptimal prices are used. The seller thus needs to balance optimizing the quality of the parameter estimates and optimizing instant revenue, i.e., exploitation and exploration.

In this study we propose a pricing policy for this dynamic pricing problem. The key idea is to use at each time period the price that is optimal with respect to current parameter estimates, with an additional constraint that ensures sufficient price dispersion. We measure the price dispersion by the smallest eigenvalue of the design matrix and show how a desired growth rate of this eigenvalue can be achieved by a simple quadratic constraint in the price-optimization problem. We study the performance of our pricing policy by providing bounds on the regret, which measures the expected revenue loss caused by using suboptimal prices.

Keywords: marketing: estimation/statistical techniques; pricing; statistics: estimation
MSC2000 subject classification: Primary: 90B60; secondary: 62L05
OR/MS subject classification: marketing: estimation/statistical techniques, pricing; statistics: estimation
History: Received April 18, 2011; revised October 9, 2012, July 17, 2013, and October 4, 2013. Published online in *Articles in Advance* February 13, 2014.

1. Introduction. For firms that sell products or deliver services, it is important to know which selling price generates the highest revenue. This price is generally unknown to the firm, but it can be learned by experimenting with the selling prices. In particular, firms that sell products via the Internet can easily change their selling prices. Since price experimentation means that suboptimal prices are chosen for some time periods, price experimentation can be costly and should be conducted properly. That means that the seller should balance between minimizing the revenue losses due to experimentation and gaining as much information as possible about the relation between price and demand. In other words, in order to learn the price that generates the highest revenue, the firm needs a pricing policy that includes price experimentation in such a way that learning and instant optimization are optimally balanced.

This problem has recently received much research attention. Under different assumptions, pricing policies have been proposed and (sometimes) performance characteristics have been proven. Parametric demand models were employed by Lobo and Boyd [52], Carvalho and Puterman [18, 17], Bertsimas and Perakis [11], Besbes and Zeevi [12], den Boer and Zwart [27], Broder and Rusmevichientong [15] and Keskin and Zeevi [44]; Bayesian models by Aviv and Pazgal [7], Araman and Caldentey [3], Farias and van Roy [32] and Harrison et al. [38]; and nonparametric demand models have been studied by Kleinberg and Leighton [46], Cope [21], Lim and Shanthikumar [50], Eren and Maglaras [31], Besbes and Zeevi [12]. We refer to den Boer [25] for a more elaborate overview of this literature.

Practically all research on this subject focuses on the single-product case. 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; not all unknown parameters of the system may be learned if one simply applies a single-product pricing policy to each individual product. This motivates the current study on dynamic pricing and learning in a setting with multiple products.

The abundance of literature on pricing-and-learning in a single-product setting contrasts with the relative scarcity of papers that consider multiple products. Exceptions are the nonparametric approach by Besbes and Zeevi [12], the robust optimization approach by Lim et al. [51], the linear demand model studied in the Master's thesis of Le Guen [49], and the work of Keskin and Zeevi [44]. The latter paper, written in parallel with our work, assumes that expected demand is a linear function of price and that the demand distribution is sub-Gaussian, and it derives sufficient conditions that guarantee single-product pricing policies to be asymptotically optimal. The authors also study the performance of so-called "orthogonal pricing policies" in a multiple-product setting. In §5.4 we compare our results with those of Keskin and Zeevi [44] in more detail.

In this paper, we study the aforementioned dynamic pricing problem with multiple products in a general parametric setting. In particular, we assume that 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 that 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, and the current knowledge of the parameter estimates is exploited by choosing the optimal prices with respect to 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 prespecified 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 that 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 $\text{Regret}(T)$, 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 cumulative 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 $\text{Regret}(T) = O(T^{2/3})$.

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 $\text{Regret}(T) = O(\sqrt{T \log(T)})$ 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 [46], Besbes and Zeevi [13], Broder and Rusmevichientong [15]). The upper bound $\text{Regret}(T) = O(\sqrt{T \log(T)})$ 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 [47] 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 [49] or Keskin and Zeevi [44], restrict to linear demand functions or sub-Gaussian demand distributions. In addition, we construct a pricing policy that facilitates learning the unknown parameters; in contrast, in a robust approach such as Lim et al. [51], no learning takes place.

Our proposed adaptive pricing policy is based on an optimization problem, Equation (12), which contains a nonconvex constraint. In §5.1 we discuss computational aspects of solving this optimization problem, and provide several suggestions to reduce the required computation time. Despite these suggestions, however, for large instances the problem may still be computationally intractable. Designing efficient numerical algorithms to obtain exact solutions (or sufficiently good heuristics) for these large instances is an open problem for future research; such algorithms will make our adaptive pricing policy also applicable for large problem instances.

The remainder of this paper is organized as follows. Section 2 introduces the model and notation, discusses some of the assumptions we make, and introduces the maximum quasi-likelihood estimator. Section 3 describes the proposed adaptive pricing policy. In §4.1 we provide an upper bound on the regret of a pricing policy, in terms of the amount of price dispersion. Section 4.2 improves these bounds in case of canonical link functions. Some auxiliary results needed to prove these regret bounds are contained in §4.3. Section 5 addresses computational aspects of the policy, discusses the quality of our regret bounds, compares our study with parallel related work and with the literature on multi-armed bandit problems, provides some context for our extension of a theorem of Lai [47], discusses possible applications to adaptive design of experiments, and shows regret bounds when the optimal price lies outside the set of admissible prices. Two numerical illustrations are provided in §6. All mathematical proofs are contained in §7.

2. Model, assumptions, and estimation method.

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, $(1 \le k \le n)$. The term $p_0(t) = 1$ is convenient for notational reasons. We assume that the prices lie in a compact, convex, nonempty 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 $0 < p_{lk} < p_{hk}$ denotes 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 $p_1 \le p_2$.

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:

$$E[D_k(\mathbf{p})] = h_k(\mathbf{p}^T \beta_k^{(0)}), \quad (\mathbf{p} \in \mathcal{P}), \tag{1}$$

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

Here for all $k = 1, \ldots, n$, the functions h_k: $\mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ and v_k: $\mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ are both thrice continuously differentiable, with $\dot{h}_k(x) = \partial h_k(x)/\partial x > 0$, $v_k(x) > 0$ for all $x \ge 0$, 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

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

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 $\beta_k \in \mathbb{R}^{n+1}$ and $\beta \in \mathbb{R}^{(n+1) \times n}$.

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

$$\mathbf{p}(\beta) = \arg\max_{\mathbf{p} \in \mathcal{P}} r(\mathbf{p}, \beta) \in \mathrm{int}(\mathcal{P}), \tag{4}$$

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

$$H(\mathbf{p}, \beta) = \left(\frac{\partial^2 r(\mathbf{p}, \beta)}{\partial p_i \partial p_j} \right)_{1 \le i, j \le n}, \tag{5}$$

is negative definite at the point $\mathbf{p}(\beta)$. In (4), and throughout this article, $\mathrm{int}(\mathcal{P})$ is defined as $\{1\} \times \mathrm{int}(\{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid (1, p_1, \ldots, p_n) \in \mathcal{P}\})$. The correct optimal price $\mathbf{p}(\beta^{(0)})$ is also denoted by $\mathbf{p}_{\mathrm{opt}}$.

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 $\mathbf{p}_{\mathrm{opt}}$. 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

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

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 $\mathrm{Regret}(T, \psi)$. 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 $\mathrm{tr}(A)$ and $\det(A)$ we denote the trace and determinant of a matrix A, with $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ 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

$$P(t) = \sum_{i=1}^{t} \mathbf{p}(i)\mathbf{p}^T(i). \tag{6}$$

Since the largest and smallest eigenvalues of $P(t)$ play an important role in the analysis, we use shorthand notation $\lambda_{\max}(t) = \lambda_{\max}(P(t))$ and $\lambda_{\min}(t) = \lambda_{\min}(P(t))$. 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 $\mathrm{Regret}(T)$ instead of $\mathrm{Regret}(T, \psi)$.

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 (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 (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 $(\beta^{(0)}_{kl} + \beta^{(0)}_{lk})_{k,\, l=1,\ldots,n}$ is negative definite.

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 [67], McCullagh [54], Godambe and Heyde [35], McCullagh and Nelder [55], Heyde [39] and Gill [34].

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 MQLE of $\beta^{(0)}_k$, denoted by $\hat{\beta}_k(t) \in \mathbb{R}^{n+1}$, is defined as a solution to the $(n+1)$-dimensional equation

$$l_{kt}(\beta_k) = \sum_{i=1}^{t} \frac{\dot{h}_k(\mathbf{p}^T(i)\beta_k)}{\sigma_k^2 v_k(h_k(\mathbf{p}^T(i)\beta_k))} \mathbf{p}(i)(d_{ki} - h_k(\mathbf{p}^T(i)\beta_k)) = 0. \tag{7}$$

The functions h_k are called *canonical* link functions if $\dot{h}_k(x) = v_k(h_k(x))$ for all $x \in \mathbb{R}$, $k = 1, \ldots, n$. This relation holds for normally distributed demand with $h_k(x) = x$, Poisson distributed demand with $h_k(x) = \exp(x)$, and Bernoulli distributed demand with $h_k(x) = \exp(x)/(1 + \exp(x))$. In case of canonical link functions, the estimation Equation (7) simplify considerably to

$$l_{kt}(\beta_k) = \sum_{i=1}^{t} \mathbf{p}(i)(d_{ki} - h_k(\mathbf{p}^T(i)\beta_k)) = 0. \tag{8}$$

Computational methods to solve the MQLE Equation (7) are discussed in Osborne [57] and Heyde and Morton [40].

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: den Boer and Zwart [27] 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. They 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 (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 nonrandom function on $\mathbb{N}$. The motivation for requiring this bound on $\lambda_{\min}(t)$ is because the expected square estimation error can be bounded from above by an expression that is inversely proportional to $\lambda_{\min}(t)$ (Propositions 3 and 4; see also den Boer and Zwart [26] and Lai and Wei [48]). The rate at which the parameter estimates $\hat{\beta}(t)$ converge to $\beta^{(0)}$ can thus be controlled by requiring a minimum growth rate on $\lambda_{\min}(t)$.

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, $\operatorname{tr}(P(t)^{-1})$. This can be justified because for any positive definite $n\times n$ matrix A,

$$\operatorname{tr}(A^{-1})^{-1} \leq \lambda_{\min}(A) \leq n \operatorname{tr}(A^{-1})^{-1}. \tag{9}$$

Thus, $\operatorname{tr}(P(t)^{-1}) = O(L_1(t)^{-1})$ is equivalent to $\lambda_{\min}(P(t)) = \Omega(L_1(t))$. The expression $\operatorname{tr}(P(t)^{-1})$ admits a recursive form via the Sherman-Morrison formula (Bartlett [8]; see Hager [36] for a historical treatment of these type of formulas). In particular, one can show

$$\operatorname{tr}(P(t+1)^{-1}) - \operatorname{tr}(P(t)^{-1}) = -\frac{\|P(t)^{-1}\mathbf{p}(t+1)\|^2}{1+\mathbf{p}^T(t+1)P(t)^{-1}\mathbf{p}(t+1)}. \tag{10}$$

If $\operatorname{tr}(P(t)^{-1}) \leq 1/L_1(t)$ and $\mathbf{p}(t+1)$ is chosen such that the right hand side of (10) satisfies a carefully chosen constraint, we can make sure that $\operatorname{tr}(P(t+1)^{-1}) \leq 1/L_1(t+1)$.

Let $\mathscr{L}$ be the class of nondecreasing differentiable functions $L\colon \mathbb{N}\to\mathbb{R}_{>0}$ such that $\dot{L}(t)=o(1)$, and $t\mapsto 1/L(t)$ is convex. Examples of functions contained in $\mathscr{L}$ 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 $L\in\mathscr{L}$, $L(t)=o(t)$, 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_{L_1}, are outlined below:

Adaptive pricing policy Φ_{L_1} for n products

Initialization: Choose $L_1 \in \mathscr{L}$.
Choose $n+1$ linearly independent initial price vectors $\mathbf{p}(1), \mathbf{p}(2), \ldots, \mathbf{p}(n+1)$ in $\mathscr{P}$.
For all $t \geq n+2$:
Estimation: For each $k=1,\ldots,n$, calculate the MQLE $\hat{\beta}_k(t)$ using the MQLE Equation (7).
Pricing:
(I) If for some k, $\hat{\beta}_k(t)$ does not exist, or $\operatorname{tr}(P(t)^{-1})^{-1} \ngeq L_1(t)$, then set $\mathbf{p}(t+1)=\mathbf{p}(1)$, $\mathbf{p}(t+2)=\mathbf{p}(2), \ldots, \mathbf{p}(t+j)=\mathbf{p}(j)$, where j is the smallest integer such that $\operatorname{tr}(P(t+j)^{-1})^{-1}\geq L_1(t+j)$.
(II) If for all k, $\hat{\beta}_k(t)$ exists, and $\operatorname{tr}(P(t)^{-1})^{-1}\geq L_1(t)$, let $\mathbf{p}_{\text{ceqp}}=\mathbf{p}(\hat{\beta}(t))$, and consider the following cases:
(IIa) If

$$\operatorname{tr}((P(t)+\mathbf{p}_{\text{ceqp}}\mathbf{p}_{\text{ceqp}}^T)^{-1})^{-1} \geq L_1(t+1), \tag{11}$$

then choose $\mathbf{p}(t+1)=\mathbf{p}_{\text{ceqp}}$.
(IIb) If (11) does not hold, then choose $\mathbf{p}(t+1)$ that maximizes

$$\max_{\mathbf{p}\in\mathscr{P}} r(\mathbf{p},\hat{\beta}(t)) \quad \text{s.t. } \frac{\|P(t)^{-1}\mathbf{p}\|^2}{1+\mathbf{p}^T P(t)^{-1}\mathbf{p}} \geq \frac{\dot{L}_1(t)}{L_1(t)^2}, \tag{12}$$

provided there is a feasible solution.
(IIc) If (11) does not hold, and (12) has no feasible solution, then set $\mathbf{p}(t+1)=\mathbf{p}(1)$, $\mathbf{p}(t+2)=\mathbf{p}(2),\ldots,\mathbf{p}(t+j)=\mathbf{p}(j)$, where j is the smallest integer such that $\|P(t+j)^{-1}\mathbf{p}\|^2\left[1+\mathbf{p}^T P(t+j)^{-1}\mathbf{p}\right]^{-1} \geq \dot{L}_1(t+j)L_1(t+j)^{-2}$ is satisfied by some $\mathbf{p}\in\mathscr{P}$.

Ad (I) and (IIc) in the policy description deal with possible nonexistence 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 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 (12) has to be satisfied. Computational aspects of solving (12) are discussed in §5.1.

For sufficiently large t, the maximization problem (12) always has a feasible solution:

Proposition 1 (Feasibility of (12)). *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*

$$\operatorname{tr}(P(t)^{-1})^{-1} \geq L_1(t), \tag{13}$$

$$\operatorname{tr}((P(t) + \mathbf{p}(\hat{\beta}(t))\mathbf{p}(\hat{\beta}(t))^T)^{-1})^{-1} < L_1(t+1), \tag{14}$$

then the set

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

is nonempty.

The following proposition states that for sufficiently large t, the adaptive pricing policy Φ_{L_1} induces a lower bound on $\operatorname{tr}(P(t)^{-1})^{-1}$ and thus by (9) also on $\lambda_{\min}(t)$.

Proposition 2 (Growth Rate of $\operatorname{tr}(P(t)^{-1})^{-1}$). *There are T_1, $C_L \in \mathbb{N}$, depending only on T_0, L_1, and $P(n+1)$, such that for all $t \geq T_1$,*

$$\operatorname{tr}(P(t)^{-1})^{-1} \geq C_L^{-1} L_1(t). \tag{15}$$

4. Bounds on the regret.

In §4.1, we provide upper bounds on the regret induced by Φ_{L_1}, for general link functions. The bounds depends on two characteristics of the 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. Lemma 1 formulates these two characteristics more precisely, and Theorem 1 applies these properties to derive an upper bound on $\text{Regret}(\Phi_{L_1}, T)$, in terms of the function L_1. It turns out that for general link functions, this bound is minimized if $L_1(t)$ grows proportionally to $t^{2/3}$, with a corresponding regret bound of $O(T^{2/3})$. We furthermore show that this regret rate is achieved by any pricing policy that satisfies the conditions of Lemma 1.

In §4.2, we consider the case of canonical link functions. 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 in §4.3 an interesting extension of Lai [47, Theorem 3] to martingale difference sequences.

4.1. General link functions.

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

$$T_{\rho,k} = \sup\{n \in \mathbb{N}: \text{there is no } \beta \in B_{\rho,k} \text{ such that } l_{kt}(\beta) = 0\}, \tag{16}$$

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

$$T_\rho = \max\{T_{\rho,1}, \ldots, T_{\rho,n}\}. \tag{17}$$

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 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 $E[T_\rho^\eta] < \infty$, 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 (7), $\lim_{t\to\infty} \hat{\beta}_k(t) = \beta_k^{(0)}$ a.s., and*

$$E[\|\hat{\beta}_k(t) - \beta_k^{(0)}\|^2 \mathbf{1}_{t > T_\rho}] = O(L_1(t)^{-1}\log(t) + tL_1(t)^{-2}). \tag{18}$$

The assertions about T_ρ follow from applying den Boer and Zwart [26, Theorem 1], for each $T_{\rho,k}$, $k = 1, \ldots, n$ separately, and noting that $T_\rho \leq \sum_{k=1}^n T_{\rho,k}$ a.s. The other statements follow from den Boer and Zwart [26, Theorem 2].

The following lemma lists a number of properties satisfied by Φ_{L_1}.

LEMMA 1. *Let $\rho \in (0, \rho_0)$ such that $\{(\beta_1, \dots, \beta_n) \in \mathbb{R}^{n\times(n+1)} \mid \beta_k \in B_{\rho,k}, k = 1, \dots, n\} \subset V$, where V is defined in §2 and ρ_0 is as in Proposition 3. Let $t_0 \in \mathbb{N}$, $L_1 \in \mathscr{L}$ such that $L_1(t) \geq ct^\alpha$ for all $t \geq t_0$ and some $c > 0$, $\alpha \in (\frac{1}{2}, 1)$. Suppose that policy Φ_{L_1} is used. Then there is a random variable T_2 taking values on $\mathbb{N}$ such that $T_2 \geq T_\rho$ a.s. and $E[T_2^{1/2}] < \infty$; in addition,*
(i) *$\lambda_{\min}(t) \geq L_1(t)$ a.s., for all $t \geq t_0$,*
(ii) *$\sum_{t=1}^T \|\mathbf{p}(t) - \mathbf{p}(\hat{\beta}(t-1))\|^2 \mathbf{1}_{t>T_2} \leq K_2 L_1(T)$ a.s., for all $T \geq t_0$ and some $K_2 > 0$.*

The following theorem provides an upper bound on the regret of Φ_{L_1}, in terms of the function L_1.

THEOREM 1. *Let $t_0 \in \mathbb{N}$, $L_1 \in \mathscr{L}$ such that $L_1(t) \geq ct^\alpha$ for all $t \geq t_0$ and some $c > 0$, $\alpha \in (\frac{1}{2}, 1)$. Then*

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

In Theorem 1, the choice $L_1(t) = ct^{2/3}$, for some $c > 0$, yields $\text{Regret}(\Phi_{L_1}, T) = O(T^{2/3})$. 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 $\tilde{L}_1 \in \mathscr{L}$ such that $L_1 = o(\tilde{L}_1)$ or $\tilde{L}_1 = o(L_1)$.

In addition, we note that the regret bound of Theorem 1 is valid for *any* pricing policy ψ that satisfies the properties of Lemma 1. More precisely, if there are $\rho \in (0, \rho_0)$ and a random variable $T_2 \geq T_\rho$ a.s. with $E[T_2^{1/2}] < \infty$, and ψ implies (i) and (ii) of Lemma 1, then Theorem 1 is also valid for ψ.

4.2. Canonical link functions. As already mentioned in §2.3, the estimation equations for $\hat{\beta}(t)$ simplify considerably if the link functions h_k are all canonical; i.e., if $\dot{h}_k = v_k \circ h_k$ for all $k = 1, \dots, n$. As a result, sharper bounds on the estimation error can be derived. In particular, in den Boer and Zwart [26, Theorem 3], it is shown that in case of canonical link functions h_k, and assuming precisely the same conditions as Proposition 3, the convergence rates (18) can be improved to

$$E[\|\hat{\beta}_k(t) - \beta_k^{(0)}\|^2 \mathbf{1}_{t>T_\rho}] = O(L_1(t)^{-1}\log(t)). \tag{19}$$

It is easy to see from the proof of Theorem 1 that these improved bounds (19) for canonical link functions imply that the regret bound of Theorem 1 can be improved to

$$\text{Regret}(\Phi_{L_1}, T) = O\left(L_1(T) + \sum_{t=1}^{T} L_1(t)^{-1}\log(t)\right), \tag{20}$$

assuming $L_1(t) \geq ct^\alpha$ for some $\alpha \in (1/2, 1)$, $c > 0$, $t_0 \in \mathbb{N}$, and all $t \geq t_0$. The choice $L_1(t) = ct^{1/2+\delta}$, for $c > 0$ and small $\delta > 0$, then implies $\text{Regret}(\Phi_{L_1}, T) = O(T^{1/2+\delta})$, which is a substantial improvement to the rate $T^{2/3}$ derived in §4.1.

However, one can show that the optimal choice that minimizes the right-hand side of (20) is $L_1(t) = c\sqrt{t\log(t)}$, $(c > 0)$, which would lead to $\text{Regret}(\Phi_{L_1}, T) = O(\sqrt{T\log(T)})$. This choice is optimal in the following sense: if $L_1(t) = c\sqrt{t\log(t)}$ and $\tilde{L}_1 \in \mathscr{L}$ 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 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 3 is still valid if $L_1(t) \geq c\sqrt{t\log(t)}$ for a sufficiently large $c > 0$. One then can show that in Theorem 1, the choice $L_1(t) = c\sqrt{t\log(t)}$ with sufficiently large c leads to $\text{Regret}(\Phi_{L_1}, T) = O(\sqrt{T\log(T)})$, when the link functions are canonical. This bound is again not only valid for the policy Φ_{L_1} but also for any pricing policy satisfying Lemma 1 with $L_1(t) = c\sqrt{t\log(t)}$ and c sufficiently large.

Proposition 4 (Strong Consistency and Convergence Rates). *Suppose there are $t_0 \in \mathbb{N}$ and $c > 0$ such that $L_1(t) \geq c\sqrt{t\log(t)}$ a.s. for all $t \geq t_0$. Then there is a $\rho_0 > 0$, and for all $\rho \in (0, \rho_0)$ there is a $c_\rho^* > 0$, such that $T_\rho < \infty$ a.s. and $E[T_\rho^\eta] < \infty$ for all $0 < \eta < (\gamma - 1)/2$, 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 (7), $\lim_{t\to\infty} \hat{\beta}_k(t) = \beta_k^{(0)}$ a.s., and*

$$E[\|\hat{\beta}_k(t) - \beta_k^{(0)}\|^2 \mathbf{1}_{t>T_\rho}] = O(L_1(t)^{-1}\log(t)). \tag{21}$$

The proof is based on Theorems 1 and 3 of den Boer and Zwart [26] and on Proposition 5 contained in the next section.

Observe again that the bound $\text{Regret}(\psi, T) = O(\sqrt{T\log(T)})$ is valid for *any* pricing policy that satisfies the properties of Lemma 1 with $L_1(t) = c\sqrt{t\log(t)}$ and c sufficiently large.

4.3. Auxiliary results. This section contains a number of auxiliary results that are needed to prove Proposition 4.

Proposition 5. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence with respect to 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 $\eta > 0$, $r > 2(\eta+1)$, and $c > 2\sigma\sqrt{\eta}$ 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 $\sup_{i\in\mathbb{N}} E[|X_i|^r] \leq C < \infty$ for some $C > 0$, then*

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

A key ingredient to Proposition 5 is the following theorem. This was proven in Lai [47, Theorem 3] for i.i.d. random variables; we extend it to martingale difference sequences.

Theorem 2. *Let $(X_i)_{i\in\mathbb{N}}$ be a martingale difference sequence with respect to 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*

$$\sum_{n=1}^\infty n^a P(|S_n| > \delta\sqrt{2n\log(n)}) < \infty, \tag{22}$$

$$\sum_{n=1}^\infty n^a P\left(\sup_{1\leq i\leq n} |S_i| > \delta\sqrt{2n\log(n)}\right) < \infty. \tag{23}$$

The proof makes use of the following result, which is based on Stout [65].

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

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

5. Discussion.

5.1. Computational aspects. If (11) is satisfied, then under some mild assumptions on the revenue function, the revenue-maximizing price $\mathbf{p}_{\text{ceqp}}$ can be determined using a gradient-ascent method. If (11) does not hold, then the additional constraint in (12) leads to a more complicated optimization problem with a nonconvex feasible set. In this section we show how an (approximate) solution can be obtained that does not affect the asymptotic growth rate of the regret.

Fix t. We assume that $\mathcal{P}$ is defined by a number of linear constraints. Write

$$A = \frac{L_1(t)^2}{\dot{L}_1(t)} P(t)^{-2} - P(t)^{-1},$$

and observe that the constraint in (12) can be rewritten as $\mathbf{p}^T A \mathbf{p} \geq 1$.

All relevant choices of L_1 in this paper, i.e., $L_1(t) = c\sqrt{t\log(t)}$ or $L_1(t) = ct^\alpha$ for some $0 < \alpha < 1$ and $c > 0$, satisfy $t(1 + \max_{\mathbf{p}\in\mathcal{P}} \|\mathbf{p}\|^2) \leq L_1(t)^2/\dot{L}_1(t)$ for sufficiently large t. This implies

$$\frac{y^T P(t) y}{y^T y} \leq \lambda_{\max}(P(t)) \leq \text{tr}(P(t)) < t\left(1 + \max_{\mathbf{p}\in\mathcal{P}} \|\mathbf{p}\|^2\right) \leq \frac{L_1(t)^2}{\dot{L}_1(t)},$$

for all nonzero $y \in \mathbb{R}^{n+1}$. It follows that for all nonzero $z \in \mathbb{R}^{n+1}$, writing $y = P(t)^{-1}z$,

$$z^T A z = y^T P(t)\left(\frac{L_1(t)^2}{\dot{L}_1(t)} P(t)^{-2} - P(t)^{-1}\right) P(t) y = y^T y \left(\frac{L_1(t)^2}{\dot{L}_1(t)} - \frac{y^T P(t) y}{y^T y}\right) > 0.$$

This implies that A is positive definite, and that the feasible region $\{\mathbf{p} \in \mathcal{P} \mid \mathbf{p}^T A \mathbf{p} \geq 1\}$ of (12) is nonconvex. Optimization problems with a nonconvex feasible region may in general be untractable. We show that in our setting, however, the optimization problem (12) can be solved exactly (in case of linear demand functions), or the optimal solution can be approximated without affecting the asymptotic growth rate of the regret (in case of nonlinear demand functions).

If all the demand functions are linear (i.e., the functions h_k all equal the identity function), then the revenue function $r(\mathbf{p}, \hat{\beta}(t))$ is a quadratic function, and (12) is a quadratic optimization problem with a single quadratic constraint $\mathbf{p}^T A \mathbf{p} \geq 1$ and m linear constraints of the form $b_j^T \mathbf{p} \leq c_j$ for some $m \in \mathbb{N}$, $b_j \in \mathbb{R}^{n+1}$, $c_j \in \mathbb{R}$, $j = 1, \ldots, m$; here the m linear constraints define $\mathcal{P}$. To solve (12), we construct for each $S \subset \{1, \ldots, m\}$ a new optimization problem P_S, given by

$$(P_S) \quad \max_{\mathbf{p} \in \mathbb{R}^{n+1}} r(\mathbf{p}, \hat{\beta}(t)) \quad \text{s.t. } \mathbf{p}^T A \mathbf{p} \geq 1 \quad \text{and} \quad b_j^T \mathbf{p} = c_j \text{ for all } j \in S. \tag{24}$$

By substituting the equality constraints $b_j^T \mathbf{p} = c_j$ $(j \in S)$ into the quadratic objective function $r(\mathbf{p}, \hat{\beta}(t))$ and the quadratic constraint $\mathbf{p}^T A \mathbf{p} \geq 1$, P_S reduces to a quadratic optimization problem with a single quadratic constraint (on a possibly lower-dimensional subspace of $\mathbb{R}^{n+1}$). This problem can be solved efficiently by application of the S-Lemma and a reduction to a semidefinite program, as shown in Boyd and Vandenberghe [14, Appendix B]. Let $\mathbf{p}_S^*$ denote an optimal solution of P_S. An optimal solution to (12) is obtained by simply maximizing $r(\mathbf{p}, \hat{\beta}(t))$ over the finite set of $\{\mathbf{p}_S^* \mid S \in \{1, \ldots, m\}\} \cap \mathcal{P}$. This finite set is nonempty since it contains an optimal solution to (12).

For nonlinear demand functions, (12) may be more difficult to solve, and we therefore propose an approximate solution. An important observation is that instead of a solution to (12), *any* choice of $\mathbf{p}(t+1)$ that satisfies $\mathbf{p}(t+1)^T A \mathbf{p}(t+1) \geq 1$ and $\|\mathbf{p}(t+1) - \mathbf{p}_{\text{ceqp}}\|^2 \mathbf{1}_{t>T_2} \leq K_2 \dot{L}_1(t) \mathbf{1}_{t>T_2}$ leads to the same regret bounds proven in Theorem 1 (here K_2 and T_2 are as in Lemma 1).

A particular feasible choice of $\mathbf{p}(t+1)$ can be obtained by overestimating the revenue function with a quadratic function. To this end, take l and L as in (31) and (32), and define

$$g(\mathbf{p}) = r(\mathbf{p}_{\text{ceqp}}, \hat{\beta}(t)) + \frac{1}{2} L (\mathbf{p} - \mathbf{p}_{\text{ceqp}})^T (\mathbf{p} - \mathbf{p}_{\text{ceqp}}).$$

Our approximate solution to (12) is given by

$$\tilde{\mathbf{p}}(t+1) \in \arg\max\{g(\mathbf{p}) \mid \mathbf{p} \in \mathcal{P},\ \mathbf{p}^T A \mathbf{p} \geq 1,\ \|\mathbf{p} - \mathbf{p}_{\text{ceqp}}\|^2 \leq K_2 \dot{L}_1(t)\}. \tag{25}$$

Observe that $r(\mathbf{p}_{\text{ceqp}}, \hat{\beta}(t))$ does not depend on $\mathbf{p}$ and that L is strictly smaller than zero. As a result, (25) is equal to

$$\tilde{\mathbf{p}}(t+1) \in \arg\min\left\{\frac{|L|}{2} \|\mathbf{p} - \mathbf{p}_{\text{ceqp}}\|^2 \,\middle|\, \mathbf{p} \in \mathcal{P},\ \mathbf{p}^T A \mathbf{p} \geq 1,\ \|\mathbf{p} - \mathbf{p}_{\text{ceqp}}\|^2 \leq K_2 \dot{L}_1(t)\right\}. \tag{26}$$

For $t > T_2$, (26) always has a feasible solution (namely, the optimal solution to (12)). The constraint $\|\mathbf{p} - \mathbf{p}_{\text{ceqp}}\|^2 \leq K_2 \dot{L}_1(t)$ is then not active, and (26) is equal to a quadratic optimization problem with linear constraints and a single quadratic constraint. This problem can be solved efficiently as described above.

The instantaneous revenue loss, caused by choosing $\tilde{\mathbf{p}}(t+1)$ instead of $\mathbf{p}(t+1)$, is bounded by

$$\begin{aligned}
& r(\mathbf{p}(t+1), \beta^{(0)}) - r(\tilde{\mathbf{p}}(t+1), \beta^{(0)}) \\
& \quad \leq \left(r(\mathbf{p}_{\text{ceqp}}, \hat{\beta}(t)) + \frac{1}{2} L (\mathbf{p}(t+1) - \mathbf{p}_{\text{ceqp}})^T (\mathbf{p}(t+1) - \mathbf{p}_{\text{ceqp}})\right) \\
& \qquad - \left(r(\mathbf{p}_{\text{ceqp}}, \hat{\beta}(t)) + \frac{1}{2} l (\tilde{\mathbf{p}}(t+1) - \mathbf{p}_{\text{ceqp}})^T (\tilde{\mathbf{p}}(t+1) - \mathbf{p}_{\text{ceqp}})\right) \\
& \quad \leq g(\mathbf{p}(t+1)) - g(\tilde{\mathbf{p}}(t+1)) + \frac{1}{2}(L - l)(\tilde{\mathbf{p}}(t+1) - \mathbf{p}_{\text{ceqp}})^T (\tilde{\mathbf{p}}(t+1) - \mathbf{p}_{\text{ceqp}}) \\
& \quad \leq \frac{1}{2}(L - l) K_2^2 \dot{L}_1(t),
\end{aligned}$$

which converges to zero as $t \to \infty$. The cumulative revenue loss after T periods, caused by this price approximation-scheme, is $O(\sum_{t=1}^{T} \dot{L}_1(t)) = O(L_1(T))$. The growth rate of the regret in Theorem 1 is thus not affected by this price approximation scheme.

Remark 1. If the number of constraints m is not too big, the solution to (24) for all $S \subset \{1, \ldots, m\}$ can be computed by brute force. This means that 2^m separate optimization problems need to be solved, which in practical applications may require too much computation time if m is large. However, by a number of observations the computation time can be reduced significantly. First, without loss of generality, we can restrict to subsets $S \subset \{1, \ldots, m\}$ with cardinality $|S| \leq n$; the reason is that a system of more than n linear equalities in n variables $\mathbf{p}_1, \ldots, \mathbf{p}_n$ either has no feasible solution or contains at most n linearly independent inequalities. By removing linear dependencies, we are left with a system with at most n equalities.

Second, we have $r(\mathbf{p}_S^*, \hat{\beta}(t)) \leq r(\mathbf{p}_{S'}^*, \hat{\beta}(t))$ whenever $S' \subset S \subset \{1, \ldots, m\}$ since adding constraints cannot improve the optimal solution. As a result, if $\mathbf{p}_{S'}^* \in \mathcal{P}$ for some S', then there is no need to solve P_S for all $S \supset S'$; moreover, for all sets $\bar{S} \in \{1, \ldots, m\}$ for which $P_{\bar{S}}$ has been solved but not yet all sets $\bar{S}' \supset \bar{S}$ and for which $r(\mathbf{p}_{\bar{S}}^*, \hat{\beta}(t)) \leq r(\mathbf{p}_{\bar{S}'}^*, \hat{\beta}(t))$, there is then no need to solve $P_{\bar{S}'}$ for all $\bar{S}' \supset \bar{S}$. These observations suggest that a branch-and-bound type of algorithm can be used (Clausen [20]). The worst-case computation time of such algorithms is typically exponential in m—just as brute-force computation of P_S for all $S \subset \{1, \ldots, m\}$—but may in practice significantly reduce the required computation time.

As an important and relevant problem for future research, we suggest designing and analyzing numerical methods to solve (12) for large values of m, for example by a branch-and-bound type of algorithm. A computational study could shine light on the relation between running time and m and give insight into the largest values of n and m for which our algorithm is usable in practice.

5.2. Quality of regret bounds for general link functions. In §4.1 we show that our adaptive pricing policy Φ_{L_1} has $\text{Regret}(\Phi_{L_1}, T) = O(T^{2/3})$, when $L_1(t) = ct^{2/3}$ for some $c > 0$. In addition we provide sufficient conditions for *any* pricing policy to achieve $O(T^{2/3})$ regret. These bounds are valid for all general link functions but can be improved to $O(\sqrt{T \log(T)})$ in case of canonical link functions, as shown in §4.2. The gap between $T^{2/3}$ and $O(\sqrt{T \log(T)})$ is caused by different bounds on the convergence rates of the maximum quasi-likelihood estimates; in particular, the term $tL_1(t)^{-2}$ in Proposition 3, Equation (18), does not appear in the corresponding Proposition 4, Equation (21).

This additional term $tL_1(t)^{-2}$ can be traced back to den Boer and Zwart [26, Theorem 2]. Because for general link functions and adaptive design no explicit form of $\hat{\beta}_t$ is available, bounds on the convergence rates of the expected square estimation error are derived indirectly via a quadratic inequality in $\hat{\beta}(t) - \beta^{(0)}$. Then Lemma 7 of den Boer and Zwart [26] is applied to derive these bounds, yielding a dependence on the first two eigenvalues of $P(t)$. In a single-product setting the second-smallest eigenvalue of $P(t)$ equals $\lambda_{\max}(P(t))$, which grows linearly in t; as a result, the term $tL_1(t)^{-1}L_2(t)^{-1} \sim L_1(t)^{-1}$ is dominated by the term $\log(t)L_1(t)^{-1}$. In the multi-product setting this is not the case, leading to the term $tL_1(t)^{-2}$ in Proposition 3. This is the main reason why in the multi-product setting with general link functions we get $\text{Regret}(T) = O(T^{2/3})$, whereas in the single-product setting with general link function we can get regret close to $\sqrt{T}$ (as in den Boer and Zwart [27]).

It is not clear if the convergence rates of den Boer and Zwart [26, Theorem 2] can be improved upon. Chang [19] claims to prove a.s. convergence rates on $\|\hat{\beta}(t) - \beta^{(0)}\|^2$ that do not include the term $tL_1(t)^{-2}$, but his proof contains a mistake (see Remark 1 of den Boer and Zwart [26]). Yin et al. [69], considering maximum quasi-likelihood estimators with adaptive design, general link functions, and multivariate response data, provide convergence rates that, in the case of bounded design, imply

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

Thus, here again a term $t\lambda_{\min}(t)^{-2}$ appears in the convergence rates.

Summarizing, the statistical literature on maximum quasi-likelihood estimators does not provide a conclusive answer to the question whether the convergence rates (18) of maximum quasi-likelihood estimators for general link functions and adaptive design are tight. This area is an interesting and important direction for future research.

5.3. Quality of regret bounds for canonical link functions. In §4.2 we show that in case of canonical link functions, our adaptive pricing policy Φ_{L_1} has $\text{Regret}(\Phi_{L_1}, T) = O(\sqrt{T \log(T)})$, when $L_1(t) = c\sqrt{t \log(t)}$ for some sufficiently large $c > 0$.

Under different sets of assumptions, it has been shown by Kleinberg and Leighton [46], Broder and Rusmevichientong [15], and Besbes and Zeevi [13] that there is no pricing policy with $\text{Regret}(T) = o(\sqrt{T})$. This means that apart from the $\sqrt{\log(T)}$ term, our adaptive policy has optimal asymptotic growth rate whenever the link functions are canonical. As a result, for many demand models that are used in practice (e.g., normally distributed demand with linear link function, Bernoulli distributed demand with logit link function, and Poisson distributed demand with exponential link function), our adaptive pricing policy has near-optimal performance.

The factor $\sqrt{\log(T)}$ represents a gap between the upper bound $\text{Regret}(\Phi_{L_1}, T) = O(\sqrt{T\log(T)})$ and the optimal growth rate $O(\sqrt{T})$ and can be traced back to two sources: Proposition 5 and den Boer and Zwart [26, Proposition 2].

Proposition 5 is a building block to prove that for sufficiently large t, a solution to the likelihood equations exists in a neighborhood of $\beta^{(0)}$. We do this by relating the implicitly defined $\hat{\beta}_k(t)$ to 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 $c > 0$. Proposition 5 shows that T is finite a.s. and has some finite moments; these properties are used to derive the desired existence properties of the quasi-likelihood estimator. 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. Any attempt to remove the $\sqrt{\log(n)}$ term here would require completely different proof techniques to deal with possible nonexistence of the maximum quasi-likelihood estimator.

The second source of the $\sqrt{\log(T)}$ term is Proposition 2 of den Boer and Zwart [26], 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 [48], who derive a.s. convergence rates, a $\log(t)$ term appears in the equations. An example provided by Nassiri-Toussi and Ren [56] shows that at least in some instances, the $\log(t)$ term is present in the asymptotic behavior of the estimates.

Summarizing, there does not seem to be a straightforward way to remove the $\sqrt{\log(T)}$-term from the regret bounds, and in fact, it is not clear if it is possible at all. In this respect, it is interesting to note that many papers on online learning problems with adaptive design report regret bounds that involve logarithmic terms; see for instance (Dani et al. [23], Bartlett and Tewari [9], Rusmevichientong and Tsitsiklis [61], Jaksch et al. [42], and Abbasi-Yadkori and Szepesvári [1]). Studying whether these logarithmic factors can be removed from the regret bounds may refine the performance analysis of many algorithms in online learning problems.

5.4. Comparison with parallel work. Keskin and Zeevi [44] is a recent study on multi-product pricing that is closely related to our work. We here provide a brief summary of similarities and differences between the two papers.

In Keskin and Zeevi [44], the authors study dynamic pricing with multiple products, under the assumptions of a linear demand function and sub-Gaussian disturbance terms. The unknown parameters of the demand function are estimated with least-squares linear regression. For a certain class of pricing policies, called "orthogonal pricing policies," conditions are derived that guarantee $\text{Regret}(T) = O(\sqrt{T}\log T)$. One of these conditions is to ensure that the smallest eigenvalue of the design matrix grows with rate $\sqrt{t}$. This is similar to our approach, and it guarantees that the parameter estimates converge at a certain rate to the true values.

A distinction between this and our work is the level of generality. Whereas we allow for a very large class of demand functions and (even heavy-tailed) noise distributions, Keskin and Zeevi [44] restrict to linear demand functions and sub-Gaussian disturbance terms. As a result, our analysis covers several often-used nonlinear demand models, such as Bernoulli distributed demand with logit link function or Poisson distributed demand with exponential link function.

5.5. Connection to multi-armed bandit problems. The pricing-and-learning problem considered in this paper is an example of a sequential decision problem under uncertainty, and as such closely related to the multi-armed bandit (MAB) problem: an archetypal problem for which a trade-off between learning and instant optimization, i.e., between exploration and exploitation, is encountered (see Bubeck and Cesa-Bianchi [16] for a recent survey). Well-known algorithms for MAB problems are the family of upper-confidence-bound (UCB) algorithms (Auer et al. [5]) or various weight-updating methods (Arora et al. [4]). Some examples of pricing problems that are modeled as an MAB problem are Rothschild [60], Xia and Dube [68], and Cope [21]. These studies assume that the set of admissible prices or actions is discrete and finite.

We allow $\mathcal{P}$ to be continuous, and this makes our study related to continuum-arm MAB problems. These problems have received considerable research attention in recent years. Performance analysis of decision policies under various assumptions are studied by, among others, Kleinberg [45], Auer et al. [6], Cope [22], Wang et al. [66], Rusmevichientong and Tsitsiklis [61], Filippi et al. [33], Abbasi-Yadkori et al. [2], and Yu and Mannor [70].

5.6. Probabilities of moderate deviations. Theorem 2 stands in a long tradition of literature that studies necessary and sufficient conditions guaranteeing

$$\sum_{n\in\mathbb{N}} a_n P(|S_n| \geq b_n) < \infty, \qquad \text{and} \qquad \sum_{n\in\mathbb{N}} a_n P\Big(\sup_{k\leq n} |S_k| \geq b_n\Big) < \infty, \tag{27}$$

where $(S_n)_{n\in\mathbb{N}}$ is a random walk and $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ are nonrandom sequences.

For example, if b_n is of the form $b_n = cn^{1/p}$, $(0 < p < 2,\ c > 0)$, and $S_n = \sum_{i=1}^n X_i$ with $(X_i)_{i\in\mathbb{N}}$ a sequence of i.i.d. zero-mean random variables, various classical results for (27) have been obtained (among others) by Hsu and Robbins [41], Erdős [29, 30], Katz [43], and Baum and Katz [10]. Recently, Stoica [64] has extended some of these results to the case where S_n is a martingale.

In Theorem 2, we consider b_n of the form $b_n = \delta\sqrt{n\log(n)}$. In case $S_n = \sum_{i=1}^n X_i$ with $(X_i)_{i\in\mathbb{N}}$ a sequence of i.i.d. zero-mean random variables, results for (27) have been obtained by Davis [24] and Lai [47]. The quantity $P(|S_n| > c\sqrt{n\log(n)})$ is then usually called a probability of moderate deviation, (see Spataru [63]). We contribute to the literature on these probabilities of moderate deviations by extending Lai [47, Theorem 3] to the case where S_n is a martingale (Theorem 2) and by showing finiteness of moments of the closely related last-times $\sup\{n\in\mathbb{N} \mid |S_n| \geq c\sqrt{n\log(n)}\}$ (Proposition 5).

Theorem 2 is not valid when $\delta \leq \sigma\sqrt{1+a}$. In fact, for δ approaching $\sigma\sqrt{1+a}$ rather precise results are proven by Spataru [62]. He shows (in our notation) that if $(X_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with $E[X_1]=0$, $E[X_1^2]=\sigma^2>0$, and $E[|X|^{2(a+2)}(\log^+|X|)^{-(a+2)}] < \infty$, then

$$\lim_{\delta\downarrow\sigma\sqrt{a+1}} \sqrt{\delta^2-\sigma^2(a+1)} \sum_{n\geq 2} n^a P\Big(|S_n| \geq \delta\sqrt{2n\log(n)}\Big) = \sigma\sqrt{\frac{1}{a+1}}, \quad \text{for all } -1 < a < -1/2.$$

Our proof of Proposition 4 can therefore not easily be extended to all $c_\rho^* > 0$. It is possible to explicitly calculate the value of c_ρ^*, although the calculation is somewhat tedious.

5.7. Application to adaptive design of experiments. In §3 we combine the Sherman-Morrison formula with the fact that $\lambda_{\min}(P(t))$ grows proportional to $\mathrm{tr}(P(t)^{-1})^{-1}$ and show that a minimal growth rate on $\lambda_{\min}(P(t))$ can be achieved by requiring a simple quadratic constraint on the control variable. This idea is related to E-optimal designs in the area of design of experiments (DoE) that aim at maximizing the smallest eigenvalue of the design matrix. In the DoE literature one typically aims at minimizing the expected squared estimation error after all experiments have been deployed; a difference in our dynamic pricing setting is that the costs incurred by the decision maker are determined by the cumulative expected square estimation errors over the whole time horizon. Our methodology may find application in several DoE problems, for example to construct adaptive E-optimal designs in nonlinear regression settings; (see Pronzato [58, Section 4] or Pronzato [59]).

5.8. Regret bounds when the optimal price is not admissible. Because we assume that $\mathbf{p}(\beta^{(0)}) \in \mathrm{int}(\mathcal{P})$ and $H(\mathbf{p},\beta^{(0)})$ is negative definite at $\mathbf{p}(\beta^{(0)})$, the instantaneous expected regret in a single period is quadratic in the deviation from the optimal price: $r(\mathbf{p}_{\text{opt}},\beta^{(0)}) - r(\mathbf{p},\beta^{(0)}) = O(\|\mathbf{p}_{\text{opt}}-\mathbf{p}\|^2)$;see Equation (37). This relation may fail to hold if $\mathbf{p}(\beta^{(0)})$ lies outside $\mathcal{P}$. Two cases can be distinguished:

(i) $\mathbf{p}(\beta) = \mathbf{p}(\beta^{(0)})$ for all β in an open neighborhood of $\beta^{(0)}$.

(ii) For any open neighborhood U of $\beta^{(0)}$, there is a $\beta \in U$ with $\mathbf{p}(\beta) \neq \mathbf{p}(\beta^{(0)})$.

Case (i) may occur, for example, when $\mathcal{P} = \{1\}\times[p_l,p_h]^n$ for some $0<p_l<p_h$ and

$$\arg\max_{\mathbf{p}\in\{1\}\times\mathbb{R}^n} r(\mathbf{p},\beta^{(0)}) \in \{1\}\times(p_h,\infty)^n.$$

In this case $\mathbf{p}(\beta^{(0)}) = (1,p_h,\ldots,p_h)$, and by continuity arguments $\mathbf{p}(\beta) = \mathbf{p}(\beta^{(0)})$ for all β in an open neighborhood of $\beta^{(0)}$. The terms $\|\mathbf{p}(\hat\beta(t-1) - \mathbf{p}_{\text{opt}}\|^2 \mathbf{1}_{t>T_2}$ in the proof of Theorem 1 vanish if ρ is chosen sufficiently small, resulting in $\mathrm{Regret}(\Phi_{L_1},T) = O(L_1(T))$. The requirement that $L_1(t)$ grows faster than $\sqrt{t}$ is still necessary to guarantee strong consistency in Proposition 3, and thus we get $\mathrm{Regret}(\Phi_{L_1},T) = O(T^{1/2+\delta})$ when $L_1(t) = ct^{1/2+\delta}$, for some $c>0$ and arbitrarily small $\delta>0$.

Case (ii) may occur for example when $n=2$, $\mathcal{P} = \{1\}\times[p_l,p_h]^2$ for some $0<p_l<p_h$, h_1 and h_2 are the identity function, and

$$\arg\max_{\mathbf{p}\in\{1\}\times\mathbb{R}^2} r(\mathbf{p},\beta^{(0)}) \in \{1\}\times(p_l,p_h)\times(p_h,\infty).$$

In this case $r(\mathbf{p}_{\text{opt}}, \beta^{(0)}) - r(\mathbf{p}, \beta^{(0)}) = O(\|\mathbf{p}_{\text{opt}} - \mathbf{p}\|)$, and $r(\mathbf{p}_{\text{opt}}, \beta^{(0)}) - r(\mathbf{p}, \beta^{(0)}) \neq O(\|\mathbf{p}_{\text{opt}} - \mathbf{p}\|^{\zeta})$ for all $\zeta > 1$.

Suppose Φ_{L_1} is used. Then by slightly modifying the proof of Theorem 1, we obtain

$$\begin{aligned} E\bigg[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\text{opt}}\|\bigg] &\leq E\bigg[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}(\hat\beta(t-1))\|\mathbf{1}_{t>T_2}\bigg] + E\bigg[\sum_{t=t_0}^{T}\|\mathbf{p}(\hat\beta(t-1))-\mathbf{p}_{\text{opt}}\|\mathbf{1}_{t>T_2}\bigg] + E\bigg[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\text{opt}}\|\mathbf{1}_{t\leq T_2}\bigg] \\ &= O\bigg(\sum_{t=t_0}^{T} E\big[\sqrt{\dot L_1(t)}\mathbf{1}_{t>T_2}\big] + \sum_{t=t_0}^{T} E\big[\|\hat\beta(t-1)-\beta^{(0)}\|\mathbf{1}_{t>T_2}\big] + \sum_{t=t_0}^{T} P(t\leq T_2)\bigg) \\ &= O\bigg(\sum_{t=t_0}^{T}\sqrt{\dot L_1(t)} + \sum_{t=t_0}^{T}\sqrt{L_1(t)^{-1}\log(t)+tL_1(t)^{-2}} + \sum_{t=t_0}^{T} E[T_2^{1/2}]t^{-1/2}\bigg), \end{aligned}$$

using

$$\begin{aligned} E[\|\hat\beta(t-1)-\beta^{(0)}\|\mathbf{1}_{t>T_2}] &= E\Big[\sqrt{\|\hat\beta(t-1)-\beta^{(0)}\|^2\mathbf{1}_{t>T_2}}\Big] \\ &\leq \sqrt{E[\|\hat\beta(t-1)-\beta^{(0)}\|^2\mathbf{1}_{t>T_2}]} = O\big(\sqrt{L_1(t)^{-1}\log(t)+tL_1(t)^{-2}}\big) \end{aligned}$$

and

$$\|\mathbf{p}(t)-\mathbf{p}(\hat\beta(t-1))\|^2\mathbf{1}_{t>T_2} = O(\dot L_1(t));$$

see Equation (30). If $L_1(t) = ct^{\alpha}$ for some $c > 0$, $\alpha \in (1/2, 1)$, we obtain $\text{Regret}(\Phi_{L_1}, T) = O(T^{(\alpha+1)/2} + T^{3/2-\alpha})$, and in this case the optimal choice of α equals 2/3, with corresponding $\text{Regret}(\Phi_{L_1}, T) = O(T^{5/6})$. For canonical link functions, the choice $L_1(t) = ct^{\alpha}$ leads to $\text{Regret}(\Phi_{L_1}, T) = O(T^{(\alpha+1)/2} + T^{1-\alpha/2}\sqrt{\log(T)})$; this bound is minimized by choosing $\alpha = 1/2 + \delta$ for $\delta > 0$ arbitrarily small, in which case $\text{Regret}(\Phi_{L_1}, T) = O(T^{3/4+\delta/2})$.

These two examples show that the regret behaves quite differently under case (i) and (ii). This is, of course, because in case (i) the value of $\beta^{(0)}$ does not have to be learned exactly: it suffices to have $\hat\beta(t)$ sufficiently close to $\beta^{(0)}$. Also observe that (ii) cannot occur in the single-product case, indicating a qualitative difference between single-product and multi-product pricing when $\mathbf{p}(\beta^{(0)}) \notin \mathcal{P}$. An interesting direction for future research is to derive lower bounds on the regret that *any* pricing policy must incur when $\mathbf{p}(\beta^{(0)}) \notin \mathcal{P}$. It has been shown in various that there is no pricing policy with $\text{Regret}(T) = o(\sqrt{T})$ when $\mathbf{p}(\beta^{(0)}) \in \text{int}(\mathcal{P})$; see Kleinberg and Leighton [46], Besbes and Zeevi [13], and Broder and Rusmevichientong [15]. It would be interesting to derive analogous results for the case $\mathbf{p}(\beta^{(0)}) \notin \mathcal{P}$.

Of course, in practical applications price managers would probably reconsider their choice of $\mathcal{P}$ if there is strong statistical evidence that $\mathbf{p}(\beta^{(0)})$ lies outside $\mathcal{P}$.

6. Numerical illustration. In this section we provide two numerical illustrations of the proposed adaptive pricing policy Φ_{L_1}. The first considers two products with Poisson distributed demand and noncanonical, linear link functions. The second instance shows that our pricing policy Φ_{L_1} can handle large instances: we consider 10 products, with normally distributed demand and canonical link functions.

6.1. Two products, Poisson distributed demand. Consider two products with Poisson distributed demand, with expectation

$$\begin{aligned} E[D_1(p_1,p_2)] &= 11.5 - 1.25p_1 + 0.34p_2, \\ E[D_2(p_1,p_2)] &= 10.22 + 0.25p_1 - 1.55p_2. \end{aligned}$$

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_{L_1} with $L_1(t) = 0.2\cdot t^{2/3}$ (note that the link functions are not canonical).

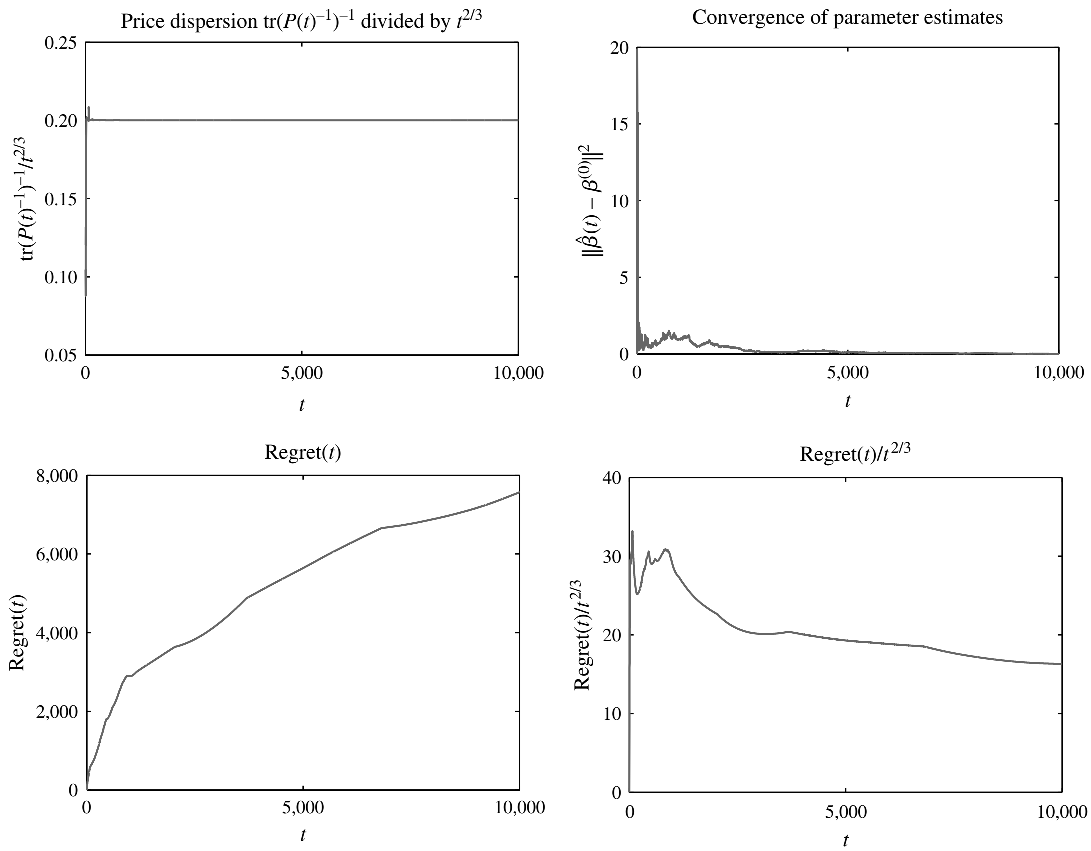

FIGURE 1. Numerical results for §6.1.

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

6.2. Ten products, normally distributed demand. We here consider a large instance with 10 products. The demand for each product k is normally distributed with expectation and variance given by

$$E[D_k(\mathbf{p})] = \beta_{k0}^{(0)} + \beta_{k1}^{(0)} p_1 + \cdots + \beta_{kn}^{(0)} p_n, \quad (k = 1, \ldots, n),$$
$$\mathrm{Var}[D_k(\mathbf{p})] = \sigma_k^2, \quad (k = 1, \ldots, n),$$

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

$$\mathbf{p}(1) = \begin{pmatrix} 1 \\ 18.59 \\ 1.81 \\ 13.09 \\ 6.11 \\ 19.32 \\ 4.23 \\ 10.65 \\ 13.27 \\ 15.64 \\ 1.76 \end{pmatrix}, \quad \mathbf{p}(2) = \begin{pmatrix} 1 \\ 4.48 \\ 1.33 \\ 5.34 \\ 9.26 \\ 10.75 \\ 14.18 \\ 1.23 \\ 14.06 \\ 18.87 \\ 8.36 \end{pmatrix}, \quad \mathbf{p}(3) = \begin{pmatrix} 1 \\ 19.04 \\ 18.34 \\ 19.61 \\ 18.98 \\ 11.24 \\ 10.47 \\ 6.34 \\ 14.4 \\ 18.44 \\ 18.63 \end{pmatrix}, \quad \mathbf{p}(4) = \begin{pmatrix} 1 \\ 15.04 \\ 3.17 \\ 14.61 \\ 5.79 \\ 16.51 \\ 17.67 \\ 1.49 \\ 9.14 \\ 17.78 \\ 14.32 \end{pmatrix},$$

$$\mathbf{p}(5) = \begin{pmatrix} 1 \\ 14.3 \\ 4.99 \\ 11.79 \\ 2.33 \\ 3.02 \\ 8.18 \\ 4.65 \\ 7.45 \\ 1.31 \\ 2.81 \end{pmatrix}, \quad \mathbf{p}(6) = \begin{pmatrix} 1 \\ 9.7 \\ 7.76 \\ 4.82 \\ 5.46 \\ 11.88 \\ 16.83 \\ 17.51 \\ 2.94 \\ 10.28 \\ 5.81 \end{pmatrix}, \quad \mathbf{p}(7) = \begin{pmatrix} 1 \\ 13.06 \\ 1.47 \\ 2.86 \\ 12.06 \\ 16.61 \\ 5.18 \\ 10.57 \\ 4.46 \\ 5.67 \\ 6.66 \end{pmatrix}, \quad \mathbf{p}(8) = \begin{pmatrix} 1 \\ 19.74 \\ 6.61 \\ 2.92 \\ 16.96 \\ 17.55 \\ 16.34 \\ 19.51 \\ 14.3 \\ 19.51 \\ 10.18 \end{pmatrix},$$

$$\mathbf{p}(9) = \begin{pmatrix} 1 \\ 9.45 \\ 18.81 \\ 2.26 \\ 2.28 \\ 4.1 \\ 12.21 \\ 1.62 \\ 11.14 \\ 19.42 \\ 10.5 \end{pmatrix}, \quad \mathbf{p}(10) = \begin{pmatrix} 1 \\ 2.1 \\ 17.23 \\ 10.77 \\ 7.21 \\ 10.89 \\ 13.56 \\ 7.34 \\ 11.81 \\ 9.82 \\ 13.74 \end{pmatrix}, \quad \mathbf{p}(11) = \begin{pmatrix} 1 \\ 3.8 \\ 7.1 \\ 3.15 \\ 6.73 \\ 2.26 \\ 9.05 \\ 6.5 \\ 5.31 \\ 12.12 \\ 7.51 \end{pmatrix}.$$

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_{L_1} with $L_1(t) = 0.05\sqrt{t\log(t)}$ (note that, in contrast with §6.1, the link functions are canonical).

The plots in Figure 2 show a sample path of $\operatorname{tr}(P(t)^{-1})^{-1}$ 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 $\operatorname{tr}(P(t)^{-1})^{-1} \geq 0.05\sqrt{t\log(t)}$ for all sufficiently large t, $\lim_{t\to\infty} \|\hat{\beta}(t) - \beta^{(0)}\|^2 = 0$, and Regret$(t) = O(\sqrt{t\log(t)})$.

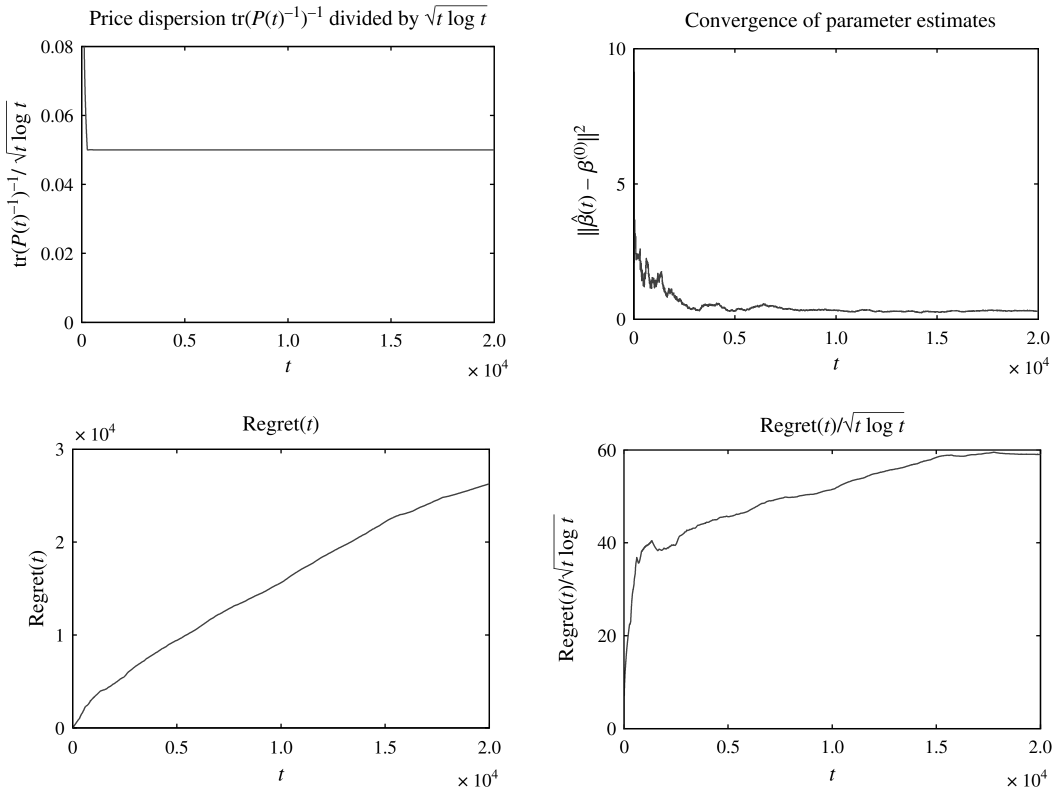

FIGURE 2. Numerical results for §6.2.

7. Proofs.

7.1. Proofs of §3.

Proof of Proposition 1. Let $t > n+1$ and assume (13) and (14). Let $\lambda_1 \geq \cdots \geq \lambda_{n+1} > 0$ 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 $\phi = (\phi_0, \phi_1, \ldots, \phi_n) \in \operatorname{int}(\mathcal{P})$ 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

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

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

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

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,

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

since

$$\lambda_{n+1} \leq (n+1)\operatorname{tr}(P(t)^{-1})^{-1} \leq (n+1)\operatorname{tr}((P(t)+\mathbf{p}(\hat{\beta}(t))\mathbf{p}(\hat{\beta}(t))^T)^{-1})^{-1} < (n+1)L_1(t+1).$$

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

$$\|P(t)^{-1}\mathbf{q}\|^2 \geq \epsilon^2(n+1)^{-2}L_1(t+1)^{-2}. \tag{28}$$

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

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

Proof of Proposition 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 $\operatorname{tr}(P(t+j)^{-1}) \geq L_1(t+j)$, in (I) or until (12) has a feasible solution, in (IIc). By Proposition 1, (IIc) does not occur for $t \geq T_0$.

In addition, since prices are merely repeated in (I), and $\operatorname{tr}(P(ct)^{-1})^{-1} = c\operatorname{tr}(P(t)^{-1})^{-1}$ 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 $\operatorname{tr}(P(T_1)^{-1})^{-1} \geq L_1(T_1)$.

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 $\operatorname{tr}(P(t)^{-1})^{-1} \geq L_1(t)$, then the following holds:

(1) If $\hat{\beta}_k(t)$ does not exist for some k, then

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

and

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

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

$$\operatorname{tr}(P(t+1)^{-1})^{-1} \geq L_1(t+1). \tag{29}$$

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

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

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

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

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

(2) Now suppose that $\text{tr}(P(t)^{-1})^{-1} \geq L_1(t)$, and $\hat{\beta}_k(t)$ does exist for all k. The case (11) is trivial; suppose that (11) does not hold. Then $\mathbf{p}(t+1)$ is determined by (IIb). The Sherman-Morrison formula (Bartlett [8]),

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

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

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

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

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

7.2. Proofs of §4.1.

Proof of Lemma 1. Ad (i) of the lemma follows from Proposition 2 and (9). For (ii), we show

$$\|\mathbf{p}(t) - \mathbf{p}(\hat{\beta}(t-1))\|^2 \mathbf{1}_{t>T_2} \leq K_2 \dot{L}_1(t) \mathbf{1}_{t>T_2} \quad \text{a.s.} \tag{30}$$

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 $\rho \in (0, \rho_0)$, let C_L, T_1 be as in Proposition 2, and set $T_2 = \max\{C_L T_\rho, T_1, T_3, T_4\}$, where T_3, T_4 are nonrandom constants specified below. Clearly, $E[T_2^\eta] < \infty$ if and only if $E[T_\rho^\eta] < \infty$, 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 $E[T_2^{1/2}] < \infty$.

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 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 2, and since $T_2 \geq \max\{C_L T_\rho, T_1\}$, it suffices to show $\text{tr}(P(C_L T_\rho)^{-1})^{-1} \geq L_1(C_L T_\rho)$. This follows since $\text{tr}(P(T_\rho + j)^{-1})^{-1} \geq L_1(T_\rho + j)$ must hold for some $1 \leq j \leq (C_L - 1)T_\rho$; see the proof of Proposition 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

$$l = \inf_{\beta \in V} \inf_{\mathbf{p} \in U_\beta} \lambda_{\min}(\nabla^2 r(\mathbf{p})) < 0, \tag{31}$$

$$L = \sup_{\beta \in V} \sup_{\mathbf{p} \in U_\beta} \lambda_{\max}(\nabla^2 r(\mathbf{p})) < 0; \tag{32}$$

in view of (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 (12), 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

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

and

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

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

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

Assertion (ii) of Lemma 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 (12), such that $\|\mathbf{p}'-\mathbf{p}(\hat{\beta}(t))\|^2 \le K_3\dot{L}_1(t)$ 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 (12).

Let $C_0 > 2$, and

$$T_3 = \sup\left\{t\in\mathbb{N}\;\middle|\;\text{there exists a } \beta\in V \text{ and } \mathbf{p}\in\mathbb{R}^{n+1}\backslash U_\beta \text{ such that } \|\mathbf{p}(\beta)-\mathbf{p}\|^2 \le C_0^2\dot{L}_1(t)\left(1+\max_{\mathbf{p}\in\mathscr{P}}\|\mathbf{p}\|^2\right)\right\},$$

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

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

Let $\lambda_1 \ge \cdots \ge \lambda_{n+1}$ 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

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

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 $\|v_{n+1}\|\le 1$ and $\|\mathbf{p}(\hat{\beta}(t))\|\le \max_{\mathbf{p}\in\mathscr{P}}\|\mathbf{p}\|$,

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

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 (12). Observe that

$$\begin{aligned}\mathbf{p}'^T P(t)^{-1}\mathbf{p}' &\le \lambda_{\max}(P(t)^{-1})\|\mathbf{p}'\|^2 \le \lambda_{\min}(P(t))^{-1}\max_{\mathbf{p}\in\mathscr{P}}\|\mathbf{p}\|^2 \le \operatorname{tr}(P(t)^{-1})\max_{\mathbf{p}\in\mathscr{P}}\|\mathbf{p}\|^2\\ &\le L_1(n+1)^{-1}\max_{\mathbf{p}\in\mathscr{P}}\|\mathbf{p}\|^2 \end{aligned}\tag{33}$$

and

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

$$= (\alpha_{n+1} + \sqrt{\dot{L}_1(t)}Cv_{n+1,1}\alpha_{n+1} - \sqrt{\dot{L}_1(t)}C)^2\lambda_{n+1}^{-2} + \sum_{i=1}^{n}(1+\sqrt{\dot{L}_1(t)}Cv_{n+1,1})^2\alpha_i^2\lambda_i^{-2}$$
$$\geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)}(v_{n+1,1}\alpha_{n+1}-1))^2\lambda_{n+1}^{-2}$$
$$\geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)}(v_{n+1,1}\alpha_{n+1}-1))^2\,\mathrm{tr}(P(t)^{-1})^2;$$

thus,

$$\|P(t)^{-1}\mathbf{p}'\|^2 \geq (\alpha_{n+1} + C\sqrt{\dot{L}_1(t)}(v_{n+1,1}\alpha_{n+1}-1))^2 L_1(t+1)^{-2}. \tag{34}$$

By construction of C,

$$\|P(t)^{-1}\mathbf{p}'\|^2 \geq (\alpha_{n+1}^2 + C^2\dot{L}_1(t)(v_{n+1,1}\alpha_{n+1}-1)^2)L_1(t+1)^{-2}. \tag{35}$$

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

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

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

$$\|P(t)^{-1}\mathbf{p}'\|^2 \geq \alpha_{n+1}^2 L_1(t+1)^{-2} \geq \frac{1}{4}C_0^2\dot{L}_1(t)L_1(t+1)^{-2} \tag{36}$$

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

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

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

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

$$r(\mathbf{p}_{\mathrm{opt}},\beta^{(0)}) - r(\mathbf{p},\beta^{(0)}) \leq \left(\sup_{\mathbf{p}\in\mathscr{P}}|\lambda_{\max}(\nabla^2 r(\mathbf{p},\beta^{(0)})|\right)\cdot\|\mathbf{p}_{\mathrm{opt}}-\mathbf{p}\|^2, \tag{37}$$

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

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 (5) of second partial derivatives with respect to $\mathbf{p}_1,\ldots,\mathbf{p}_n$ 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 [28]) 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,

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

for all $\beta\in V$ and some nonrandom constant $K_1 > 0$.

Choose $\rho\in(0,\rho_0)$ and T_2 as in Lemma 1. Then $\hat{\beta}(t)\in V$ for all $t > T_\rho$, $T_2\geq T_\rho$ a.s., and

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

By application of Lemma 1, we have

$$E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\mathrm{opt}}\|^2\right]$$
$$\leq E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\mathrm{opt}}\|^2\mathbf{1}_{t>T_2}\right] + E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\mathrm{opt}}\|^2\mathbf{1}_{t\leq T_2}\right]$$
$$\leq 2E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}(\hat{\beta}(t-1))\|^2\mathbf{1}_{t>T_2}\right] + 2E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(\hat{\beta}(t-1))-\mathbf{p}_{\mathrm{opt}}\|^2\mathbf{1}_{t>T_2}\right] + E\left[\sum_{t=t_0}^{T}\|\mathbf{p}(t)-\mathbf{p}_{\mathrm{opt}}\|^2\mathbf{1}_{t\leq T_2}\right]$$
$$\leq 2K_2L_1(T) + 2K_1E\left[\sum_{t=t_0}^{T}\|\hat{\beta}(t-1)-\beta^{(0)}\|^2\mathbf{1}_{t>T_2}\right] + \sum_{t=t_0}^{T}\max_{\mathbf{q},\mathbf{q}'\in\mathscr{P}}\|\mathbf{q}-\mathbf{q}'\|^2 P(t\leq T_2)$$
$$= O\left(L_1(T) + \sum_{t=1}^{T}(L_1(t)^{-1}\log(t) + tL_1(t)^{-2} + E[T_2^{1/2}]t^{-1/2})\right).$$

Since $\sum_{t=1}^{T} E[T_2^{1/2}]t^{-1/2} = O(T^{1/2}) = o(L_1(T))$, it follows by (37) that

$$\text{Regret}(\Phi_{L_1}, T) = O\left(L_1(T) + \sum_{t=1}^{T}(L_1(t)^{-1}\log(t) + tL_1(t)^{-2})\right).$$

7.3. Proofs of §4.2.

Proof of Proposition 4. The assertions on T_ρ are proven in den Boer and Zwart [26, Theorem 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

$$\begin{aligned} T_{A[i]} &= \sup\left\{n \geq n_0 \,\middle|\, \rho^{-1}A_n[i] - \frac{1}{d+2d^2}\frac{c_2}{2}L_1(n)\right\}, \\ T_{B[i,j]} &= \sup\left\{n \geq n_0 \,\middle|\, B_n[i,j] - \frac{1}{d+2d^2}\frac{c_2}{2}L_1(n)\right\}, \\ T_{J[i,j]} &= \sup\left\{n \geq n_0 \,\middle|\, \rho J_n[i,j] - \frac{1}{d+2d^2}\frac{c_2}{2}L_1(n)\right\}, \end{aligned} \tag{38}$$

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 certain martingales with square-integrable differences. It is shown that these last-times are a.s. finite and have finite η-th moments, using their Proposition 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 $c > 0$, $\frac{1}{2} < \alpha \leq 1$.

To prove the current proposition, we need to show that the last-times of (38) are a.s. finite and have finite η-th moment, when $L_1(t) = c\sqrt{t\log(t)}$. We do that using Proposition 5, a refinement of Proposition 1 of Boer and Zwart [26], which considers last-times of the form

$$\sup\{n \in \mathbb{N} \mid |S_n| \geq c\sqrt{n\log(n)}\}.$$

Fix $\rho \in (0, \rho_0)$; let $\sigma_C > 0$ be such that

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

and let $c_2 = \inf_{\mathbf{p}\in\mathcal{P},\,\beta\in B_\rho} \dot{h}(\mathbf{p}^T\beta)$ (see the definition of c_2 in Equation (7) of den Boer and Zwart [26]). If we set $c_\rho^* = ((d+2d^2)/(c_2/2)) \cdot 2\sigma_C\sqrt{\eta}$, then our Proposition 5 implies that the last-times

$$\begin{aligned} T_{A[i]} &= \sup\left\{n \geq n_0 \,\middle|\, \rho^{-1}A_n[i] - \frac{1}{d+2d^2}\frac{c_2}{2}c\sqrt{n\log(n)}\right\}, \\ T_{B[i,j]} &= \sup\left\{n \geq n_0 \,\middle|\, B_n[i,j] - \frac{1}{d+2d^2}\frac{c_2}{2}c\sqrt{n\log(n)}\right\}, \\ T_{J[i,j]} &= \sup\left\{n \geq n_0 \,\middle|\, \rho J_n[i,j] - \frac{1}{d+2d^2}\frac{c_2}{2}c\sqrt{n\log(n)}\right\} \end{aligned} \tag{39}$$

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

The asymptotic existence, strong consistency, and mean square bounds for $\hat{\beta}_k(t)$ then follow directly from den Boer and Zwart [26, Theorems 2, 3].

7.4. Proofs of §4.3.

Proof of Proposition 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 $n > n'$,

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

where (1) follows from den Boer and Zwart [26, Lemma 4], and (2) from the definition of n'.

By Chebyshev's and Rosenthal's inequality (see, e.g., Hall and Heyde [37, Theorem 2.12]), there is a $C_2 > 0$ such that for all $k > e$, $c > 0$,

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

Consequently,

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

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

$$\sum_{j\geq\log_2(n)} P(|S_{2^j}| \geq c'\sqrt{2^{j-1}\log(2^{j-1})}) \tag{40}$$

$$= \int_{j\geq\log_2(n)} P\bigg(|S_{2^j}| \geq c'\frac{\sqrt{j-1}}{\sqrt{2j}}\sqrt{2^j\log(2^j)}\bigg)\,dj \tag{41}$$

$$= \int_{k\geq n} P\bigg(|S_k| \geq c'\sqrt{\frac{1}{2} - \frac{\log(2)}{2\log(k)}}\sqrt{k\log(k)}\bigg)\frac{1}{k\log(2)}\,dk \tag{42}$$

$$\leq \sum_{k\geq n} P\bigg(|S_k| \geq c'\sqrt{\frac{1}{2} - \frac{\log(2)}{2\log(n')}}\sqrt{k\log(k)}\bigg)\frac{1}{k\log(2)} \tag{43}$$

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

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

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

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

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

$$\sum_{k\geq 1} k^{\eta-1} P(|S_k| \geq \delta\sqrt{2k\log(k)}) < \infty. \tag{44}$$

We can choose n' sufficiently large such that $\delta > \sigma\sqrt{\eta}$. Then (44) follows from Theorem 2 with $a = \eta - 1$.

Proof of Lemma 2. Apply Lemma 1 of Stout [65] on the sequence $(X_i/(\sigma\sqrt{n}))_{1\leq i\leq n}$, with $l = 0$. Then for all $0 \leq \lambda c \leq 1$,

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

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

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

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

Proof of Theorem 2. The proof of Theorem 2 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 [53, 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 [53, 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 $n \geq n_1$,

$$\begin{aligned} P(S_n > \delta\sqrt{2n\log(n)}) &\leq P\left(S_n - \text{med}(S_n) > \delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n}\right) \\ &\leq 2P(S_n^s > \delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n}) \\ &\leq 2P(S_n^s > \delta_1\sqrt{2n\log(n)}). \end{aligned} \tag{45}$$

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

$$\sum_{n=1}^\infty n^a P(S_n^s > \delta_1\sqrt{2n\log(n)}) \leq \sum_{n=1}^\infty n^a P(S_n^s > \delta_1\sqrt{2n\log(n)}, n > T_{\neq}) \tag{46}$$

$$+ \sum_{n=1}^\infty n^a P(n \leq T_{\neq}). \tag{47}$$

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 with respect to $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$, 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 $\delta_2 > (1-\kappa/2)^{-1/2}\sigma\sqrt{1+a}$, 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 2,

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

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

$$\sum_{n=1}^{\infty} n^a P(\tilde{S}_n^s > \delta_2\sqrt{2n\log(n)}) = O\left(\sum_{n=1}^{\infty} n^{a-\delta_2^2\sigma^{-2}(1-\kappa/2)}\right) < \infty \text{ a.s.} \tag{48}$$

This proves that the right-hand side of (46) is finite a.s.

For the term (47), we have

$$\sum_{n=1}^{\infty} n^a P(T_{\neq} \geq n) = \sum_{n=1}^{\infty} n^a P(\exists\, k \geq n\colon |X_k|/g(k) > 1)$$
$$\leq \sum_{n=1}^{\infty} n^a \sum_{j\geq\lfloor\log_{(2)}(n)\rfloor} P(\exists\, k\colon 2^j \leq k < 2^{j+1}\colon |X_k|/g(k) > 1)$$
$$\leq \sum_{n=1}^{\infty} n^a \sum_{j\geq\lfloor\log_{(2)}(n)\rfloor} P\left(\sup_{1\leq k\leq 2^{j+1}} |X_k| > g(2^{j+1})\right)$$
$$\leq \sum_{n=1}^{\infty} n^a \sum_{j\geq\lfloor\log_{(2)}(n)\rfloor} \frac{C}{g(2^{j+1})^p}, \tag{49}$$

by Doob's inequality for martingales. Furthermore,

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

where we applied a change of variables $k = 2^{j+1}$. Combining this with (49), it follows from the assumption $p > 2(a+2)$ that

$$\sum_{n=1}^{\infty} n^a P(T_{\neq} \geq n) = O\left(\sum_{n=1}^{\infty} n^{a+1-p/2}(\log(n))^{p/2}\right) < \infty. \tag{50}$$

It follows from Equations (45), (46) (47), (48), and (50) 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, (22) follows. For (23), 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 $n \geq n_3$,

$$P\left(\sup_{1\leq i\leq n}|S_i| > \delta\sqrt{2n\log(n)}\right) \leq 2P(|S_n| \geq \delta\sqrt{2n\log(n)} - \sqrt{2\sigma^2 n})$$
$$\leq 2P\left(|S_n| \geq \frac{1}{2}\delta_3\sqrt{2n\log(n)}\right),$$

using Lemma 4 of den Boer and Zwart [26]. Now (23) follows from (22).

Acknowledgments. The author would like to thank Bert Zwart, Rob van der Mei, Kacha Dzhaparidze, Cor Hurkens, and Rudi Pendavingh for useful discussions and remarks, and the anonymous referees and associate editor for several constructive comments and suggestions. This work was written while affiliated with Centrum Wiskunde and Informatica (CWI) Amsterdam, Eindhoven University of Technology, and University of Amsterdam.

References

[1] Abbasi-Yadkori Y, Szepesvári C (2011) Regret bounds for the adaptive control of linear quadratic systems. Kakade S, von Luxburg U, eds. *24th Annual Conf. Learn. Theory* (*COLT*), Vol. 19, 1–26.
[2] Abbasi-Yadkori Y, Pal D, Szepesvári C (2011) Improved algorithms for linear stochastic bandits. Shawe-Taylor J, Zemel RS, Bartlett P, Pereira FCN, Weinberger KQ, eds. *Advances in Neural Information Processing Systems* 24, 2312–2320.
[3] Araman VF, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. *Oper. Res.* 57(5):1169–1188.
[4] Arora S, Hazan E, Kale S (2012) The multiplicative weights update method: A meta-algorithm and applications. *Theory Comput.* 8:121–164.
[5] Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multi-armed bandit problem. *Machine Learn.* 47(2):235–256.
[6] Auer P, Ortner R, Szepesvári C (2007) Improved rates for the stochastic continuum-armed bandit problem. Bshouty N, Gentile C, eds. *Learning Theory*, Lecture Notes in Computer Science, Vol. 4539 (Springer, Berlin), 454–468.
[7] Aviv Y, Pazgal A (2005) A partially observed Markov decision process for dynamic pricing. *Management Sci.* 51(9):1400–1416.
[8] Bartlett MS (1951) An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statist.* 22(1):107–111.
[9] Bartlett PL, Tewari A (2009) REGAL: A regularization based algorithm for reinforcement learning in weakly communicating MDPs. *Proc. Twenty-Fifth Conf. Uncertainty in Artificial Intelligence UAI '09* (AUAI Press, Arlington, VA), 35–42.
[10] Baum LE, Katz M (1965) Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.* 120(1):108–123.
[11] Bertsimas D, Perakis G (2006) Dynamic pricing: A learning approach. *Mathematical and Computational Models for Congestion Charging* (Springer, New York), 45–79.
[12] Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Oper. Res.* 57(6):1407–1420.
[13] Besbes O, Zeevi A (2011) On the minimax complexity of pricing in a changing environment. *Oper. Res.* 59(1):66–79.
[14] Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York).
[15] Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Oper. Res.* 60(4):965–980.
[16] Bubeck S, Cesa-Bianchi N (2012) Regret analysis of stochastic and non-stochastic multi-armed bandit problems. *Foundations Trends Machine Learn.* 58(1):1–122.
[17] Carvalho AX, Puterman ML (2005) Dynamic optimization and learning: How should a manager set prices when the demand function is unknown? Discussion Papers 1117, Instituto de Pesquisa Economica Aplicada—IPEA, Brasilia.
[18] Carvalho AX, Puterman ML (2005) Learning and pricing in an internet environment with binomial demand. *J. Revenue Pricing Management* 3(4):320–336.
[19] Chang YI (1999) Strong consistency of maximum quasi-likelihood estimate in generalized linear models via a last time. *Statist. Probab. Lett.* 45(3):237–246.
[20] Clausen J (1999) Branch and bound algorithms—Principles and examples, http://janders.eecg.toronto.edu/1387/readings/b_and_b.pdf.
[21] Cope E (2007) Bayesian strategies for dynamic pricing in e-commerce. *Naval Res. Logist.* 54(3):265–281.
[22] Cope EW (2009) Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Trans. Automatic Control* 54(6):1243–1253.
[23] Dani V, Hayes TP, Kakade SM (2008) Stochastic linear optimization under bandit feedback. *21st Annual Conf. Learn. Theory (COLT)* (Omnipress, Madison, WI), 355–366.
[24] Davis JA (1968) Convergence rates for probabilities of moderate deviations. *Ann. Math. Statist.* 39(6):2016–2028.
[25] den Boer AV (2013) Dynamic pricing and learning: Historical origins, current research, and new directions. Working paper, http://ssrn.com/abstract=2334429.
[26] den Boer AV, Zwart B (2011) Mean square convergence rates for maximum quasi-likelihood estimators. Submitted for publication, https://www.researchgate.net/publication/257985753_Mean_square_convergence_rates_for_maximum_quasi-likelihood_estimators.
[27] den Boer AV, Zwart B (2013) Simultaneously learning and optimizing using controlled variance pricing. *Management Sci.*, ePub ahead of print December 10, http://dx.doi.org/10.1287/mnsc.2013.1788.
[28] Duistermaat JJ, Kolk JAC (2004) *Multidimensional Real Analysis: Differentiation*, Cambridge Studies in Advanced Mathematics (86) (Cambridge University Press, Cambridge, UK).
[29] Erdős P (1949) On a theorem of Hsu and Robbins. *Ann. Math. Statist.* 20(2):286–291.
[30] Erdős P (1950) Remark on my paper "On a theorem of Hsu and Robbins." *Ann. Math. Statist.* 21(1):138.
[31] Eren SS, Maglaras C (2010) Monopoly pricing with limited demand information. *J. Revenue Pricing Management* 9:23–48.
[32] Farias VF, van Roy B (2010) Dynamic pricing with a prior on market response. *Oper. Res.* 58(1):16–29.
[33] Filippi S, Cappe O, Garivier A, Szepesvari C (2010) Parametric bandits: The generalized linear case. Lafferty J, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, eds. *Advances in Neural Information Processing Systems* 23 (Curran Associates, Red Hook, NY), 586–594.
[34] Gill J (2001) *Generalized Linear Models: A Unified Approach* (Sage Publications, Thousand Oaks, CA).
[35] Godambe VP, Heyde CC (1987) Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* 55(3):231–244.
[36] Hager WW (1989) Updating the inverse of a matrix. *SIAM Rev.* 31(2):221–239.
[37] Hall P, Heyde CC (1980) *Martingale Limit Theory and Its Application*, Probability and Mathematical Statistics (Academic Press, New York).
[38] Harrison JM, Keskin NB, Zeevi A (2012) Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Sci.* 58(3):570–586.
[39] Heyde CC (1997) *Quasi-Likelihood and Its Application*, Springer Series in Statistics (Springer Verlag, New York).

[40] Heyde CC, Morton R (1996) Quasi-likelihood and generalizing the EM algorithm. *J. Roy. Statist. Soc. Ser. B* (*Methodological*) 58(2):317–327.
[41] Hsu PL, Robbins H (1947) Complete convergence and the law of large numbers. *Proc. Natl. Acad. Sci. USA* 33(2):25–31.
[42] Jaksch T, Ortner R, Auer P (2010) Near-optimal regret bounds for reinforcement learning. *J. Machine Learn. Res.* 11:1563–1600.
[43] Katz ML (1963) The probability in the tail of a distribution. *Ann. Math. Statist.* 34(1):312–318.
[44] Keskin NB, Zeevi A (2013) Dynamic pricing with an unknown linear demand model: Asymptotically optimal semi-myopic policies. Working paper, University of Chicago, Chicago, http://faculty.chicagobooth.edu/bora.keskin/pdfs/DynamicPricingUnknownDemandModel.pdf.
[45] Kleinberg R (2005) Nearly tight bounds for the continuum-armed bandit problem. Saul LK, Weiss Y, Bottou L, eds. *Advances in Neural Information Processing Systems* 17 (MIT Press, Cambridge, MA), 697–704.
[46] Kleinberg R, Leighton T (2003) The value of knowing a demand curve: Bounds on regret for online posted-price auctions. *Proc. 44th IEEE Sympos. Foundations of Comput. Sci.* (IEEE Computer Society, Washington, DC), 594–605.
[47] Lai TL (1974) Limit theorems for delayed sums. *Ann. Probab.* 2(3):432–440.
[48] Lai TL, Wei CZ (1982) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* 10(1):154–166.
[49] Le Guen T (2008) Data-driven pricing. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA. Retrieved April 4, 2011, http://hdl.handle.net/1721.1/45627.
[50] Lim AEB, Shanthikumar JG (2007) Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res.* 55(2):198–214.
[51] Lim AEB, Shanthikumar JG, Watewai T (2008) Robust multi-product pricing. Working paper, University of California, Berkeley, http://dx.doi.org/10.2139/ssrn.1078012.
[52] Lobo MS, Boyd S (2003) Pricing and learning with uncertain demand. Working paper, University of Stanford, Stanford, http://www.stanford.edu/~boyd/papers/pdf/pric_learn_unc_dem.pdf.
[53] Loève M (1977) *Probability Theory I*, 4th ed. (Springer Verlag, New York).
[54] McCullagh P (1983) Quasi-likelihood functions. *Ann. Statist.* 11(1):59–67.
[55] McCullagh P, Nelder JA (1983) *Generalized Linear Models* (Chapman & Hall, London).
[56] Nassiri-Toussi K, Ren W (1994) On the convergence of least squares estimates in white noise. *IEEE Trans. Automatic Control* 39(2):364–368.
[57] Osborne MR (1992) Fisher's method of scoring. *Internat. Statist. Rev./Revue Internationale de Statistique* 60(1):99–117.
[58] Pronzato L (2008) Optimal experimental design and some related control problems. *Automatica* 44(2):303–325.
[59] Pronzato L (2009) Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statist. Probab. Lett.* 79(21):2307–2313.
[60] Rothschild M (1974) A two-armed bandit theory of market pricing. *J. Econom. Theory* 9(2):185–202.
[61] Rusmevichientong P, Tsitsiklis JN (2010) Linearly parameterized bandits. *Math. Oper. Res.* 35(2):395–411.
[62] Spataru A (2004) Precise asymptotics for a series of T.L. Lai. *Proc. Amer. Math. Soc.* 132(11):3387–3395.
[63] Spataru A (2009) Improved convergence rates for tail probabilities. *Bull. Transilvania University of Braşov—Ser. III: Math., Informatics, Phys.* 2(51):137–142.
[64] Stoica G (2007) Baum-Katz-Nagaev type results for martingales. *J. Math. Anal. Appl.* 336(2):1489–1492.
[65] Stout WF (1970) A martingale analogue of Kolmogorov's law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 15(4):279–290.
[66] Wang Y, Audibert JY, Munos R (2009) Algorithms for infinitely many-armed bandits. Koller D, Schuurmans D, Bengio Y, Bottou L, eds. *Advances in Neural Information Processing Systems* 21 (Curran Associates, Red Hook, NY), 1729–1736.
[67] Wedderburn RWM (1974) Quasi-likelihood functions, generalized, linear models, and the Gauss-Newton method. *Biometrika* 61(3):439–447.
[68] Xia CH, Dube P (2007) Dynamic pricing in e-services under demand uncertainty. *Production Oper. Management* 16(6):701–712.
[69] Yin C, Zhang H, Zhao L (2008) Rate of strong consistency of maximum quasi-likelihood estimator in multivariate generalized linear models. *Comm. Statist.—Theory Methods* 37(19):3115–3123.
[70] Yu JY, Mannor S (2011) Unimodal bandits. Getoor L, Scheffer T, eds. *Proc. 28th Internat. Conf. Machine Learn., Bellevue, Washington*, (Omnipress, Madison, WI), 41–48.