Data-Driven Pricing of Demand Response

Kia Khezeli Eilyan Bitar

***Abstract*—We consider the setting in which an electric power utility seeks to curtail its peak electricity demand by offering a fixed group of customers a uniform price for reductions in consumption relative to their predetermined baselines. The underlying demand curve, which describes the aggregate reduction in consumption in response to the offered price, is assumed to be affine and subject to unobservable random shocks. Assuming that both the parameters of the demand curve and the distribution of the random shocks are initially unknown to the utility, we investigate the extent to which the utility might dynamically adjust its offered prices to maximize its cumulative *risk-sensitive payoff* over a finite number of T days. In order to do so effectively, the utility must design its pricing policy to balance the tradeoff between the need to learn the unknown demand model (exploration) and maximize its payoff (exploitation) over time. In this paper, we propose such a pricing policy, which is shown to exhibit an expected payoff loss over T days that is at most $O(\sqrt{T})$, relative to an oracle who knows the underlying demand model. Moreover, the proposed pricing policy is shown to yield a sequence of prices that converge to the oracle optimal prices in the mean square sense.**

I. Introduction

The ability to implement residential *demand response* (DR) programs at scale has the potential to substantially improve the efficiency and reliability of electric power systems. In the following paper, we consider a class of DR programs in which an electric power utility seeks to elicit a reduction in the aggregate electricity demand of a fixed group of customers, during peak demand periods. The class of DR programs we consider rely on non-discriminatory, price-based incentives for demand reduction. That is to say, each participating customer is remunerated for her reduction in electricity demand according to a uniform price determined by the utility.

There are several challenges a utility faces in implementing such programs, the most basic of which is the prediction of how customers will adjust their aggregate demand in response to different prices – the so-called aggregate demand curve. The extent to which customers are willing to forego consumption, in exchange for monetary compensation, is contingent on variety of idiosyncratic and stochastic factors – the majority of which are initially unknown or not directly measurable by the utility. The utility must, therefore, endeavor to learn the behavior of customers over time through observation of aggregate demand reductions in response to its offered prices for DR. At the same time, the utility must set its prices for DR in such a manner as to promote increased earnings over time. As we will later establish, such tasks are inextricably linked, and give rise to a trade-off between *learning* (exploration) and *earning* (exploitation) in pricing demand response over time.

Contribution and Related Work: We consider the setting in which the electric power utility is faced with a demand curve that is affine in price, and subject to unobservable, additive random shocks. Assuming that both the parameters of the demand curve and the distribution of the random shocks are initially unknown to the utility, we investigate the extent to which the utility might dynamically adjust its offered prices for demand curtailment to maximize its cumulative risk-sensitive payoff over a finite number of T days. We define the utility's payoff on any given day as the largest return the utility is guaranteed to receive with probability no less than $1-\alpha$. Here, $\alpha \in (0,1)$ encodes the utility's sensitivity to risk. In this paper, we propose a causal pricing policy, which resolves the tradeoff between the utility's need to learn the underlying demand model and maximize its cumulative risk-sensitive payoff over time. More specifically, the proposed pricing policy is shown to exhibit an expected payoff loss over T days – relative to an oracle that knows the underlying demand model – which is at most $O(\sqrt{T})$. Moreover, the proposed pricing policy is shown to yield a sequence of offered prices, which converges to the sequence of oracle optimal prices in the mean square sense.

There is a related stream of literature in operations research [1]–[4], which considers a similar setting in which a monopolist endeavors to sell a product over multiple time periods – with the aim of maximizing its cumulative expected revenue – when the underlying demand curve (for that product) is unknown and subject to exogenous shocks. What distinguishes our formulation from this prevailing literature is the explicit treatment of risk-sensitivity in the optimization criterion we consider, and the subsequent need to design pricing policies that not only learn the underlying demand curve, but also learn the shock distribution.

Focusing explicitly on demand response applications, there are several related papers in the literature, which formulate the problem of eliciting demand response under uncertainty within the framework of multi-armed bandits [5]–[8]. In this setting, each arm represents a customer or a class of customers. Taylor and Mathieu [5] show that, in the absence of exogenous shocks on load curtailment, the optimal policy is indexable. Kalathil and Rajagopal [6] consider a similar multi-armed bandit setting in which a customer's load curtailment is subject to an exogenous shock, and attenuation due to fatigue resulting from repeated requests for reduction in demand over time. They propose a policy, which ensures that the T-period regret is bounded from above by $O(\sqrt{T \log T})$. There is a related stream of literature, which treats the problem of pricing demand response under uncertainty using techniques from online learning [9]–[12]. Perhaps closest to the setting considered

Supported in part by NSF grants ECCS-1351621, CNS-1239178, IIP-1632124, US DoE under the CERTS initiative, and the Atkinson Center for a Sustainable Future.

Kia Khezeli and Eilyan Bitar are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 14853, USA. Emails: {kk839, eyb5}@cornell.edu

in this paper, Jia et al. [10] consider the problem of pricing demand response when the underlying demand function is unknown, affine, and subject to normally distributed random shocks. With the aim of maximizing the utility's expected surplus, they propose a stochastic approximation-based pricing policy, and establish an upper bound on the T-period regret that is $O(\log T)$. There is another stream of literature, which considers an auction-based approach to the procurement of demand response [13]–[19]. In such settings the primary instrument for analysis is game-theoretic in nature.

Organization: The rest of the paper is organized as follows. In Section II, we develop the demand model and formulate the utility's pricing problem for demand response. In Section III, we outline a scheme for demand model learning. In Section IV, we propose a pricing policy and analyze its performance according to the T-period regret. Finally, Section VI concludes the paper. All mathematical proofs are omitted due to space constraints. They can be found in [20].

II. Model

A. Responsive Demand Model

We consider a class of demand response (DR) programs in which an electric power utility seeks to elicit a reduction in peak electricity demand from a fixed group of N customers over multiple time periods (e.g., days) indexed by $t = 1, 2, \ldots$. The class of DR programs we consider rely on uniform price-based incentives for demand reduction. Specifically, prior to each time period t, the utility broadcasts a single price $p_t \geq 0$ (\$/kWh), to which each participating customer i responds with a reduction in demand D_{it} (kWh) – thus entitling customer i to receive a payment in the amount of $p_t D_{it}$.[1] We model the response of each customer i to the posted price p_t at time t according to a linear demand function given by

$$D_{it} = a_i p_t + b_i + \varepsilon_{it}, \quad \text{for} \quad i = 1, \ldots, N$$

where $a_i \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are model parameters *unknown to the utility*, and ε_{it} is an unobservable demand shock, which we model as a random variable. *Its distribution is also unknown to the utility*. We define the aggregate response of customers at time t as $D_t := \sum_{i=1}^{N} D_{it}$, which satisfies

$$D_t = a p_t + b + \varepsilon_t, \tag{1}$$

where the aggregate model parameters and shock are defined as $a := \sum_{i=1}^{N} a_i$, $b := \sum_{i=1}^{N} b_i$, and $\varepsilon_t = \sum_{i=1}^{N} \varepsilon_{it}$. To simplify notation in the sequel, we write the deterministic component of aggregate demand as $\lambda(p, \theta) := ap + b$, where $\theta := (a, b)$ denotes the aggregate demand parameters.

We assume throughout the paper that $a \in [\underline{a}, \overline{a}]$ and $b \in [0, \overline{b}]$, where the model parameter bounds are assumed to be known and satisfy $0 < \underline{a} \leq \overline{a} < \infty$ and $0 \leq \overline{b}$. Such assumptions are natural, as they ensure that the price elasticity of aggregate demand is strictly positive and bounded, and that reductions in aggregate demand are guaranteed to be nonnegative in the absence of demand shocks. We also assume that the sequence of shocks $\{\varepsilon_t\}$ are independent and identically distributed random variables, in addition to the following technical assumption.

Assumption 1. The aggregate demand shock ε_t has a bounded range $[\underline{\varepsilon}, \overline{\varepsilon}]$, and a cumulative distribution function F, which is bi-Lipschitz over this range. Namely, there exists a real constant $L \geq 1$, such that for all $x, y \in [\underline{\varepsilon}, \overline{\varepsilon}]$, it holds that

$$\frac{1}{L} |x - y| \leq |F(x) - F(y)| \leq L |x - y| .$$

There is a large family of distributions respecting Assumption 1 including uniform and doubly truncated normal distributions. Moreover, the assumption that the aggregate demand shock takes bounded values is natural, given the inherent physical limitation on the range of values that demand can take. And, technically speaking, the requirement that F be bi-Lipschitz is stated to ensure Lipschitz continuity of its inverse, which will prove critical to the derivation of our main results. Finally, we note that the utility need not know the parameters specified in Assumption 1.

B. Utility Model and Pricing Policies

We consider a setting in which the utility seeks to reduce its peak electricity demand over multiple days, indexed by t. Accordingly, we let c_t (\$/kWh) denote the wholesale price of electricity during peak demand hours on day t. And, we assume that c_t is known to the utility prior to its determination of the DR price p_t in each period t. Upon broadcasting a price p_t to its customer base, and realizing an aggregate demand reduction D_t, the utility derives a net reduction in its peak electricity cost in the amount of $(c_t - p_t)D_t$. Henceforth, we will refer to the net savings $(c_t - p_t)D_t$ as the *revenue* derived by the utility in period t.

The utility is assumed to be *sensitive to risk*, in that it would like to set the price for DR in each period t to maximize the revenue it is guaranteed to receive with probability no less than $1 - \alpha$. Clearly, the parameter $\alpha \in (0, 1)$ encodes the degree to which the utility is sensitive to risk. Accordingly, we define the *risk-sensitive revenue* derived by the utility in period t given a posted price p_t as

$$r_\alpha(p_t) = \sup \{x \in \mathbb{R} : \mathbb{P}\{(c_t - p_t)D_t \geq x\} \geq 1 - \alpha\} . \tag{2}$$

The risk measure specified in (2) is closely related to the standard concept of *value at risk* commonly used in mathematical finance. Conditioned on a fixed price p_t, one can reformulate the expression in (2) as

$$r_\alpha(p_t) = (c_t - p_t)(\lambda(p_t, \theta) + F^{-1}(\alpha)), \tag{3}$$

where $F^{-1}(\alpha) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$ denotes the α-quantile of the random variable ε_t. It is immediate to see from the simplified expression in (3) that $r_\alpha(p_t)$ is strictly concave in p_t. Let p_t^* denote the *optimal price*, which maximizes the risk-sensitive revenue in period t. Namely,

$$p_t^* := \arg\max\{r_\alpha(p_t) : p_t \in [0, c_t]\}.$$

[1] A customer's reduction in demand is measured relative to a predetermined baseline. The question as to how such a baseline is calculated is beyond the scope of this paper, and is left as a direction for future research.

Its explicit solution is readily derived from the corresponding first order optimality condition, and is given by

$$p_t^* = \frac{c_t}{2} - \frac{b + F^{-1}(\alpha)}{2a}.$$

We define the *oracle risk-sensitive revenue* accumulated over T time periods as

$$R^*(T) := \sum_{t=1}^{T} r_\alpha(p_t^*).$$

The term oracle is used, as $R^*(T)$ equals the maximum risk-sensitive revenue achievable by the utility over T periods if it were to have *perfect knowledge* of the demand model.

In the setting considered in this paper, we assume that both the demand model parameters $\theta = (a, b)$ and the shock distribution F are *unknown to the utility* at the outset. As a result, the utility must attempt to learn them over time by observing aggregate demand reductions in response to offered prices. Namely, the utility must endeavor to learn the demand model, while simultaneously trying to maximize its risk-sensitive returns over time. As we will later see, such task will naturally give rise to a trade-off between *learning* (exploration) and *earning* (exploitation) in pricing demand response over time. First, we describe the space of feasible pricing policies.

We assume that, prior to its determination of the DR price in period t, the utility has access to the entire history of prices and demand reductions until period $t-1$. We, therefore, define a *feasible pricing policy* as an infinite sequence of functions $\pi = (p_1, p_2, \dots)$, where each function in the sequence is allowed to depend only on the past history. More precisely, we require that the function p_t be measurable according to the σ-algebra generated by the history of past decisions and demand observations $(p_1, \dots, p_{t-1}, D_1, \dots, D_{t-1})$ for all $t \geq 2$, and that p_1 be a constant function. The *expected risk-sensitive revenue* generated by a feasible pricing policy π over T time periods is defined as

$$R^\pi(T) := \mathbb{E}^\pi \left[\sum_{t=1}^{T} r_\alpha(p_t) \right],$$

where expectation is taken with respect to the demand model (1) under the pricing policy π.

C. Performance Metric

We evaluate the performance of a feasible pricing policy π according to the T-period *regret*, which we define as

$$\Delta^\pi(T) := R^*(T) - R^\pi(T).$$

Naturally, pricing policies yielding a smaller regret are preferred, as the oracle risk-sensitive revenue $R^*(T)$ stands as an upper bound on the expected risk-sensitive revenue $R^\pi(T)$ achievable by any feasible pricing policy π. Ultimately, we seek a pricing policy whose T-period regret is sublinear in the horizon T. Such a pricing policy is said to have *no-regret*.

Definition 1 (No-Regret Pricing)**.** A feasible pricing policy π is said to exhibit *no-regret* if $\lim_{T\to\infty} \Delta^\pi(T)/T = 0$.

III. Demand Model Learning

Clearly, the ability to price with no-regret will rely centrally on the rate at which the unknown parameters, θ, and quantile function, $F^{-1}(\alpha)$, can be learned from the market data. In what follows, we describe a basic approach to model learning built on the method of least squares estimation.

A. Parameter Estimation

Given the history of past decisions and demand observations $(p_1, \dots, p_t, D_1, \dots, D_t)$ through period t, define the *least squares estimator* (LSE) of θ as

$$\theta_t := \arg\min \left\{ \sum_{k=1}^{t} (D_k - \lambda(p_k, \vartheta))^2 \ : \ \vartheta \in \mathbb{R}^2 \right\},$$

for time periods $t = 1, 2, \dots$. The LSE at period t admits an explicit expression of the form

$$\theta_t = \left(\sum_{k=1}^{t} \begin{bmatrix} p_k \\ 1 \end{bmatrix} \begin{bmatrix} p_k \\ 1 \end{bmatrix}^\top \right)^{-1} \left(\sum_{k=1}^{t} \begin{bmatrix} p_k \\ 1 \end{bmatrix} D_k \right), \qquad (4)$$

provided the indicated inverse exists. It will be convenient to define the 2×2 matrix

$$\mathscr{J}_t := \sum_{k=1}^{t} \begin{bmatrix} p_k \\ 1 \end{bmatrix} \begin{bmatrix} p_k \\ 1 \end{bmatrix}^\top = \begin{bmatrix} \sum_{k=1}^t p_k^2 & \sum_{k=1}^t p_k \\ \sum_{k=1}^t p_k & t \end{bmatrix}.$$

Utilizing the definition of the aggregate demand model (1), in combination with the expression in (4), one can obtain the following expression for the parameter estimation error:

$$\theta_t - \theta = \mathscr{J}_t^{-1} \left(\sum_{k=1}^{t} \begin{bmatrix} p_k \\ 1 \end{bmatrix} \varepsilon_k \right). \qquad (5)$$

Remark 1 (The Role of Price Dispersion)**.** The expression for the parameter estimation error in (5) reveals how consistency of the LSE is reliant upon the asymptotic spectrum of the matrix $\mathscr{J}_t$. Namely, the minimum eigenvalue of $\mathscr{J}_t$, must grow unbounded with time, in order that the parameter estimation error converge to zero in probability. In [3, Lemma 2], the authors establish a sufficient condition for such growth. Specifically, they prove that the minimum eigenvalue of $\mathscr{J}_t$ is bounded from below (up to a multiplicative constant) by the *sum of squared price deviations* defined as $J_t := \sum_{k=1}^t (p_k - \overline{p}_t)^2$, where $\overline{p}_t := (1/t)\sum_{k=1}^t p_k$. The result is reliant on the assumption that the underlying pricing policy π yield a bounded sequence of prices $\{p_t\}$. An important consequence of such a result is that it reveals the explicit role that *price dispersion* (i.e., exploration) plays in facilitating consistent parameter estimation.

Finally, given the underlying assumption that the unknown model parameters θ belong to a compact set defined $\Theta := [\underline{a}, \overline{a}] \times [0, \overline{b}]$, one can improve upon the LSE at time t by projecting it onto the set Θ. Accordingly, we define the *truncated least squares estimator* as

$$\widehat{\theta}_t := \arg\min \{ \|\vartheta - \theta_t\|_2 : \vartheta \in \Theta \} \qquad (6)$$

Clearly, we have that $\|\widehat{\theta}_t - \theta\|_2 \leq \|\theta_t - \theta\|_2$. In the following section, we describe an approach to estimating the underlying quantile function using the parameter estimator defined in (6).

B. Quantile Estimation

Building on the parameter estimator specified in Equation (6), we construct an estimator of the unknown quantile function $F^{-1}(\alpha)$ according to the empirical quantile function associated with the demand estimation residuals. Namely, in each period t, define the sequence of *residuals* associated with the estimator $\widehat{\theta}_t$ as

$$\widehat{\varepsilon}_{k,t} := D_k - \lambda(p_k, \widehat{\theta}_t),$$

for $k = 1, \ldots, t$. Define their *empirical distribution* as

$$\widehat{F}_t(x) := \frac{1}{t} \sum_{k=1}^{t} \mathbb{1}\{\widehat{\varepsilon}_{k,t} \leq x\},$$

and their corresponding *empirical quantile function* as $\widehat{F}_t^{-1}(\alpha) = \inf\{x \in \mathbb{R} : \widehat{F}_t(x) \geq \alpha\}$ for all $\alpha \in (0, 1)$. It will be useful in the sequel to express the empirical quantile function in terms of the order statistics associated with sequence of residuals. Essentially, the *order statistics* $\widehat{\varepsilon}_{(1),t}, \ldots, \widehat{\varepsilon}_{(t),t}$ are defined as a permutation of $\widehat{\varepsilon}_{1,t}, \ldots, \widehat{\varepsilon}_{t,t}$ such that $\widehat{\varepsilon}_{(1),t} \leq \widehat{\varepsilon}_{(2),t} \leq \cdots \leq \widehat{\varepsilon}_{(t),t}$. With this concept in hand, the empirical quantile function can be equivalently expressed as

$$\widehat{F}_t^{-1}(\alpha) = \widehat{\varepsilon}_{(i),t} \tag{7}$$

where the index i is chosen such that $\frac{i-1}{t} < \alpha \leq \frac{i}{t}$. It is not hard to see that $i = \lceil t\alpha \rceil$. Using Equation (7), one can relate the quantile estimation error to the parameter estimation error according to the following inequality

$$\begin{aligned} &|\widehat{F}_t^{-1}(\alpha) - F^{-1}(\alpha)| \\ &\qquad \leq |F_t^{-1}(\alpha) - F^{-1}(\alpha)| + \left(1 + |p_{(i)}|\right) \|\widehat{\theta}_t - \theta\|_1, \end{aligned} \tag{8}$$

where F_t^{-1} is defined as the empirical quantile function associated with the sequence of demand shocks $\varepsilon_1, \ldots, \varepsilon_t$. Their empirical distribution is defined as $F_t(x) := \frac{1}{t} \sum_{k=1}^{t} \mathbb{1}\{\varepsilon_k \leq x\}$.

The inequality in (8) reveals that consistency of the quantile estimator (7) is reliant upon consistency of the both the *parameter estimator* and the *empirical quantile function* defined in terms of the sequence of demand shocks. Consistency of the former is established in Lemma 1 under a suitable choice of a pricing policy, which is specified in Equation (11). Consistency of the latter is clearly independent of the choice of pricing policy. In what follows, we present a bound on the rate of its convergence in probability.

Proposition 1. There exists a finite positive constant μ_1 such that

$$\mathbb{P}\{|F_t^{-1}(\alpha) - F^{-1}(\alpha)| > \gamma\} \leq 2\exp(-\mu_1 \gamma^2 t) \tag{9}$$

for all $\gamma > 0$ and $t \geq 2$.

Proposition 1 is similar in nature to [21, Lemma 2], which provides a bound on the rate at which the empirical distribution function converges to the true cumulative distribution function in probability. The combination of Assumption 1 with [21, Lemma 2] enables the derivation of the bound in (9).

IV. A No-Regret Pricing Policy

Building on the approach to demand model learning in Section III, we construct a DR pricing policy, which is guaranteed to exhibit *no-regret*.

A. Policy Design

We begin with a description of a natural approach to pricing, which interleaves the model estimation scheme defined in Section III with a *myopic* approach to pricing. That is to say, at each stage $t + 1$, the utility estimates the demand model parameters and quantile function according to (6) and (7), respectively, and sets the price according to

$$\widehat{p}_{t+1} = \frac{c_{t+1}}{2} - \frac{\widehat{b}_t + \widehat{F}_t^{-1}(\alpha)}{2\widehat{a}_t}. \tag{10}$$

Under such pricing policy, the utility essentially treats its model estimate in each period as if it were correct, and disregards the subsequent impact of its choice of price on its ability to accurately estimate the demand model in future time periods. A danger inherent to a myopic approach such as this is that the resulting price sequence may fail to elicit information from demand at a rate, which is fast enough to enable consistent model estimation. As a result, the model estimates may converge to incorrect values. Such behavior is well documented in the literature [2]–[4], and is commonly referred to as *incomplete learning*.

In order to prevent the possibility of incomplete learning in the setting considered in this paper, we propose a pricing policy, which is guaranteed to elicit information from demand at a sufficient rate through perturbations to myopic price (10). The pricing policy we propose is defined as

$$p_{t+1} = \begin{cases} \widehat{p}_{t+1}, & t \text{ odd} \\ \widehat{p}_t + \frac{1}{2}(c_{t+1} - c_t) + \delta_{t+1}, & t \text{ even}, \end{cases} \tag{11}$$

where $\delta_t := \operatorname{sgn}(c_t - c_{t-1}) \cdot t^{-1/4}$. We refer to (11) as the *perturbed myopic policy*. In defining the sign function, we require that $\operatorname{sgn}(0) = 1$. Roughly speaking, the sequence of myopic price offsets are chosen to decay at a rate, which is slow enough to ensure consistent model learning, but not so slow as to preclude a sublinear growth rate for regret.

The perturbed myopic policy (11) differs from the myopic policy (10) in two ways. First, the model parameter estimate, $\widehat{\theta}_t$, and quantile estimate, $\widehat{F}_t^{-1}(\alpha)$, are updated at every other time step. Second, to enforce sufficient price exploration, an offset is added to the myopic price at every other time step. In Section IV-B, we will show that the combination of these two features is enough to ensure consistent parameter estimation and a sublinear growth rate for the T-period regret, which is bounded from above by $O(\sqrt{T})$.

B. A Bound on Regret

Given the demand model considered in this paper, the T-period regret can be expressed as

$$\Delta^{\pi}(T) = a \sum_{t=1}^{T} \mathbb{E}^{\pi}\left[(p_t - p_t^*)^2\right] \tag{12}$$

under any pricing policy π. It becomes apparent, upon examination of Equation (12), that the rate at which regret grows is directly proportional to the rate at which pricing errors accumulate. We, therefore, proceed in deriving a bound on the rate at which the absolute pricing error $|p_t - p_t^*|$ converges to zero in probability, under the perturbed myopic policy.

First, it is not difficult to show that, under the perturbed myopic policy (11), the absolute pricing error incurred in each period t is upper bounded by

$$|p_{t+1} - p_{t+1}^*| \leq \kappa_1 \|\widehat{\theta}_t - \theta\|_1 \ + \ \kappa_2 |\widehat{F}_t^{-1}(\alpha) - F^{-1}(\alpha)| \ + \ |\delta_{t+1}| \tag{13}$$

where $\kappa_1 := \max\{\frac{1}{2\underline{a}}, \frac{\overline{b}+|F^{-1}(\alpha)|}{2a\underline{a}}\}$ and $\kappa_2 := \frac{1}{2\underline{a}}$. The upper bound in (13) is intuitive as it consists of three terms: the parameter estimation error, the quantile estimation error, and the myopic price offset – each of which represents a rudimentary source of pricing error.

One can further refine the upper bound in (13), by leveraging on the fact that, under the perturbed myopic policy, the generated price sequence is uniformly bounded. That is to say, $|p_t| \leq \overline{p}$ for all time periods t, where

$$\overline{p} := \frac{1}{2} \max \left\{ \overline{c} - \frac{\underline{\varepsilon}}{\underline{a}} \ , \ \overline{c} - \frac{\underline{\varepsilon}}{\overline{a}} \ , \ \frac{\overline{b} + \overline{\varepsilon}}{\underline{a}} \right\}.$$

Combining this fact with the previously derived upper bound on the quantile estimation error in (8), we have that

$$|p_{t+1} - p_{t+1}^*| \leq \kappa_3 \|\widehat{\theta}_t - \theta\|_1 \ + \ \kappa_2 |F_t^{-1}(\alpha) - F^{-1}(\alpha)| \ + \ |\delta_{t+1}| \tag{14}$$

where $\kappa_3 := \kappa_1 + \kappa_2(1 + \overline{p})$.

Consistency of the perturbed myopic policy depends on the asymptotic behavior of each term in (14). Among them, only the parameter estimation error depends on the choice of pricing policy. The price offset converges to zero by construction, and consistency of the empirical quantile function is established in Proposition 1. The following Lemma establishes a bound on the rate at which the parameter estimates converges to the true model parameters in probability.

Lemma 1 (Consistent Parameter Estimation)**.** There exist finite positive constants μ_2 and μ_3 such that, under the perturbed myopic policy (11),

$$\mathbb{P}\{\|\widehat{\theta}_t - \theta\|_1 > \gamma\} \leq 2\exp(-\mu_2\gamma^2(\sqrt{t} - 1)) + 2\exp(-\mu_3\gamma^2 t)$$

for all $\gamma > 0$ and $t \geq 2$.

The following Theorem provides an upper bound on the T-period regret.

Theorem 1 (Sublinear Regret)**.** There exist finite positive constants C_0, C_1, C_2, and C_3 such that, under the perturbed myopic policy (11), the T-period regret is bounded by

$$\Delta^{\pi}(T) \leq C_0 + C_1\sqrt{T} + C_2\sqrt[4]{T} + C_3\log(T)$$

for all $T \geq 2$.

In proving Theorem 1, we also show that the perturbed myopic policy (11) yields a sequence of market prices p_t, which converges to the optimal price sequence p_t^* in the mean square sense. It is also worth noting that the setting considered in this paper includes as a special case the single product setting considered in [3]. The order of the upper bound on regret derived in this paper, $O(\sqrt{T})$, is a slight improvement on the order of the bound derived in [3, Theorem 2], $O(\sqrt{T}\log T)$, as it eliminates the multiplicative factor of $\log(T)$.

V. Case Study

In this section, we compare the performance of the myopic policy (10) against the perturbed myopic policy (11) with a numerical example. We consider the setting in which there are $N = 1000$ customers participating in the DR program. For each customer i, we select a_i uniformly at random from the interval $[0.04, 0.20]$,[2] and independently select b_i according an exponential distribution (with mean equal to 0.01) truncated over interval $[0, 0.1]$. Parameters are drawn independently across customers. For each customer i, we take the the demand shock to be distributed according to a normal distribution with zero-mean and standard deviation equal to 0.04, truncated over the interval $[-0.4, 0.4]$. We consider a utility with risk sensitivity equal to $\alpha = 0.1$. In other words, the utility seeks to maximize the revenue it is guaranteed to receive with probability 0.9 or greater. Finally, we take the wholesale price of electricity to be fixed at $c_t = 1.5$ \$/kWh for all times t.

A. Discussion

Because the wholesale price of electricity is fixed over time, the parameter and quantile estimates represent the only source of variation in the sequence of prices generated by the myopic policy. Due to the combined structure of the myopic policy and the least squares estimator, the value of each new demand observation rapidly diminishes over time, which, in turn, manifests in a rapid convergence of the myopic price process. The resulting lack exploration in the sequence of myopic prices results in incomplete learning, which is seen in Figure 1. Namely, the sequence of myopic prices converges to a value, which differs form the oracle optimal price. As a consequence, the myopic policy incurs a T-period regret that grows linearly with time, as is observed in Figure 2.

On the other hand, the price offset δ_t generates enough variation in sequence of prices generated by the perturbed myopic policy to ensure consistent model estimation. This, in turn, results in convergence of the sequence of posted prices to the oracle optimal price. This, combined with the fact that the price offset δ_t vanishes asymptotically, ensures sublinearity of the resulting T-period regret, as is observed in Figure 2.

VI. Conclusion

In this paper, we propose a data-driven approach to pricing demand response with the aim of maximizing the risk sensitive revenue derived by the utility. The pricing policy we propose

[2]This range of parameter values is consistent with the range of demand price elasticities observed in several real-time pricing programs operated in the United States [22], [23].

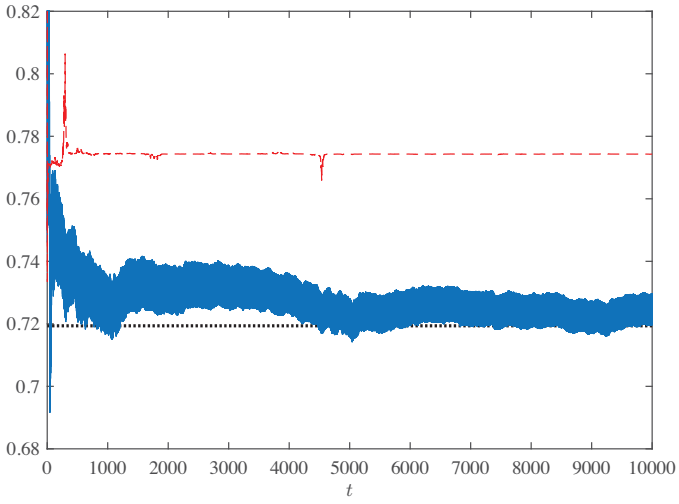

Fig. 1. A sequence of prices ($/kWh) generated by the perturbed myopic policy (——), the myopic policy (- - -), and the oracle policy (·····).

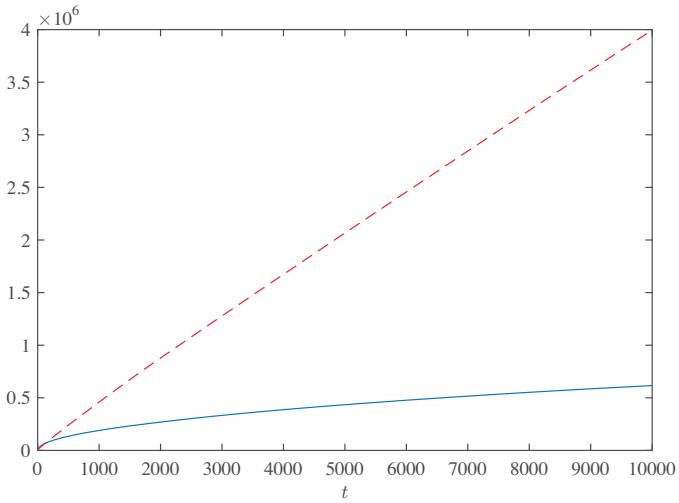

Fig. 2. Regret of the perturbed myopic policy (——) and the myopic policy (- - -).

has two key features. First, the unknown demand model parameters are estimated using a least squares estimator. Second, the proposed policy incorporates an explicit price offset to ensure sufficient exploration in the sequence of prices it generates. We show that these two features together guarantee complete learning. Moreover, we show that the order of regret associated with the proposed policy is no worse than $O(\sqrt{T})$.

REFERENCES

[1] O. Besbes and A. Zeevi, "On the (surprising) sufficiency of linear models for dynamic pricing with demand learning," *Management Science*, vol. 61, no. 4, pp. 723–739, 2015.

[2] A. V. den Boer and B. Zwart, "Simultaneously learning and optimizing using controlled variance pricing," *Management science*, vol. 60, no. 3, pp. 770–783, 2013.

[3] N. B. Keskin and A. Zeevi, "Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies," *Operations Research*, vol. 62, no. 5, pp. 1142–1167, 2014.

[4] T. Lai and H. Robbins, "Iterated least squares in multiperiod control," *Advances in Applied Mathematics*, vol. 3, no. 1, pp. 50–73, 1982.

[5] J. A. Taylor and J. L. Mathieu, "Index policies for demand response," *Power Systems, IEEE Transactions on*, vol. 29, no. 3, pp. 1287–1295, 2014.

[6] D. Kalathil and R. Rajagopal, "Online learning for demand response," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2015, pp. 218–222.

[7] S. Jain, B. Narayanaswamy, and Y. Narahari, "A multiarmed bandit incentive mechanism for crowdsourcing demand response in smart grids," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[8] Q. Wang, M. Liu, and J. L. Mathieu, "Adaptive demand response: Online learning of restless and controlled bandits," in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*. IEEE, 2014, pp. 752–757.

[9] R. Gomez, M. Chertkov, S. Backhaus, and H. J. Kappen, "Learning price-elasticity of smart consumers in power distribution systems," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*. IEEE, 2012, pp. 647–652.

[10] L. Jia, L. Tong, and Q. Zhao, "An online learning approach to dynamic pricing for demand response," *arXiv preprint arXiv:1404.1325*, 2014.

[11] D. O. Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 409–414.

[12] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, "Real-time load elasticity tracking and pricing for electric vehicle charging," *Smart Grid, IEEE Transactions on*, vol. 6, no. 3, pp. 1303–1313, 2015.

[13] E. Bitar and Y. Xu, "On incentive compatibility of deadline differentiated pricing for deferrable demand," in *Decision and control (CDC), 2013 IEEE 52nd annual conference on*. IEEE, 2013, pp. 5620–5627.

[14] ——, "Deadline differentiated pricing of deferrable electric loads," *Smart Grid, IEEE Transactions on*, to appear, 2016.

[15] W. Lin and E. Bitar, "Forward electricity markets with uncertain supply: Cost sharing and efficiency loss," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 1707–1713.

[16] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 320–331, 2010.

[17] W. Saad, Z. Han, H. V. Poor, and T. Bacsar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 86–105, 2012.

[18] Y. Xu, N. Li, and S. H. Low, "Demand response with capacity constrained supply function bidding," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1377–1394, March 2016.

[19] H. Tavafoghi and D. Teneketzis, "Optimal contract design for energy procurement," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 62–69.

[20] K. Khezeli and E. Bitar, "Risk-sensitive learning and pricing for demand response," *in prepration*.

[21] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

[22] Q. QDR, "Benefits of demand response in electricity markets and recommendations for achieving them," *US Dept. Energy, Washington, DC, USA, Tech. Rep*, 2006.

[23] A. Faruqui and S. Sergici, "Household response to dynamic pricing of electricity: a survey of 15 experiments," *Journal of regulatory Economics*, vol. 38, no. 2, pp. 193–225, 2010.