This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

Risk-Sensitive Learning and Pricing for 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 pricing policy that 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.**

***Index Terms*—Demand response, electricity markets, dynamic pricing, online learning.**

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 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 . Here, encodes the utility's sensitivity to risk. In this paper, we propose a causal pricing policy, which resolves the trade-off 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 days – relative to an oracle that knows the underlying demand model – which is at most . 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 and adaptive control [1]–[5], 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 [6]–[9]. In this setting, each arm represents a customer or a class of customers. Taylor and Mathieu [6] show that, in the absence of exogenous shocks on load curtailment, the optimal policy is indexable. Kalathil and Rajagopal [7] consider a similar multi-armed bandit

Supported in part by NSF grant ECCS-1351621, NSF grant CNS-1239178, NSF grant IIP-1632124, US DoE under the CERTS initiative, and the Simons Institute for the Theory of Computing.

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

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 guarantees that the -period regret is bounded from above by . There is a related stream of literature, which treats the problem of pricing demand response under uncertainty using techniques from online learning [10]–[13]. Perhaps closest to the setting considered in this paper, Jia et al. [11] 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 -period regret that is of the order . There is another stream of literature, which considers an auction-based approach to the procurement of demand response [14]–[20]. 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. We investigate the behavior of the proposed pricing policy with a numerical case study in Section VI. All mathematical proofs are presented in the Appendix to the paper.

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 customers over multiple time periods (e.g., days) indexed by . The class of DR programs we consider rely on uniform price-based incentives for demand reduction.[1] Specifically, prior to each time period , the utility broadcasts a single price ($/kWh), to which each participating customer responds with a reduction in demand (kWh) – thus entitling customer to receive a payment in the amount of .[2]

We model the response of each customer to the posted price at time according to a linear demand function given by

for

where and are model parameters *unknown to the utility*, and is an unobservable demand shock, which we model as a random variable with zero mean.[3] *Its distribution is also unknown to the utility*. We define the aggregate response of customers at time as $\sum$, which satisfies

(1)

Here, the aggregate model parameters and shock are defined as $\sum$, $\sum$, and $\sum$. To simplify notation in the sequel, we write the deterministic component of aggregate demand as , where denotes the aggregate demand function parameters.

We assume throughout the paper that and , where the model parameter bounds are assumed to be known and satisfy and . 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 are independent and identically distributed random variables, in addition to the following technical assumption.

Assumption 1. The aggregate demand shock has a bounded range , and a cumulative distribution function , which is bi-Lipschitz over this range. Namely, there exists a real constant , such that for all , it holds that

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 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 electric power utility need not know the parameters specified in Assumption 1, beyond the assumption of their boundedness.

Remark 1 (On the Linearity Assumption)**.** While the assumption of linearity in the underlying demand model might appear restrictive at first glance, there are several sensible arguments in support of its adoption. First, the assumption of linearity is routinely employed in the revenue management and pricing literature [3], [11], [27]–[30], as it serves to facilitate theoretical analyses, thereby bringing to light key features of the problem and its solution structure. More practically, if the range of allowable prices is sufficiently limited, then it is reasonable to assume that the underlying (possibly nonlinear) demand function is well approximated by an affine function over that range. And, in the specific context of pricing for DR programs, it is reasonable to expect that the electric power utility, being a regulated company, will face restrictions on the range of prices that it can offer to customers. Finally, there are recent results in the revenue management literature [1], which demonstrate how the assumption of a linear demand model might be dynamically *adapted* to price in environments where the true demand function is nonlinear. The generalization of such techniques to accommodate the risk-sensitive criteria

[1]This class of DR programs falls within the more general category of programs that rely on *peak time rebates* (PTR) as incentives for demand reduction [21].

[2]A customer's reduction in demand is measured against a predetermined baseline. The question as to how such baselines might be reliably inferred is a challenging and active area of research [22]–[26]. Expanding our model to make endogenous the calculation of customer baselines is left as a direction for future research.

[3]We note that the assumption that be zero-mean is without loss of generality.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

considered in this paper (cf. Equation (2)) represents an interesting direction for future research.

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 . Accordingly, we let (\$/kWh) denote the wholesale price of electricity during peak demand hours on day . And, we let (\$/kWh) denote the retail price of electricity, i.e., the fixed price that customers are charged for their electricity consumption. For the remainder of the paper, it will be convenient to work with the difference between the wholesale and retail prices of electricity on each day , which we denote by . We assume throughout the paper that − for all days , where − .[4] In addition, we assume that is known to the utility prior to its determination of the DR price in each period . Upon broadcasting a price to its customer base, and realizing an aggregate demand reduction , the utility derives a net reduction in its peak electricity cost in the amount of . Henceforth, we will refer to the net savings as the *revenue* derived by the utility in period .

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

(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 , one can reformulate the expression in (2) as

(3)

where denotes the - quantile of the random variable . It is immediate to see from the simplified expression in (3) that is strictly concave in . Let denote the *oracle optimal price*, which maximizes the risk-sensitive revenue in period . Namely,

The optimal price is readily derived from the corresponding first order optimality condition, and is given by

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

$$\sum$$

[4]Implicit in this requirement is the assumption that − for all days . The lower bound on implies that the utility will only call for a demand reduction on those days in which the wholesale market manifests in prices that exceed the fixed retail price for electricity. The upper bound on implies the enforcement of a *price cap* in the wholesale market.

The term oracle is used, as equals the maximum risk-sensitive revenue achievable by the utility over 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 and the shock distribution 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 , the utility has access to the entire history of prices and demand reductions until period . We, therefore, define a *feasible pricing policy* as an infinite sequence of functions , where each function in the sequence is allowed to depend only on the past history. More precisely, we require that the function be measurable according to the - algebra generated by the history of past decisions and demand observations for all , and that be a deterministic constant. The *expected risk-sensitive revenue* generated by a feasible pricing policy over time periods is defined as

$$\left[\sum\right]$$

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

C. Performance Metric

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

Naturally, pricing policies yielding a small regret are preferred, as the oracle risk-sensitive revenue stands as an upper bound on the expected risk-sensitive revenue achievable by any feasible pricing policy . Ultimately, we seek a pricing policy whose -period regret is sublinear in the horizon . Such a pricing policy is said to have *no-regret*.

Definition 1 (No Regret Pricing)**.** A feasible pricing policy is said to exhibit *no-regret* if .

Implicit in the goal of designing a no-regret policy is that the sequence of prices that it generates should converge to the oracle optimal price sequence.

III. Demand Model Learning

Clearly, the ability to price with no-regret will rely centrally on the rate at which the unknown parameters, , and quantile function, , can be learned from the market data. In what follows, we describe a basic approach to learning the demand model using the method of least squares estimation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

A. Parameter Estimation

Given the history of past decisions and demand observations through period , define the *least squares estimator* (LSE) of as

$$\left\{ \sum \right\}$$

for time periods . The LSE at period admits an explicit expression of the form

$$\left(\sum [\ ][\ ] \right) \left(\sum [\ ] \right) \quad (4)$$

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

$$\mathscr{J} \quad \sum [\ ][\ ] \quad \begin{bmatrix} \sum & \sum \\ \sum & \end{bmatrix}$$

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:

$$\mathscr{J} \left(\sum [\ ] \right) \quad (5)$$

Remark 2 (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 $\mathscr{J}$. Namely, the minimum eigenvalue of $\mathscr{J}$, 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 $\mathscr{J}$ is bounded from below (up to a multiplicative constant) by the *sum of squared price deviations* defined as

$$\sum \quad -$$

where $-$ $\sum$. The result is reliant on the assumption that the underlying pricing policy yields a bounded sequence of prices . 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 belong to a compact set defined $\_\ -\ -$, one can improve upon the LSE at time by projecting it onto the set . Accordingly, we define the *truncated least squares estimator* as

$$\widehat{\ } \quad (6)$$

Clearly, we have that $\widehat{\ }$. 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 according to the empirical quantile function associated with the demand estimation residuals. Namely, in each period , define the sequence of *residuals* associated with the estimator $\widehat{\ }$ as

$$\widehat{\ } \qquad \widehat{\ }$$

for . Define their *empirical distribution* as

$$\widehat{\ } \quad - \sum \mathbb{1} \ \widehat{\ }$$

and their corresponding *empirical quantile function* as $\widehat{\ }$ $\widehat{\ }$ for all .
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* $\widehat{\ }$ $\widehat{\ }$ are defined as a permutation of $\widehat{\ }$ $\widehat{\ }$ such that $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ With this concept in hand, the empirical quantile function can be equivalently expressed as

$$\widehat{\ } \qquad \widehat{\ } \quad (7)$$

where the index is chosen such that — -. It is not hard to see that . Using Equation (7), one can relate the quantile estimation error to the parameter estimation error according to the following inequality

$$\widehat{\ } \qquad (\qquad) \ \widehat{\ } \quad (8)$$

where is defined as the empirical quantile function associated with the sequence of demand shocks . Their empirical distribution is defined as

$$- \sum \mathbb{1}$$

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 we specify 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. Let . It holds that

$$(9)$$

for all and .

Proposition 1 is similar in nature to [31, 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 [31, Lemma 2] enables the derivation of the upper bound in Proposition 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

IV. Design of Pricing Policies

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. Myopic Policy

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 , the utility estimates the demand model parameters and quantile function according to (6) and (7), respectively, and sets the price according to

$$ \qquad (10)$$

Under this 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 to pricing 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 Section VI, we provide a numerical example, which demonstrates the occurrence of incomplete learning under the myopic pricing policy (10).

B. Perturbed Myopic Policy

In order to prevent the possibility of incomplete learning, we propose a pricing policy that is guaranteed to elicit information from demand at a sufficient rate through carefully designed perturbations to the myopic pricing policy (10). The pricing policy we propose is defined as

$$\begin{cases} & \text{odd} \\ & \text{even} \end{cases} \qquad (11)$$

where is a user specified positive constant, and

$$\operatorname{sgn}$$

We refer to the policy (11) as the *perturbed myopic policy*.[5]

The perturbed myopic policy differs from the myopic policy in two important ways. First, the model parameter estimate, , and quantile estimate, , 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. Roughly speaking, the sequence of myopic price offsets is chosen to decay at a rate, which is slow enough to ensure consistent model learning, but not so slow as to preclude a sub-linear growth rate for regret. In Section V, we will show that the combination of these features is enough to ensure consistent parameter estimation and a sub-linear growth rate for the -period regret, which is bounded from above by .

[5] In defining the sign function, we require that sgn .

Remark 3 (On the Perturbation Order)**.** We briefly describe the rationale behind the selection of the order of the perturbation sequence as . First, notice from Equation (12) that the regret incurred by any feasible pricing policy is equal to the sum of the squared pricing errors generated by the policy. Combining this expression with the upper bound on the absolute pricing error induced by the perturbed myopic policy in (14), it becomes clear to see the conflicting effects that the perturbation sequence has on regret. On the one hand, an increase in the order of the perturbation sequence will tend to reduce the growth rate of regret by increasing the rate at which the parameter estimation error converges to zero. On the other hand, an increase in the order of the perturbation sequence will tend to have the counterproductive effect of increasing the growth rate of regret by increasing the rate at which the deliberate pricing errors accumulate. A tradeoff, therefore, emerges in selecting the order of the perturbation sequence. In Appendix B, we show that among all perturbation sequences that are polynomial in , perturbation sequences of the order are optimal in the sense of minimizing the asymptotic order of our upper bound on regret.

V. A Bound on Regret

Given the demand model considered in this paper, one can express the -period regret as

$$\sum \left[\quad \right] \qquad (12)$$

under any pricing policy . 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 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 even time period is upper bounded by

$$ \qquad (13)$$

where and . The pricing error incurred during odd time periods is similarly bounded, sans the explicit dependency on the myopic price perturbation. 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 perturbation – 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 sequence of prices is uniformly bounded. That is to say, for all time periods , where

$$ $$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

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

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

for even time periods t, 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 perturbation 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 converge 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),

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

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

The following Theorem establishes an $O(\sqrt{T})$ upper bound on the T-period regret.

Theorem 1 (Sub-linear Regret). The T-period regret incurred by the perturbed myopic policy (11) satisfies

$$\Delta^\pi(T) \leq C_0 + C_1\sqrt{T} + C_2\sqrt[4]{T} + C_3\log(T), \tag{15}$$

for all $T \geq 2$. Here, $C_0, C_1, C_2,$ and C_3 are finite positive constants.[6]

Remark 4 (Tuning the Parameter, ρ). As one might expect, the coefficients $C_0, C_1, C_2,$ and C_3 depend on the user specified parameter ρ. Accordingly, it is natural to ask as to whether or not it is tractable to calculate a value for ρ, which minimizes the upper bound on regret in (15), given a fixed horizon T. The short answer is yes. A cursory examination of the coefficient Equations (30)-(33) reveals the upper bound on regret in (15) to be a strictly convex and differentiable function in the parameter ρ over the positive real numbers. This renders its minimization in the parameter ρ a straightforward task. That is to say, given a fixed horizon T, one can readily calculate

$$\rho^*(T) := \operatorname{argmin}\{C_0 + C_1\sqrt{T} + C_2\sqrt[4]{T} + C_3\log(T) : \rho \in \mathbb{R}\}, \tag{16}$$

using one of variety of first-order, second-order, or bisection-based numerical methods. In Section VI, we conduct a numerical study to asses the performance of the perturbed myopic policy when its tuning parameter is selected according to Equation (16). The numerical results suggest that a selection of the tuning parameter according to $\rho = \rho^*(T)$ manifests in a corresponding T-period regret that is comparable to the minimum achievable regret over all possible tuning parameters $\rho \in \mathbb{R}_+$. We refer the reader to Figure 1 for a graphical illustration of this comparison.

[6]We refer the reader to Equations (30) -(33) for the exact specification of the coefficients $C_0, C_1, C_2,$ and C_3.

In the process of proving Theorem 1, we also show that the perturbed myopic policy generates a sequence of market prices $\{p_t\}$ that converges to the oracle optimal price sequence $\{p_t^*\}$ in the mean square sense. More formally, we have the following corollary.

Corollary 1 (Price Consistency). The sequence of prices $\{p_t\}$ generated by the perturbed myopic policy (11) satisfies

$$\lim_{t\to\infty} \mathbb{E}\left[(p_t - p_t^*)^2\right] = 0,$$

where $\{p_t^*\}$ denotes the oracle optimal price sequence.

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 $\log(T)$.

A. The Exploratory Effect of Wholesale Price Variation

Thus far in this paper, we have made no assumption on the nature of variation in the sequence of wholesale electricity prices $\{w_t\}$. In particular, all of the previously stated results hold for any sequence of time-varying wholesale electricity prices. This includes the special case in which the wholesale price of electricity is constant across time, i.e., $w_t = w$ for all time periods t. It is, however, natural to inquire as to how the degree of variation in the sequence of wholesale prices might impact the performance of the pricing policies considered in this paper.

First, it is straightforward to see from Equation (10) that variation in the sequence of wholesale prices induces equivalent variation in the sequence of myopic prices. Such variation in the myopic price sequence is most naturally interpreted as a form of *costless exploration*. In the following result, we establish a sufficient condition on the variation of wholesale prices, which eliminates the need for external perturbations to the myopic price sequence (i.e., setting $\rho = 0$), while guaranteeing an upper bound on the resulting T-period regret that is $O(\log T)$.

Theorem 2 (Logarithmic Regret). Assume that there exists a finite positive constant $\sigma > 0$ such that

$$|w_t - w_{t-1}| \geq \sigma, \tag{17}$$

for all time periods t.[7] It follows that the T-period regret incurred by the perturbed myopic policy (11), with $\rho = 0$, satisfies

$$\Delta^\pi(T) \leq M_0 + \frac{M_1}{\sigma^2} + \left(\frac{M_1}{\sigma^2} + M_2\right)\log(T), \tag{18}$$

for all $T \geq 2$. Here, $M_0, M_1,$ and M_2 are finite positive constants[8], which are independent of the parameter σ.

Several comments are in order. First, under the additional assumption of persistent wholesale price variation (17), we

[7]Note that Assumption (17) in Theorem 2 implies that $|c_t - c_{t\ 1}| \geq \sigma$.

[8]We refer the reader to Equations (36) -(38) for the exact specification of the coefficients $M_0, M_1,$ and M_2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

establish in Theorem 2 an improvement upon the original order of regret stated in Theorem 1 from to . However, as one might expect, the magnitude of the upper bound on regret in (18) scales in a manner that is inversely proportional to . As a result, the upper bound on the -period regret goes to infinity as goes to zero, and, therefore, provides little useful information when is small.

VI. Case Study

We conduct a numerical analysis to compare the performance of the myopic policy (10) against the perturbed myopic policy (11) over a time horizon of . Given this time horizon, we set the tuning parameter according to Equation (16) as . We consider the setting in which there are customers participating in the DR program. For each customer , we select uniformly at random from the interval , and independently select according an exponential distribution (with mean equal to) truncated over interval . Parameters are drawn independently across customers.[9] For each customer , we take the demand shock to be distributed according to a normal distribution with zero-mean and standard deviation equal to , truncated over the interval . We consider a utility with risk sensitivity equal to . In other words, the utility seeks to maximize the revenue it is guaranteed to receive with probability no less than 0.9. Finally, we set the retail price of electricity to ($/kWh), and set the wholesale price of electricity to ($/kWh) for all days . Such values are consistent with the average residential retail and peak wholesale prices of electricity in the state of New York in 2016 [34], [35].

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 sequence of prices generated under the myopic policy. The resulting lack of exploration in the sequence of myopic prices results in incomplete learning, which is seen in Figure 3a. Namely, the sequence of myopic prices converges to a value, which substantially differs form the oracle optimal price. As a consequence, the myopic policy incurs a -period regret that grows linearly with the horizon , as is observed in Figure 3.

On the other hand, the sequence of perturbations generate enough variation in the 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 vanishes at a sufficiently fast rate, ensures sublinearity in the growth rate of the corresponding -period regret, as is observed in Figure 3.

[9] It is worth noting that the range of parameter values considered in this numerical study is consistent with the range of demand price elasticities observed in several real-time pricing programs conducted in the United States [32], [33].

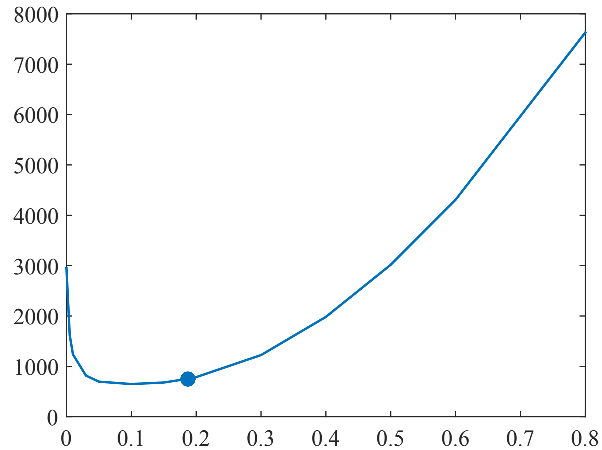


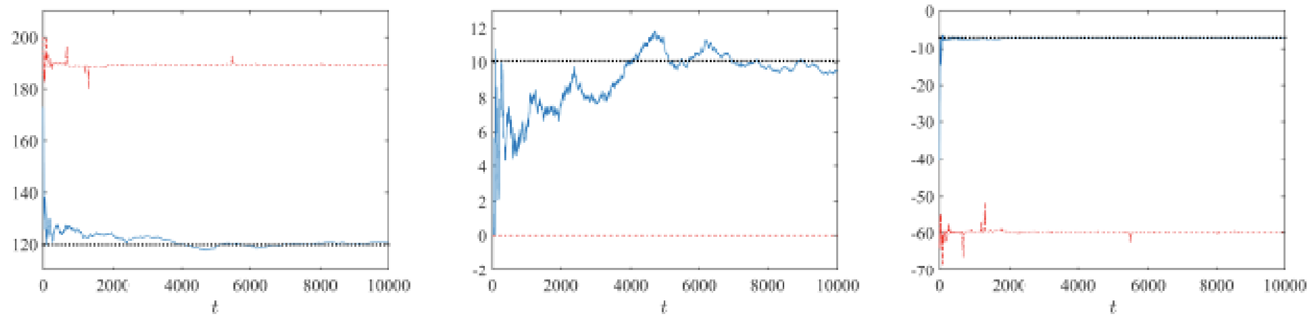
Fig. 1: For a fixed horizon , the figure above plots the regret incurred under the perturbed myopic policy as function of the tuning parameter . The bullet ( ) indicates the regret incurred under the perturbed myopic policy when the tuning parameter is selected according to Equation (16), i.e., .

For the chosen horizon of , we depict in Figure 1 the sensitivity of the regret incurred under the perturbed myopic policy to the choice of tuning parameter . The bullet point specifies the regret incurred when the policy's tuning parameter is selected according to Equation (16), i.e., . For the time horizon considered, the numerical results indicate that a selection of the tuning parameter according to Equation (16) results in a corresponding regret that is comparable to the minimum achievable regret over all possible tuning parameters .

VII. 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 electric power utility. The perturbed myopic pricing policy we propose has two key features. First, the unknown demand model parameters are estimated using a least squares estimator. Second, the proposed policy implements a sequence of perturbations to the myopic price sequence to ensure sufficient exploration in the sequence of prices it generates. The price perturbation sequence is designed to decay at a rate, which is slow enough to ensure complete learning of the underlying demand model, but not so slow as to preclude a sub-linear growth rate for regret. In particular, the proposed pricing policy is proven to exhibit a -period regret that is no greater than . As a direction for future research, it would be interesting to investigate the generalization of the pricing algorithms developed in this paper to accommodate the treatment of nonlinear and possibly time-varying demand functions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

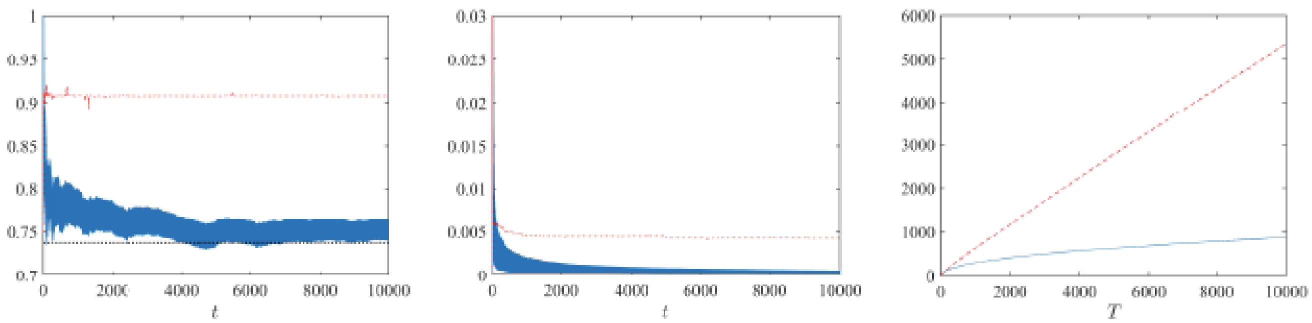


(a) Sequences of demand parameter $\widehat{a}_t$.

(b) Sequences of demand parameter $\widehat{b}_t$.

(c) Sequences of quantile function $\widehat{F}_t^{-1}(\alpha)$.

Fig. 2: (a)-(b) Sample paths of the parameter estimates, and (c) sample path of the shock quantile estimates under the *myopic policy* (- - -), the *perturbed myopic policy* (——), and the *oracle policy* (-·-·-).



(a) Sequences of posted prices.

(b) Mean squared pricing error.

(c) T-period Regret.

Fig. 3: (a) Sample path of posted prices, (b) mean squared pricing error, and (c) regret under the *myopic policy* (- - -), the *perturbed myopic policy* (——), and the *oracle policy* (-·-·-).

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] A. V. den Boer, "Dynamic pricing and learning: historical origins, current research, and new directions," *Surveys in operations research and management science*, vol. 20, no. 1, pp. 1–18, 2015.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

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

[19] 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.

[20] H. Tavafoghi and D. Teneketzis, "Optimal contract design for energy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

procurement," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 62–69.

[21] A. Faruqui, R. Hledik, and J. Tsoukalis, "The power of dynamic pricing," *The Electricity Journal*, vol. 22, no. 3, pp. 42–56, 2009.

[22] H. Chao, "Demand response in wholesale electricity markets: the choice of customer baseline," *Journal of Regulatory Economics*, vol. 39, no. 1, pp. 68–88, 2011.

[23] C. Chelmis, M. R. Saeed, M. Frincu, and V. Prasanna, "Curtailment estimation methods for demand response: Lessons learned by comparing apples to oranges," in *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*. ACM, 2015, pp. 217–218.

[24] ConEdison, "Energy efficiency and demand management procedure - general calculating customer baseline load," 2013.

[25] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote, "Statistical analysis of baseline load models for non-residential buildings," *Energy and Buildings*, vol. 41, no. 4, pp. 374–381, 2009.

[26] D. Muthirayan, D. Kalathil, K. Poolla, and P. Varaiya, "Mechanism design for self-reporting baselines in demand response," in *American Control Conference (ACC), 2016*. American Automatic Control Council (AACC), 2016, pp. 1446–1451.

[27] D. Bertsimas and P. Vayanos, "Data-driven learning in dynamic pricing using adaptive optimization," *Preprint*, 2015.

[28] L. Jia and L. Tong, "Day ahead dynamic pricing for demand response in dynamic environments," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 5608–5613.

[29] K. T. Talluri and G. J. Van Ryzin, *The theory and practice of revenue management*. Springer Science & Business Media, 2006, vol. 68.

[30] J. B. Taylor, "Asymptotic properties of multiperiod control rules in the linear regression model," *International Economic Review*, pp. 472–484, 1974.

[31] 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.

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

[33] 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.

[34] EIA, "Electric power monthly," *U.S. Department of Energy*, February 2017.

[35] NYISO. (2016) Markets and operational data. [Online]. Available: www.nyiso.com/public/markets_operations/market_data/pricing_data

[36] Y. Chow and T. Lai, "Limiting behavior of weighted sums of independent random variables," *The Annals of Probability*, pp. 810–824, 1973.

APPENDIX

In the following proofs, we consider a more general form of the perturbation as sgn where is allowed to be an arbitrary constant in the interval . Ultimately, we will prove that a choice of minimizes the asymptotic order of the upper bound on regret, which we establish in (28).

APPENDIX A
PROOF OF LEMMA 1

It is straightforward to show that the parameter estimation error is bounded from above by

$$\widehat{\quad} \qquad - \tag{19}$$

where

$$-\sum \qquad \text{and} \qquad \sum \tag{20}$$

Recall that $\sum$. It follows that

$$\widehat{\quad} \qquad -$$

We will upper bound each term separately to establish the desired result. In doing so, we will rely on the following Lemma – which we state without proof, as it follows from a direct application of the Chernoff bound together with Hoeffding's Lemma.[10]

Lemma 2. Let be an infinite sequence of zero mean independent random variables, satisfying , almost surely, for all . Let be an infinite sequence of real numbers, and define the sequence of random variables

$$\left(\sum \quad\right) \Big/ \left(\sum \quad\right)$$

For all and , it holds that

$$\left(\frac{\;\overline{\quad}\;}{\overline{\quad}} \sum \quad \right)$$

First term: By setting and , a direct application of Lemma 2 yields

$$\left(\frac{\quad}{- \quad -} \right) \tag{21}$$

Second term: By setting and , a direct application of Lemma 2 yields

$$\left[\quad \left(\frac{\quad}{- \quad -} \right) \right] \tag{22}$$

The equality follows from the law of total probability, and the inequality follows from monotonicity of the expectation operator.

We now bound the random process from below by a deterministic sequence. Fix . A direct substitution of the perturbed myopic policy yields

$$\sum \left\{ \widehat{\quad} \qquad \left(\widehat{\quad} \quad - \quad \right) \right\}$$

The above inequality can be further relaxed to eliminate its explicit dependency on the (random) price process. Namely, it is straightforward to show that

$$- \sum \frac{\quad}{\quad} \quad - \sum \tag{23}$$

One can further relax inequality (23) by using the facts that

$$\sum \frac{\quad}{\quad} \quad \int \frac{\quad}{\quad} \qquad \frac{\quad}{\quad}$$

and

$$\left(\quad \frac{\quad}{\quad} \right)$$

[10]We refer the reader to [36, Lemma 4] for a proof of a similar result.

It follows that

$$\sum \tag{24}$$

where is defined as

$$\sum$$

An application of the inequality in (24) to (22) yields

$$\tag{25}$$

Together, the upper bounds in (21) and (25) give

$$\tag{26}$$

where and . To complete the proof, we set . For this choice of , we have that .

Appendix B
Proof of Theorem 1

We introduce an additional assumption on the variation in the sequence of wholesale electricity prices.[11] Namely, let be nonnegative constant such that for all . Ultimately, we will establish the desired result for , the setting considered in the statement of the Theorem.

We begin with the following upper bound on the -period regret.

$$\sum \qquad \sum^{\frac{T+1}{2}} \qquad \sum^{\frac{T+1}{2}}$$

where the constant is defined as

Recall that is assumed to be a deterministic constant. The inequality is immediate, and the last equality follows from the fact that for any pair of scalars . We establish the following technical Lemma to bound the expectation in the above inequality. Its proof is postponed to Appendix C.

[11]Such assumption will prove useful in facilitating the proof of Theorem 2.

Lemma 3. Under the perturbed myopic policy (11), it holds that

where , and are nonnegative constants.

A direct application of Lemma 3 yields

$$\sum^{\frac{T+1}{2}} \tag{27}$$

$$\sum \tag{28}$$

Inequality (27) follows from the facts that , , and . Inequality (28) follows from the definition of and the assumption that for all .

For , it is straightforward to show that a choice of minimizes the asymptotic order of the upper bound (28) with respect to the horizon . Setting yields

$$\sum \tag{29}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2700458, IEEE Transactions on Smart Grid

Inequality (29) follows from the bounds

and

Taking the limit of the upper bound in (29) as goes to zero yields

Finally, we define the nonnegative constants , , , and as follows to conclude the proof.

(30)

(31)

(32)

(33)

Appendix C
Proof of Lemma 3

First, we use the fact that for a continuous nonnegative random variable , it holds that .

(34)

(35)

Inequality (34) follows from triangle inequality. The equality in (35) follows from the fact that for all , it holds that

Now, recall that

The above inequality, together with the inequality in (26) and Proposition 1, yield

Applying the above bound to (35), and explicitly calculating the resulting integral yields

Note that, in calculating the aforementioned integral, we used the identity:

Finally, we define nonnegative constants , , , and as follows to conclude the proof.

Appendix D
Proof of Theorem 2

Inequality (28) is a valid upper bound the -period regret incurred by perturbed myopic policy, under the assumption that . By setting , the upper bound simplifies to

We define the nonnegative constants , , and as follows to conclude the proof.

(36)

(37)

(38)

Note that the above constants are specified in such a manner as to be independent of the parameter .

Kia Khezeli has been pursuing the Ph.D. degree in Electrical and Computer Engineering from Cornell University since 2014. He received the B.S. degree in Electrical Engineering from Sharif University of Technology in 2012, and the M.S. degree in Electrical and Computer Engineering from McMaster University in 2014. His research interests include probability theory, learning, and stochastic optimization with particular application in managing uncertainty generated by renewable resources in electric power systems. He is a recipient of the Jacobs Fellowship, the Outstanding Thesis Award (McMaster), and the National Elite Foundation Fellowship (Iran).

Eilyan Bitar currently serves as an Assistant Professor and the David D. Croll Sesquicentennial Faculty Fellow in the School of Electrical and Computer Engineering at Cornell University, Ithaca, NY, USA. Prior to joining Cornell in the Fall of 2012, he was engaged as a Postdoctoral Fellow in the department of Computing and Mathematical Sciences at the California Institute of Technology and at the University of California, Berkeley in Electrical Engineering and Computer Science, during the 2011-2012 academic year. His current research examines the operation and economics of modern power systems, with an emphasis on the design of markets and optimization methods to manage uncertainty in renewable and distributed energy resources. He received the B.S. and Ph.D. degrees in Mechanical Engineering from the University of California at Berkeley in 2006 and 2011, respectively.

Dr. Bitar is a recipient of the NSF Faculty Early Career Development Award (CAREER), the John and Janet McMurtry Fellowship, the John G. Maurer Fellowship, and the Robert F. Steidel Jr. Fellowship.