Aggregate Modeling of Fast-Acting Demand Response and Control Under Real-Time Pricing

David P. Chassin[1,2] (dchassin@slac.stanford.edu)

and

Daniel Rondeau[1] (rondeau@uvic.ca)

[1] *University of Victoria, Victoria, BC (Canada)*

[2] *SLAC National Accelerator Laboratory, Menlo Park, CA (USA)*

Abstract

This paper develops and assesses the performance of a short-term demand response (DR) model for utility load control with applications to resource planning and control design. Long term response models tend to under estimate short-term demand response when induced by prices. This has two important consequences. First, planning studies tend to undervalue DR and often overlook its benefits in utility demand management program development. Second, when DR is not overlooked, the open-loop DR control gain estimate may be too low. This can result in overuse of load resources, control instability and excessive price volatility. Our objective is therefore to develop a more accurate and better performing short-term demand response model. We construct the model from first principles about the nature of thermostatic load control and show that the resulting formulation corresponds exactly to the Random Utility Model employed in economics to study consumer choice. The model is tested against empirical data collected from field demonstration projects and is shown to perform better than alternative models commonly used to forecast demand in normal operating conditions. The results suggest that (1) existing utility tariffs appear to be inadequate to incentivize demand response, particularly in the presence of high renewables, and (2) existing load control systems run the risk of becoming unstable if utilities close the loop on real-time prices.

Highlights:

- Demand elasticity for fast-acting demand response load under real-time pricing

- Validated first-principles logistic demand curve matches random utility model

- Logistic demand curve suitable for diversified aggregate loads market-based transactive control systems

Keywords: electricity demand response, demand elasticity, real-time pricing, indirect load control, transactive systems, random utility model.

---

1. Introduction

In 2003 Economics Nobel Laureate Vernon Smith published an editorial with Lynne Kiesling in the Wall Street Journal [1] summarizing the consensus in the wake of the California Electricity crisis. In their view the crisis was in part precipitated by the lack of customer engagement in electricity pricing mechanisms [2]. Reflecting on the technical and regulatory supply-side response to the crisis, they wrote "What is inadequately discussed, let alone motivated, is the [other] option – demand response". It is now widely accepted that demand response can mitigate the market power of energy suppliers. More importantly, demand response presents a real opportunity for improvement in electricity planning and operations. Research on short-term demand response in particular has increased as the growth of intermittent wind

and solar resources further exacerbates the problem of managing the balance between supply and demand in power systems [3].

Historically, demand response programs have taken the form of so-called "demand side management" (DSM) activities. DSM seeks to alter electricity demand load shapes to make them better match the available supply and reduce load peaks so as to defer costly capacity expansion investments. Traditional DSM programs include increased building and appliance efficiency standards, as well as equipment replacement/retrofit programs.

Load shifting has also long been recognized as a second approach to modulate demand response. Whereas traditional energy efficiency programs aim to reduce overall consumption, load shifting focuses specifically on changing the time of day when energy is used in order to favor times when costs are lower. Programs that focus on load shifting typically require mechanisms such as time-of-use (TOU) pricing or real-time pricing to induce transient changes in consumer behavior, such as those described by Vardakas [4]. TOU and seasonal rates focus on the customer's response to simple static price signals [5]. The Electric Power Research Institute (EPRI) carried out a major study of the top five experiments in the United States in the early 1980s and concluded that consumers responded to higher prices by shifting some of their load to off-peak periods [6]. Later experiments produced similar results. The City of Anaheim Public Utilities conducted a residential dynamic pricing experiment and found that for a peak-time rebate of $0.35/kWh they could reduce electricity use by 12% during critical-peak days [7]. California's Advanced Demand Response System pilot program used a critical peak pricing (CPP) tariff using the GoodWatts system to obtain peak reductions as high as 51% on event days with a CPP rate and 32% on non-event days with TOU rate. Enabling technology was identified as an important driver for load reductions [8]. This observation was also made in the Olympic Peninsula Project, where both TOU and real-time price (RTP) tariff were tested [9]. Similar results over a large number of studies have been widely reported and are summarized in a survey published by Faruqui et al. [8].

Since the introduction of *homeostatic utility control* by Schweppe et al. [10], it has been understood that key system state variables such as frequency and voltage in large-scale interconnections could be regulated using price signals. Prices have since been used primarily to schedule and dispatch generation resources using power markets [11]. Both energy efficiency programs and time-of-use rates have consistently been shown to effectively reduce loads on time-scales greater than one hour [12].

To avoid unfair pricing in the presence of demand response, David et al. [13] and later Kirschen et al. [14] examined how the elasticity of demand could be considered in wholesale scheduling systems. Initial work applying market-based mechanisms to building control systems showed that the notion of market-based demand control was feasible and effective for more granular systems [15]. The general concept of transactive control was initially proposed [16, 17] as a method of coordinating very large numbers of small resources using market-like signals at the electricity distribution level. The theory is essentially the same as for wholesale markets. However, realizations can be quite different insofar as more frequently updated price signals are typically used to manage distribution system constraints such as feeder capacity limits. These prices can dispatch both distributed generation, energy storage and demand response resources at much higher temporal and physical granularity than is possible with wholesale markets.

A number of previous studies have considered the operational impact of using retail price signals for controlling load in electric power systems. Glavitsch et al [18] showed that nodal pricing could find a socially optimal operating point for power markets . Following up on this work Alvarado [19] considered the question of whether power systems could be controlled entirely using prices, and found that price signals could indeed work. But the results came with some caveats, the most significant of which is the question of stability of the feedback control over the entire system.

The feasibility of transactive control methods has been proven out through a number of field demonstrations of automated distributed generation. The 2007 Olympic Peninsula demonstration [20] and the 2013 Columbus Ohio demonstration [21] are two examples of demand response control systems that dispatch distribution-level resources in quasi real-time using price signals. These experiments yielded a trove of

high-resolution data about the behavior of load resources in response to short-term price variations.

Overall, two important lessons have been learned from decades of utility research, development, and field experimentation with demand response [4]:

1. Consumer interest and sustained participation is essential to the success of demand response programs. Too many programs showed too little consumer interest and participation. This drives up program costs and reduces effectiveness. Tools to keep customers engaged and responsive to utility priorities are needed. Substantive contract diversity and meaningful incentives need to be available for customers to choose and actively engage in programs.

2. Programs should not provide rewards and incentives on the basis of complex baseline or reference models. Mechanisms that provide or enable endogenous sources of counter factual prices and quantities should be preferred by utilities.

Although transactive systems are similar to wholesale markets, the price signals are applied to different resources, affect consumer needs differently and are applied at much higher temporal and physical granularity than is possible in wholesale markets. Short-term consumer response to price variations is also understood to be quite distinct from long-term demand response. Long-term demand response is typically associated with changes in consumer behavior and the conversion to more energy efficient houses and appliances.

On the other hand, short-term demand response is primarily in the form of time-shifting and often requires automation. Short term demand response can be very different from long-term demand response because controllable load resources can be quickly exhausted, leading to control saturation. As a result, short and long term consumer responses are not generally comparable. In practice, long-term demand response models tend to underestimate the magnitude of the controllable resources and overestimate their endurance [20, 21] . This has two important consequences: i) Planning studies tend to undervalue the potential contribution of short-term demand response system and is often overlooked in utility program development; and ii) when it is not overlooked, the open-loop control gain is underestimated, resulting in over-control, instability and excessive price volatility.

The lack of solid theoretical basis for short-term performance claims has emerged as a significant challenge [22, 23]. Using static long-term own-price elasticities can be expected to give rise to erroneous short-term demand response control because short-term elasticities are more often substitution elasticities where the substitute is obtained in time rather than by an alternative product, a distinction which was made evident by Fan's study of Australian price elasticities [24], among others. Own-price elasticities represent averages over long periods of time. These averages may fail to capture the magnitude and variability possible at any given time. For example, Reiss and White [25] developed a household electricity demand model for assessing the effects of rate structure change in California and found that a small fraction of households respond to the price changes with elasticities range as large as -2, which far exceeds the average long-term elasticity of -0.14 found in Faruqui's survey of DR programs [12]. Unfortunately, computing the elasticity of demand for short-term demand response to real-time prices has proved challenging because the counterfactual price and demand are difficult to determine in the absence of a short-term feedback signal that elucidates the loads' willingness to pay [26]. Thus, more accurate models of short-term demand response are necessary for utility load control planning and design if these systems are to be deployed effectively.

The first contribution of this paper is the development of an analytical function for short-term demand in residential thermostatic loads that are responsive to real-time prices. The development of the demand function reflects first principles regarding the nature of thermostatic load control. We show that this model reduces to the Random Utility Model (RUM) employed in economics to study consumer choices and the valuation of non-market goods [27].

The second contribution of the paper is a validation of the model against data obtained from the Olympic and Columbus field demonstrations conducted by the US Department of Energy. These demonstration

projects implemented residential level thermostatic inputs to a price-based market clearing transactive control system on a five minute time-scale. The results of the field demonstrations show that customers could exhibit positive short-term demand response to short-term price variations. We show that the demand model can be easily calibrated to give an accurate representation of the market data from these experiments.

As a third contribution, the model is compared to four alternative models of demand response (DR): no DR, half DR, full DR, and demand elasticity from Faruqui's 2010 survey of DR programs [8]. For all models, we compute the error in predicting the amount of load shed at the 5-minute real-time price produced by the double-auctions of the Olympic and Columbus experiments. The results show that the RUM outperforms the alternative models for common "steady-state" demand conditions. In more extreme situations where load state diversity is low or when large price deviations occur over a very short time frame, the performance of the RUM is comparable to that of the competing models.

The derivation of the short-term demand response model is presented in Section 2, followed by its validation with field demonstration data in Section 3. A discussion follows in Section 4 on what the newly gained understanding of short term demand response might mean in terms of technology development, consumer acceptance, regulatory policy, and research opportunities.

2. Random Utility Model

Short-term consumer response to price variations is generally regarded as quite distinct from long-term demand response. The primary difference stems from the fact that long-term demand response is typically associated with enduring changes in consumer habits, whereas short-term demand response usually requires automation to support temporary changes in consumer behavior when prices are high. As a result, the consumer's considerations when whether and how to respond are not generally comparable, nor are they necessarily mutually exclusive.

Unfortunately, most studies of demand response in the electricity sector have focused on the static long-term elasticity of consumer demand [12]. Lacking alternative sources for short-term demand elasticity measures, utilities tend to use existing long-term elasticities as the basis for load control program evaluation and control systems design. Two important consequences arise from any discrepancy between the two elasticities.

1. Over or under estimation of the program value: if the short-term elasticity is greater than the long-term elasticity, then an indirect load control program would tend to be under-valued and would be less likely adopted.

2. Over or under estimation of control gain: if the short-term elasticity is greater than the long-term elasticity, then any attempt to mitigate instability from feedback signals would likely underestimate the open-loop control gain and would result in incorrect design of the closed-loop control system. This can potentially lead to less stable system operations and higher price volatility.

These discrepancies, and the results of the Olympic and Columbus studies [20] [21] showing that customers could exhibit positive short-term demand response, are the primary motivations for developing a new model of short term electricity demand.[1]

---

[1]It is worth noting that the effect of short-term demand response does not necessarily result in lower energy demand. It often results in lower peak load at the expense of increased total energy use. The reason is that consumers tend to respond to variations in price about a mean or expected price, increasing demand when prices are low and decreasing it when prices are high. Comfort is achieved by "storing" thermal resources (i.e., heat or cool) during low price periods and releasing it during high price period. Because the store/release process is not expected to be 100% efficient [28] any price-induced reduction in peak is typically associated with an increase in total energy use.

2.1. First-principles Model

We model the general behavior of thermostats governed by consumer preference based on an engineering model of houses' responses in the short-term given a consumer's static setting for comfort. The average duty cycle is based on the fraction of time the system is *on* relative to the total cycle time. For a thermostat operating within its deadband, the fraction is very closely approximated by

$$\rho = \frac{t_{on}}{t_{on} + t_{off}} \tag{1a}$$

where

$$t_{on} = \frac{1}{r} \ln \frac{T_{set} - \frac{1}{2}T_{hys} - T_{on}}{T_{set} + \frac{1}{2}T_{hys} - T_{on}} \tag{1b}$$

and

$$t_{off} = \frac{1}{r} \ln \frac{T_{set} + \frac{1}{2}T_{hys} - T_{off}}{T_{set} - \frac{1}{2}T_{hys} - T_{off}}. \tag{1c}$$

In these equations, r is the indoor air temperature decay rate time constant, T_{set} is the temperature set-point, T_{hys} is the thermostat's hysteresis, T_{on} is the steady-state temperature when the heating system is *on* (or cooling is *off*), and T_{off} is the steady-state temperature when the heating system is *off* (or cooling is *on*). This duty cycle corresponds to the probability that we observe a device to be *on* at any given time.

The transactive control system assumes thermostats submit bids such that the probability of clearing a lower retail electricity price is the duty cycle required to maintain consumer comfort. Thus the system can be expected to run with the duty cycle needed while preferentially running when retail electricity prices are lower. This comfort tracking cost minimizing strategy is embodied in the bid-response function

$$P = \bar{P} + K \frac{\tilde{P}}{T_{obs} - T_{set}} \tag{2}$$

where $\bar{P}$ is the expectation value for the clearing price, $\tilde{P}$ is the standard deviation, T_{obs} is the actual indoor air temperature, and K is the customer's comfort control setting. The consumer-controlled variable K expresses how sensitive the household is to the trade-off between money (the cost of energy) and comfort (distance from ideal temperature). A high value of K signals a higher sensitivity to price fluctuations. It gives the customer more opportunities to reduce costs at the expense of reduced short-term comfort, embodied by the thermostat's tracking error $T_{obs} - T_{set}$.

For a population of N thermostats with mean duty cycle $\bar{\rho}$, the probability of finding k devices in the *on* state is proportional to the binomial distribution of the count

$$g(k) = \frac{N!}{k!(N-k)!} \rho^k (1-\rho)^{N-k}. \tag{3}$$

The design of thermostatic loads requires significant oversizing of equipment, so we assume that on peak the mean duty cycle $\bar{\rho} = 0.5$. Given the mean thermostatic device load $\bar{q}$ we can define the load deviation $Q = (k - \frac{1}{2}N)\bar{q}$ from the most probable load $\bar{Q}$ and the total thermostatic load $Q_R = N\bar{q}$. The aggregate load entropy is then given by

$$\lim_{\rho \to 0.5} \sigma(Q) = \sigma_0 - 2\frac{Q^2}{Q_R} \tag{4}$$

where

$$\sigma_0 = \frac{Q_R}{\bar{q}} \ln \frac{Q_R}{2\bar{q}} - \frac{1}{2} \ln 2\pi$$

is the maximum entropy corresponding to the most probable load $\bar{Q} = \bar{\rho} Q_R$. The probability of observing

any given load Q is the probability

$$2^{-N} g(k) = \frac{1}{1 + e^{-2\sigma(Q)}}. \tag{5}$$

Using the standard definition of demand elasticity for a load Q at price P we obtain

$$\eta = \frac{P}{Q} \frac{\partial Q}{\partial P} = P \frac{\partial \ln Q}{\partial P} = P \frac{\partial \sigma(Q)}{\partial P}. \tag{6}$$

Observe that the maximum demand responsiveness dQ/dP occurs when the entropy is at its maximum. Therefore at the most probable price and load it must be that

$$\left. \frac{\partial^2 Q}{\partial P^2} \right|_{\substack{P = \bar{P} \\ Q = \bar{Q}}} = \left. \frac{\partial \sigma(Q)}{\partial Q} \right|_{Q = \bar{Q}} = 0.$$

Integrating twice we find

$$\sigma(\bar{Q}) = a + b\bar{P}.$$

where a and b are unknown constants which must be determined from boundary conditions or from a fit to data. We substitute this result into Eq. (5) and deduce that the load as a function of price is

$$Q(P) = \frac{Q_R}{1 + e^{a+bP}} + Q_U \tag{7}$$

where Q_U is the unresponsive load not subject to price-responsive behavior.

2.2. Model Assumptions

This model has the same form as McFadden's random utility model [27]. The RUM has been used extensively in economics to study consumer choice and in the valuation of non-market goods [29]. That it can essentially be derived independently from the first principles of thermostatic controls establishes a strong tie between the engineering approach and state of the art economic modeling of consumer preferences.[2]

The random utility model (and thus subsequent manipulations below) makes two important assumptions.

1. A consumers' choice is a discrete event in the sense that a device acting on the consumer's behalf must make an all-or-nothing decision. The consumer can either run or not run an air-conditioner. The device cannot be run at part-load for the next interval.

2. The consumer's (or device's) attraction to a particular choice is affected by a random error with a type 1 extreme value (Gumbel) distribution. In this case we use the term *attraction* in the retailing sense but we could just as well use the term *utility* to be consistent with economic theory. The randomness of the utility's observation of the current comfort preference is assumed to arise from the devices acting on behalf of consumers. The devices rationally choose the outcomes with the highest utility based on the consumer's indicated preference for comfort.

Although the random utility model has been derived using various methods, McFadden points out that according to Luce [30] it is axiomatic that the relative odds in a binary choice will remain the same for independent alternatives when additional alternatives become available. Therefore, the selection probability can always be written in the form

$$P_i \frac{e^{\nu(z_i)}}{\sum_{n=1}^{N} e^{\nu(z_n)}}$$

---

[2]McFadden received the 2000 Nobel prize in economics for his pioneering work on economic choices.

where $\nu(z)$ are scale functions of the stimulus z. When ν is linear in parameters, this result is the multinomial logit formula found in the standard statistical literature.

In the absence of prior knowledge of the quantities demanded by consumers, the derivation of the aggregate demand curve is based on this discrete choice statistic for consumer demand [31]. Thermostats act as agents on behalf of consumers whose utility is assumed to be comprised of an observable component based on their comfort setting, which follows an extreme value distribution and an unobservable component that has zero mean from the perspective of the market. Thermostats are expected to bid a price that will give the heating/cooling system a probability of running satisfying the duty cycle required to maintain indoor air temperature. A thermostat's response to a market clearing price is an exclusive choice based solely on the bid it submitted [32]. For a dichotomous choice, the reasoning is as follows: U is the consumer benefit (*utility* in economic theory) that the thermostat obtains from taking a particular action given the consumer's preferences. This net benefit depends on an unobservable characteristic α that has a zero mean distribution and an observable characteristic β that is a known decreasing function of price. The net benefit is defined as $U = \beta x + \alpha$ where x is the binary choice ($x = 1$ or 0) β represent the marginal utility of the anticipated change in comfort and cost of electricity, and α represents all other unobservable variables that can give rise to error in choices. The action corresponding to $x = 1$ is taken if $U > 0$.

From a logistic regression we find that the probability of taking the action is then

$$\rho\{x\} = \frac{1}{1 + e^{-\beta x}}. \tag{8}$$

The optimal consumer bid from Eq. (2) is the utility maximizing price. The random utility model gives the same result as the load probability in Eq. (5).

2.3. Equilibrium Demand Response

The transactive control system used in the demonstration projects is quiescent when load state diversity is maximized and the total load is steady. This steady-state condition occurs when the distribution of bids is symmetric about the mean price with the same relative variance. We assume that states are uniformly distributed over the thermostat deadbands because without price disturbances, thermostatic devices follow the standard duty cycle regime. Devices will be *on* for the time required to rise from the *on* boundary of the thermostat deadband to the *off* boundary and *off* for the time required to go the other way. Undiversified states will tend to randomize under the influence of diverse physical parameters and state diversity grows until the thermostats settle into the equilibrium demand regime. After re-diversification, prices return to obeying the logistic distribution of Equation (8). We rescale the physical quantities for an arbitrary system with Q_U unresponsive load and Q_R responsive load at the prices p. Finally we rewrite Eq. (8) to obtain Eq. (7) again, where a and b are the demand curve's shape parameters.

The most probable demand elasticity at steady state occurs at maximum diversity when $p = -\frac{a}{b}$ and

$$\hat{\eta}_D = \eta_D\left(-\frac{a}{b}\right) = \frac{a}{2} \tag{9}$$

as shown in Figure 1 (left). Using this we can estimate the demand function parameters for any set of N bids by fitting a linear function to the bids within the central 60% of the demand response range, i.e., from $0.2Q_R$ to $0.8Q_R$. The most probable price $\hat{P}$ is found at the mid-point $\hat{Q} = Q_U + \frac{1}{2}Q_R$. The demand elasticity $\eta_D = 2\hat{P}/Q_R d$ where d is the demand response slope obtained from the linear fit. From this we find the curve parameters

$$a = 2\eta_D \quad \text{and} \quad b = -a/\hat{P} \tag{10}$$

which can be obtained by applying the definition of elasticity to Eq. (7) such that $\eta_D = \frac{p}{Q(p)}\frac{dQ(p)}{dp}\big|_{p=\hat{P}}$ and observing that the most probable price occurs when $p = -a/b$.
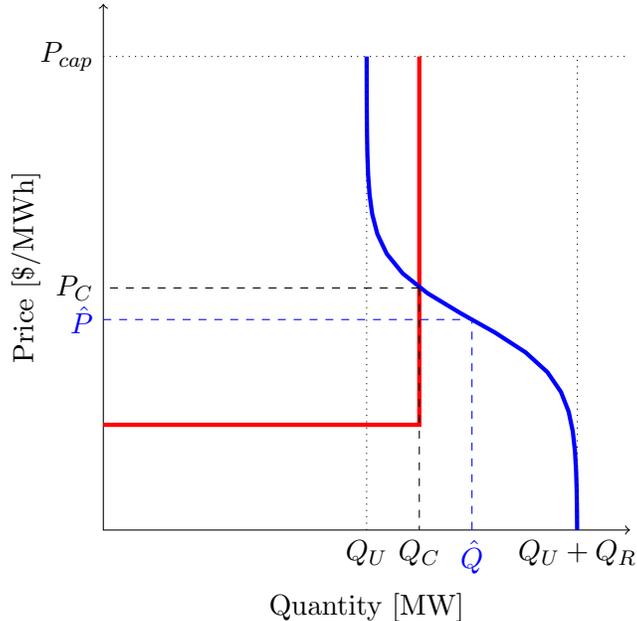
Figure 1: Demand curves (left) and demand elasticities for steady state thermostatic loads

3. Model Validation

To test the validity of the Random Utility Model, we compare its performance to that of models currently employed by utilities. The four comparator models are based on three different levels of static participation from the installed demand response capacity (i.e., 0%, 50% and 100% of $Q_R - Q_U$), as well as the -0.14 long-term demand elasticity proposed by Faruqui [12]. The key difference between the different models used by utilities is the assumption made about the short term elasticity of demand:

1. **Zero elasticity (No DR)** may be preferred when the demand response program is not expected to be a significant fraction of the total load. However, in such a case the open-loop control gain is essentially zero, and both program valuation and feedback control design are not possible. Depending on the shape of the demand curve, zero elasticity may arise at more than one quantity.

2. **Maximum elasticity (Full DR)** may be preferred when the purpose of the study is to design the feedback control so as to avoid instability. The maximum elasticity would correspond to the maximum open-loop gain and the choice of closed-loop gain would therefore be such that instability could be avoided for all physically realizable load conditions.

3. **Most probable elasticity (Half DR)** may be preferred when the purpose of the study is to evaluate a load control program's long-term value. Depending on the shape of the demand curve, the most probable elasticity may also be the maximum elasticity.

We supplement these standard models by adding a fourth comparator:

4. **Elasticity $\eta = -0.14$**. This last model imposes the long term price elasticity estimated by Faruqui [12].

It can be difficult to augment these static elasticity models with what is known about consumer choices in energy purchasing. In a study published in 2000 [33], customers were found to consider a long-term marginal energy price increase more onerous when the price is low than when it is high. However, energy consumers do not go so far as to consider price changes strictly in proportional terms either. In addition, the authors noted that bonus or coupon inducements can affect consumer choices. But these inducements
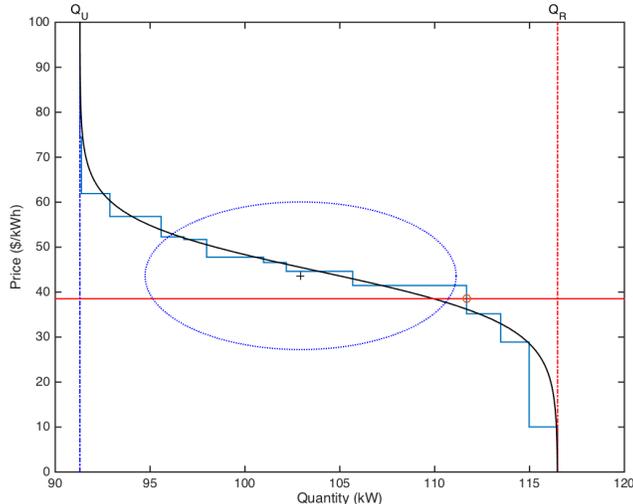
8

Figure 2: Example of demand function model validation with Columbus demonstration data. The bids shown are from 2013-06-22 22:45 EDT. The clearing price P_C and quantity Q_C are indicated by the circle. The expected price $\bar{P}$ and quantity $\bar{Q}$ are indicated by the plus sign. The standard deviation of price $\tilde{P}$ and quantity $\tilde{Q}$ are indicated by the ellipse.

Table 1: Feeder characteristics

| | | | Demand Response | |
Feeder Id	Customers	Bids	kW	(% peak load)
Olympic (heating)				
–	38	1174923	1193	(52.4)
Columbus (cooling)				
120	11	281045	20	(0.7)
140	30	699241	45	(1.0)
160	53	1478148	83	(1.2)
180	8	210241	12	(0.2)

are only relevant to long-term decisions such as tariff or supplier choice. Other variables often considered by consumers include contract duration (where long terms are viewed negatively), variable rates (also viewed negatively) with shorter-term fluctuations being viewed more negatively.

The demand function model proposed in Section 2 was validated using data obtained from the Pacific Northwest SmartGrid Demonstration Testbed (Olympic) project and the Northeast Columbus gridSMART Demonstration (Columbus) project. Bids received in both projects were fit using the method described above. An example of this fitting process is shown in Figure 2. The Olympic data set includes 103842 market clearing events from April 01, 2006 to March 31, 2007. A total of 1174923 bids were received from 38 customers. Real-time prices were computed from 5-minute double auction on a single feeder. The Columbus data set includes four separate feeders with market clearing events from June 01, 2013 to September 30, 2013. The data for both studies is summarized in Table 1.

The performance of the RUM and four alternative models are evaluated by comparing the quantity predicted by each model at the observed market price to the actual quantity observed. The -0.14 long-term elasticity is unlikely to be a good approximation for fast-acting demand response, but it is provides a clear indication of the errors potentially introduced when using long term elasticities in studies of short term demand response.

The results of the Olympic experiment are presented in Table 2. We observe that the random utility model outperforms the alternative models for all performance metrics. The difference is particularly significant for the mean error and bias error, but less significant for the standard deviation. Results for

9

Table 2: Olympic data analysis results

Model	Error Mean	Bias	Standard Deviation
Random Utility	−0.35%	1.74%	7.24%
Half DR	−0.67%	5.63%	10.10%
Full DR	−7.38%	6.83%	14.14%
No DR	6.04%	5.49%	8.89%
$\eta = -0.14$	6.62%	6.83%	14.14%

Table 3: Columbus analysis results for Feeders 120, 140, 160 and 180

Feeder 120 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.78%	0.24%	0.31%
Half DR	0.75%	0.25%	0.33%
Full DR	0.53%	0.24%	0.30%
No DR	0.98%	0.29%	0.37%
$\eta = -0.14$	14.53%	0.24%	0.30%

Feeder 140 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.97%	0.29%	0.36%
Half DR	0.99%	0.33%	0.41%
Full DR	0.71%	0.29%	0.37%
No DR	1.27%	0.38%	0.48%
$\eta = -0.14$	14.71%	0.29%	0.37%

Feeder 160 Model	Error Mean	Bias	Standard Deviation
Random Utility	1.00%	0.28%	0.36%
Half DR	1.00%	0.33%	0.41%
Full DR	0.81%	0.31%	0.39%
No DR	1.19%	0.36%	0.45%
$\eta = -0.14$	14.81%	0.31%	0.39%

Feeder 180 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.61%	0.22%	0.26%
Half DR	0.44%	0.18%	0.24%
Full DR	0.34%	0.15%	0.20%
No DR	0.53%	0.22%	0.28%
$\eta = -0.14$	14.34%	0.15%	0.20%

the full Columbus data sets are presented in Table 3. They are mixed. We note that the Full DR model outperforms the RUM for mean error. On the other hand, the RUM outperforms the Full DR model for bias error and the standard deviation on Feeder 160 (the largest and most diverse in terms of number of consumers). The picture that emerges out of Columbus is that the RUM model performs best in some

Table 4: Columbus analysis results for only non-experiment days

Feeder 120 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.78%	0.25%	0.31%
Half DR	0.74%	0.26%	0.33%
Full DR	0.52%	0.24%	0.30%
No DR	0.97%	0.29%	0.37%
$\eta = -0.14$	14.52%	0.24%	0.30%

Feeder 140 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.79%	0.21%	0.26%
Half DR	0.77%	0.21%	0.28%
Full DR	0.52%	0.20%	0.26%
No DR	1.02%	0.24%	0.31%
$\eta = -0.14$	14.52%	0.20%	0.26%

Feeder 160 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.67%	0.24%	0.30%
Half DR	0.59%	0.30%	0.38%
Full DR	0.34%	0.28%	0.37%
No DR	0.84%	0.33%	0.42%
$\eta = -0.14$	14.34%	0.28%	0.37%

Feeder 180 Model	Error Mean	Bias	Standard Deviation
Random Utility	0.69%	0.19%	0.22%
Half DR	0.58%	0.18%	0.24%
Full DR	0.45%	0.15%	0.21%
No DR	0.70%	0.21%	0.27%
$\eta = -0.14$	14.45%	0.15%	0.21%

instances but is comparable to the static models overall.

There is, however, an important qualifier to the Columbus results. Saturation of demand response (either all *on* or all *off*) is associated with diminishing load state diversity. This violates the assumption of the RUM. For this reason, we expect feeders that frequently saturate the demand response resource control not to be well represented by the RUM. These feeders would produce data more often consistent with the No DR or Full DR models (depending on conditions). Unlike the Olympic study, the experimental protocol for the Columbus study deliberately probed these limits of control every other day. This resulted in frequent loss of load state diversity in violation of the zero-mean assumption.

The experimental protocol in the Columbus demonstration is expected to have introduced additional errors. This has been verified to a first order by analyzing the data excluding the experiment days. The results are shown in Table 4, where the error on Feeder 160 is reduced from 1.00% to 0.67%. However, a second-order effect is now observed insofar as the Full DR model seems to still perform better than the random utility model with the error reduced from 0.81% to 0.34%. This can be explained by the second-day recovery during which thermostats receive relatively lower prices compared to the previous

day and tend to respond more aggressively to them. This hypothesis cannot be directly verified as the experiment protocol did not normally include a third day during which neither an experiment nor a recovery was taking place.

The model as presented is valid only for steady-state conditions. We therefore ought to consider whether transient demand response behavior influences the accuracy of the random utility model. Two factors are known to influence the magnitude of the demand response transient: (i) the fraction of devices that respond to the price signal, and (ii) the diversity of device states when a price signal is received.

The results suggest that although the random utility model is valid for predicting steady-state demand response behavior, its accuracy is limited in the wake of large magnitude price fluctuations that tend to drive a significant majority of responding devices to a single common state. This kind of state degeneracy violates the parameter distribution assumptions of the random utility model and reduces its accuracy for predicting the load after an abrupt large-magnitude change in price is observed.

In general, we expect normal utility operations to be more like the Olympic conditions than the Columbus conditions. We thus conclude that while the Columbus results only weakly support the random utility model, they mainly point out the importance of the steady-state assumptions in Section 2.

4. Discussion

The availability of a more accurate model of aggregated demand response can be expected to support a wide range of new work on controllable load using real-time pricing. Long-term demand response behavior models do not support the design and analysis of fast-acting demand response as well as the proposed short-term model. In this section we discuss the advantages of using short-term demand response models. As an equilibrium model, the random utility model is expected to be valid for both small signal control stability analysis as well as certain tariff design problems.

4.1. Model Limitations

The random utility model is not necessarily valid in its current form for large price disturbance. Roozbenahi et al. examined the feedback stability question in the context of wholesale markets [34] and found that real-time wholesale prices could create an unstable closed-loop feedback system for both ex-ante and ex-post settlement systems. It was established that the absence of an inelastic component in demand contributed to instability, supporting the intuition that increasing feedback gain from price-responsive demand is a concern. Static demand elasticity was also found to lead to loss of efficiency. In follow-up work on price volatility, the authors found that although demand bidding mechanisms eliminate the exogenous feedback delays inherent in settlement-based systems, there remain endogenous load dynamics that can cause bidding mechanisms to exhibit instability [35] and consequently more sophisticated models of demand and consumer response to real-time price dynamics may be required.

The random utility model does not account for the aggregate equilibrium duty cycle of thermostatic loads. In its simplest form, the model assumes a 50% effective duty cycle. The duty cycle tends to skew the demand curve away from the more probable load $\hat{Q} = Q_U + \frac{1}{2}Q_R$. Incorporating this effect would most likely improve the model, particularly with respect to bias errors and the standard deviation.

When a significant price deviation occurs relative to the natural diversity state, the loads enter a transient response regime. If we compare maximum entropy from Eq. (5) to the minimum elasticity from Eq. (6) we find that they occur at different prices. Specifically, load state diversity is maximized when $P = \bar{P}$, but demand elasticity

$$\eta(P) = \frac{-bPe^{a+bP}}{1 + e^{a+bP}}.$$

is minimized when
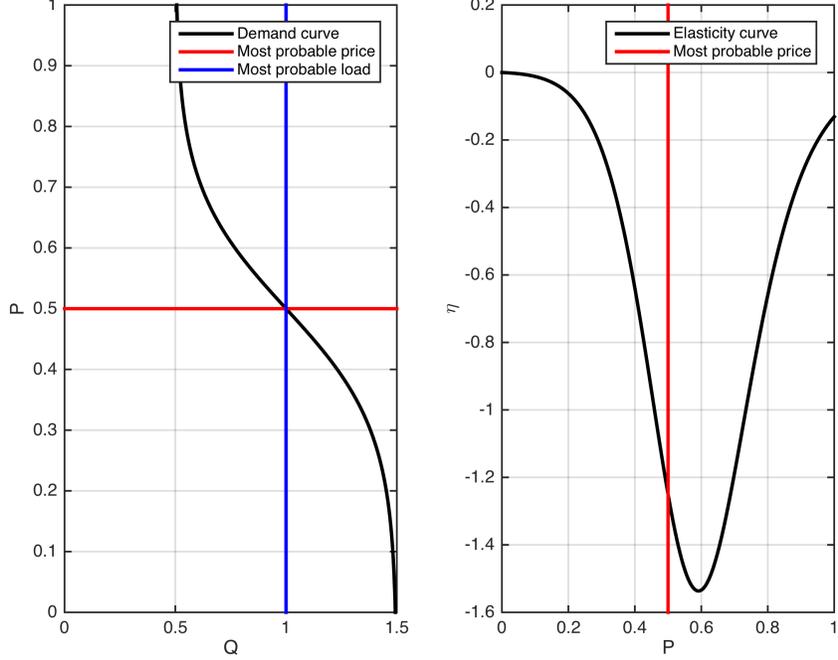
$$P = \bar{P}\frac{W(e^{2\eta_D-1}) + 1}{2\eta_D}$$

Figure 3: Demand (left) and elasticity (right) curves for a nominal case with $\eta_D = -2.5$ and $\bar{P} = 0.5$.

where W is the Lambert W-function

$$W(z) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} z^n.$$

For values of z approaching zero this function is well approximated by z, giving:

$$P \approx \bar{P} \frac{e^{2\eta_D - 1} + 1}{2\eta_D}$$

The price at which elasticity is minimized is always greater than $\bar{P}$ for all $0 < Q < Q_R$ when $Q_U > 0$. This implies that in equilibrium demand elasticity will tend to increase with prices, as illustrated in Figure 3. Under such conditions thermostatic devices no longer follow the equilibrium duty cycle regime and their states diverge from the equilibrium distribution. Consequently their bids depart from the logistic probabilities and no longer follow the bid price distribution of Equation (8).

In addition, if the price deviates too quickly, then a diabatic response governs the change in state diversity. As diversity decreases the equilibrium price moves further from the most probable price and the elasticity of demand changes significantly. Decreasing elasticity is observed when loss of diversity favors loads that are *on* and the bid distribution skews left. The distribution skews to the right with increased elasticity when diversity favors loads that are *off*. Note that the periodic behavior of thermostatic loads means that diversity is expected to fluctuate in such a way that elasticity oscillates with damping of about $a/2$ arising from the diversity in the thermal properties of the home and a frequency related to the population average cycling time of the heating/cooling systems.

Device state diversity appears to be a key characteristic that governs demand response. True state diversity can be measured by taking the weighted generalized mean M_{q-1} of the proportional occupancy of states in the population of responsive devices, and then taking the reciprocal of this quantity to obtain
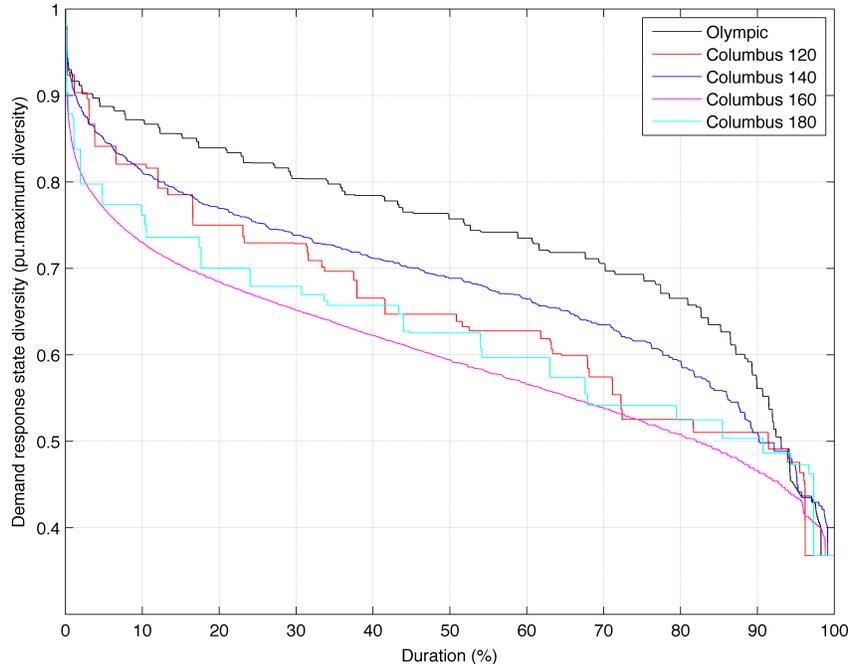
13

Figure 4: Demand response state diversity duration curves for the Olympic feeder and Columbus feeder numbers 120, 140, 160, and 180.

the density of devices in states. The diversity of order-q is then defined as

$$D_q \equiv \frac{1}{M_{q-1}} = \left(\sum_{n=1}^{N} p_n^q n \right)^{1-q^{-1}}. \tag{11a}$$

In the limit of $q = 1$, the first-order mean occupancy is well-defined and its natural logarithm converges to

$$H = - \sum_{n=1}^{N} p_n \ln p_n \tag{11b}$$

This is simply the Shannon entropy calculated using natural logarithms instead of base-2 logarithms [36].

Given the linear relationship of bid price to the device state, we use the bid price entropy norm as a measure of state diversity, as shown in Table 5. A higher bid price entropy is associated with a higher state diversity. This further explains why the Olympic results fit the equilibrium state assumptions of the random utility model so much better than the Columbus results.

If the demand response resource is very limited it can be expected to saturate more often and lead to reduced diversity and reduced entropy. This effect is illustrated by the reduced diversity duration curves of the Columbus experiments shown in Figure 4. In such cases, we expect the full DR and no DR models to be as accurate as the random utility model when DR is called and released, respectively and the half DR model to be more accurate when DR is not called. This condition is more clearly evident in the Columbus data where the total resource was relatively small compared to the total feeder capacity. The observed fluctuations in unresponsive load result in large changes in demand response and lead to DR control saturation, thus making the alternative models satisfactory when compared to the random utility model.

14

Table 5: Bid price entropy statistics

| System | Load State Entropy | |
	Mean	Standard Deviation
Olympic	0.68	0.19
Columbus 120	0.54	0.20
Columbus 140	0.60	0.18
Columbus 160	0.47	0.17
Columbus 180	0.51	0.17

4.2. Technical and Regulatory Impacts

Interest in transactive control has also resulted in renewed discussion of tariff design and rate-making processes for utilities that wish to adopt the real-time pricing strategy. The Columbus demonstration included a rate case for the real-time price double auction called "RTPda" that was approved by the Public Utility Commission of Ohio [37, 38]. However, real-time price tariffs and rate designs can be difficult to study when loads respond to hourly or sub-hourly prices. These tariffs present new challenges for utilities and regulators alike. Concerns have also been expressed regarding customer acceptance of real-time prices [12] but utilities could offer a portfolio of tariffs, including a real-time price, from which customer may choose. The rate at which tariffs are adopted then becomes a conventional portfolio optimization problem where opt-in/opt-out incentives and penalties are offered to achieve tariff adoption mixtures that meet utility and regulator objectives [39].

The implications of adopting the random utility demand response model for both utilities and regulations have yet to be explored fully. However, some initial thoughts may foster discussion and suggest further research on methods to effectively incentivizes enduring and sustainable demand response from residential customers.

First, we believe that the existence of upper and lower price asymptotes is often overlooked in demand response analysis for electricity loads. We recognize that linearization of the demand function is often necessary. But the asymptotes require us to acknowledge the existence of hard upper and lower constraints on short-term responsive load. When prices deviate from the most probable price, the elasticity of demand approaches zero. Thus an important change occurs in fast-acting demand response when prices induce load to move into either curtailment or pre-heat/pre-cool/recovery regimes: load diversity is decreased and important endogenous load oscillations can be induced, which are independent of the oscillations induced by control feedback. Unless technical steps are taken to dampen endogenous load oscillations before they occur, price oscillations can emerge which can only be mitigated by reducing the feedback from load and waiting for load state diversity to be restored. It is therefore incumbent on utilities to identify all the conditions under which instability can emerge and implement either market and/or load control mechanisms to mitigate them.

Second, we believe that existing residential electricity tariffs do not adequately support the development of fast-acting demand response or the mechanisms needed to mitigate the potential instabilities associated with demand response. This is particularly a concern in the presence of supply resources at the distribution level, such as rooftop photovoltaic panels, and significant amounts of battery storage, such as so-called vehicle-to-grid discharging. Clearly the availability of such resources can be easily incorporated into the market mechanism demonstrated in both the Olympic and Columbus trials. However, the marginal cost of most renewable retail resources is effectively zero, which not only can give rise to revenue adequacy problems for utilities, but can also effectively shut off the very price signaling mechanism needed to control load. Strategic bidding may be necessary for these resources to elicit non-zero price signals, as was demonstrated in the Olympic study to manage minimum and maximum generation runtime limits. However strategic bidding at the retail level entails an entirely new class of regulatory problems that may require mitigation strategies that are not part of current tariff design and approval procedures.

Demand response in residential settings presents an additional challenge to utilities. Historically, the

cost-per-point and cost-per-megawatt for the supporting infrastructure and management of these systems has been significantly higher than for commercial and industrial customers. Recent industry estimates suggest that customer premises portal and in-home energy management systems will cost between $US150 and $US300 by 2030 [40]. But life-cycle cost analysis for transactive technology is not generally available yet, in part due to the lack of simulation tools that can properly evaluate the benefits of the technology [41]. Fast-acting demand response requires higher communication rates than hourly or day-ahead critical peak pricing mechanisms used for commercial and industrial demand management. Automated metering infrastructure has offered the promise of fast and accurate communications with residential loads. But much of that promise has yet to be analysed in detail or realized in practice either with incentive-based or even direct load control mechanisms.

The cost per unit of power reduced of behind-the-meter infrastructure for residential loads has been typically difficult to compare to the cost per energy saved using long-term demand response programs. The additional costs per customer tends to delay adoption of residential demand response programs until after the more cost-effective available commercial and industrial demand resources have been exhausted. The advent of smart thermostats like the NEST$^{\text{TM}}$ and potentially implicitly smart loads like electric-vehicle chargers can be expected to increase the available low-cost responsive load in the residential settings [42].

5. Conclusions

We have developed a logistic demand curve for short term electricity consumption derived from the first principles of controllable thermostatic electric loads operating under the transactive control paradigm. We have shown that this model corresponds to the Random Utility Model commonly used in the economics of consumer choice. The model's performance is compared to results from two US Department of Energy demonstration projects in which short-term demand response data was obtained. We find that the random utility model predicts the total demand response to smaller price fluctuations very well, but that model performance degrades as the magnitude and frequency of price excursions increases and as the diversity of load states decreases. We conclude that the random utility model is suitable for demand response studies that utilize steady state conditions for most situations with only infrequent and modest price excursions. Future work will investigate models that account for varying effective duty cycles for large populations of thermostatic loads and account for the effect of large fast changes in prices.

The proposed model supports previous claims that additional research will be required to mitigate the potential instabilities that may emerge when employing real-time pricing signals for closed-loop feedback control of fast-acting residential demand response. This work will necessarily require a contribution from the field of control theory, while maintaining strong support from economists with an interest in mechanism design.

The model also highlights emerging challenges for tariff design and rate approval processes, particularly in cases where significant distributed generation and storage resources participate in price-discovery alongside fast-acting demand response. The question of incentive compatible retail market design can be expected to become more important as new tariffs and rate structures are developed by utilities and regulators.

In its present form the random utility model provides a robust framework that is well-founded in the engineering principles of how thermostatic devices behave in price-based control environments. By joining the engineering and economic behavior of such devices, the random utility model is set to become an essential element in the planning, design and eventual deployment of large-scale load control strategies.

References

[1] V. Smith and L. Kiesling, "Demand, Not Supply," *Wall Street Journal Online*, vol. 8, no. 20, pp. 1–3, 2003.

[2] S. Borenstein, "The trouble with electricity markets: understanding california's restructuring disaster," *Journal of economic perspectives*, pp. 191–211, 2002.

[3] W. W. Hogan, "Electricity market design and efficient pricing: Applications for new england and beyond," *The Electricity Journal*, vol. 27, no. 7, pp. 23–49, 2014.

[4] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, 2015.

[5] A. Faruqui and J. R. Malko, "The residential demand for electricity by time-of-use: a survey of twelve experiments with peak load pricing," *Energy*, vol. 8, no. 10, pp. 781–795, 1983.

[6] D. W. Caves, L. R. Christensen, and J. A. Herriges, "Consistency of residential customer response in time-of-use electricity pricing experiments," *Journal of Econometrics*, vol. 26, no. 1, pp. 179–203, 1984.

[7] F. Wolak, "Residential customer response to real-time pricing: the anaheim critical-peak pricing experiment. work. pap. 151, energy inst. cent. study energy mark," *Univ. Calif. Berkeley*, 2006.

[8] 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.

[9] L. Kiesling, "Digital technology, demand response, and customer choice: Efficiency benefits," in *NARUC Winter Meetings*, 2008.

[10] F. Schweppe, R. Tabors, J. Kirtley, H. Outhred, F. Pickel, and A. Cox, "Homeostatic Utility Control," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 3, pp. 1151–1163, 1980.

[11] S. Stoft, *Power System Economics*. Oxford Institute for Energy Studies, 2002.

[12] A. Faruqui, R. Hledik, and J. Tsoukalis, "The Power of Dynamic Pricing," *The Electricity Journal*, vol. 22, pp. 42–56, Apr. 2009.

[13] A. David and Y. Li, "Effect of inter-temporal factors on the real time pricing of electricity," *IEEE Transactions on Power Systems (Institute of Electrical and Electronics Engineers);(United States)*, vol. 8, no. 1, 1993.

[14] D. S. Kirschen, G. Strbac, P. Cumperayot, and D. de Paiva Mendes, "Factoring the elasticity of demand in electricity prices," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 612–617, 2000.

[15] B. A. Huberman and S. H. Clearwater, "A multi-agent system for controlling building environments," in *ICMAS*, pp. 171–176, 1995.

[16] D. P. Chassin, "The abstract machine model for transaction-based system control," Technical Report PNNL-12082, Pacific Northwest National Laboratory, November 2002.

[17] D. P. Chassin, J. M. Malard, C. Posse, A. Gangopadhyaya, N. Lu, S. Katipamula, and J. V. Mallow, "Modeling power systems as complex adaptive systems," Technical Report PNNL-14987, Pacific Northwest National Laboratory, December 2004.

[18] H. Glavitsch and F. Alvarado, "Management of multiple congested conditions in unbundled operation of a power system," in *Power Industry Computer Applications., 1997. 20th International Conference on*, pp. 374–380, IEEE, 1997.

[19] F. L. Alvarado, "Is system control entirely by price feasible?," in *null*, p. 53b, IEEE, 2003.

[20] D. J. Hammerstrom, R. Ambrosio, T. a. Carlon, J. G. Desteese, R. Kajfasz, R. G. Pratt, and D. P. Chassin, "Pacific Northwest GridWise(TM) Testbed Demonstration Projects Part I . Olympic Peninsula Project," Tech. Rep. PNNL-17167, Pacific Northwest National Laboratory, October 2007.

[21] S. Widergren, A. Somani, K. Subbarao, C. Marinovici, J. Fuller, J. Hammerstrom, and D. P. Chassin, "AEP Ohio gridSMART Demonstration Project Real-Time Pricing Demonstration Analysis," Tech. Rep. PNNL-23192, Pacific Northwest National Laboratory, February 2014.

[22] "U.S. Department of Energy LDRD Project List – FY 2014," tech. rep., United States Department of Energy, 2014.

[23] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, 2009.

[24] S. Fan and R. J. Hyndman, "The price elasticity of electricity demand in south australia," *Energy Policy*, vol. 39, no. 6, pp. 3709–3719, 2011.

[25] P. C. Reiss and M. W. White, "Household electricity demand, revisited," *The Review of Economic Studies*, vol. 72, no. 3, pp. 853–883, 2005.

[26] D. P. Chassin and L. Kiesling, "Decentralized coordination through digital technology, dynamic pricing, and customer-driven control: The gridwise testbed demonstration project," *The Electricity Journal*, vol. 21, no. 8, pp. 51–59, 2008.

[27] D. L. McFadden, "Quantal choice analaysis: A survey," in *Annals of Economic and Social Measurement, Volume 5, number 4*, pp. 363–390, NBER, 1976.

[28] I. Beil, I. Hiskens, and S. Backhaus, "Frequency regulation from commercial building hvac demand response," *Proceedings of the IEEE*, vol. 104, pp. 745–757, April 2016.

[29] K. G. Pillai and C. Hofacker, "Calibration of consumer knowledge of the web," *International Journal of Research in Marketing*, vol. 24, no. 3, pp. 254–267, 2007.

[30] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Wiley, New York, 1959.

[31] K. Train, *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. The MIT Press, 1993.

[32] D. P. Chassin, J. Stoustrup, P. Agathoklis, and N. Djilali, "A new thermostat for real-time price demand response: Cost, comfort and energy impacts of discrete-time control without deadband," *Applied Energy*, vol. 155, pp. 816 – 825, 2015.

[33] K. E. T. Andrew A. Goett, Kathleen Hudson, "Customers' choice among retail energy suppliers: The willingness-to-pay for service attributes," *The Energy Journal*, vol. 21, no. 4, pp. 1–28, 2000.

[34] M. Roozbehani, M. Dahleh, and S. Mitter, "On the stability of wholesale electricity markets under real-time pricing," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, pp. 1911–1918, IEEE, 2010.

[35] M. Roozbehani, M. A. Dahleh, and S. K. Mitter, "Volatility of power grids under real-time pricing," *Power Systems, IEEE Transactions on*, vol. 27, no. 4, pp. 1926–1940, 2012.

[36] M. O. Hill, "Diversity and evenness: a unifying notation and its consequences," *Ecology*, vol. 54, no. 2, pp. 427–432, 1973.

[37] A. Ohio, "gridSMART(SM) Demonstration Project," 2016. www.smartgrid.gov/project/aep_ohio_gridsmartsm_dem

[38] S. E. Widergren, M. C. Marinovici, J. C. Fuller, K. Subbarao, D. P. Chassin, and A. Somani, "Customer engagement in aep gridsmart residential transactive system," tech. rep., Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2014.

[39] D. P. Chassin, Y. Sun, and A. Somani, "Optimization of customer subscription rates to electric utility tariffs," in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pp. 2604–2609, IEEE, 2015.

[40] C. Gellings, "Estimating the costs and benefits of the smart grid," Technical Report 1022519, Electric Power Research Institute, March 2011.

[41] D. J. Hammerstrom, C. D. Corbin, N. Fernandez, J. S. Homer, A. Makhmalbaf, R. G. Pratt, and A. Somani, "Valuation of transactive systems," Technical Report PNNL-25323, Pacific Northwest National Laboratory, May 2016.

[42] S. Behboodi, D. P. Chassin, C. Crawford, and N. Djilali, "Electric vehicle participation in transactive power systems using real-time retail prices," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 2400–2407, IEEE, 2016.

Addendum

Derivation of Equation (4)

We take of the logarithm of Eq. (3) to obtain the entropy

$$\sigma = \ln N! - \ln k! - \ln(N-k)! + k\ln\rho + (N-k)\ln(1-\rho).$$

We use the Stirling approximation to expand the factorials and obtain the entropy

$$\sigma = \sigma_0 + (k + \tfrac{1}{2})\ln k - (N - k + \tfrac{1}{2})\ln(N-k) - k(N-k)\rho(1-\rho)$$

with

$$\sigma_0 = N\ln\tfrac{1}{2}N - \tfrac{1}{2}\ln 2\pi.$$

for large N. Using $\rho = \bar{\rho} = \frac{1}{2}$ and the mean device load q we define the load deviation $\Delta Q = (k - \frac{1}{2}N)q$ and the total thermostatic load $Q_R = Nq$. Then we obtain

$$\sigma = \sigma_0 - 2\Delta Q^2/Q_R$$

with

$$\sigma_0 = \frac{Q_R}{q}\ln\frac{Q_R}{2q} - \tfrac{1}{2}\ln 2\pi. \qquad \square$$

Derivation of Equation (5)

We compute the probability of observing the state k in terms of entropy, such that

$$\text{Prob}[k] = 2^{-N} g(k) = 2^{-N} e^{\sigma(Q)}.$$

We expand the total number of states k as a function of the count of devices *on* plus the number of devices *off*, i.e.,

$$2^N = e^{\sigma(Q)} + e^{Q_R - Q}$$

such that

$$\text{Prob}[k] = \frac{e^{\sigma(Q)}}{e^{\sigma(Q)} + e^{\sigma(Q_R - Q)}}$$

$$= \frac{1}{1 + e^{-\sigma(Q)\sigma(Q_R - Q)}}$$

which for $\bar{\rho} = 0.5$ we can simplify as

$$\text{Prob}[k] = \frac{1}{1 + e^{-2\sigma(Q)}}. \qquad \square$$