Interruptible Electricity Contracts from an Electricity Retailer's Point of View: Valuation and Optimal Interruption
Author(s): Ross Baldick, Sergey Kolos and Stathis Tompaidis
Source: *Operations Research*, Vol. 54, No. 4 (Jul. - Aug., 2006), pp. 627-642
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/25147000
Accessed: 24/09/2014 10:54

OPERATIONS RESEARCH
Vol. 54, No. 4, July–August 2006, pp. 627–642
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5404 | 0627

DOI 10.1287/opre.1060.0303

Interruptible Electricity Contracts from an Electricity Retailer's Point of View: Valuation and Optimal Interruption

Ross Baldick
Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas 78712,
ross.baldick@engr.utexas.edu

Sergey Kolos
Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712,
skolos@mail.utexas.edu

Stathis Tompaidis
Department of Information, Risk and Operations Management, McCombs School of Business, University of Texas at Austin,
Austin, Texas 78712, stathis.tompaidis@mccombs.utexas.edu

We consider interruptible electricity contracts issued by an electricity retailer that allow for interruptions to electric service in exchange for either an overall reduction in the price of electricity delivered or for financial compensation at the time of interruption. We provide a structural model to determine electricity prices based on stochastic models of supply and demand. We use stochastic dynamic programming to value interruptible contracts from the point of view of an electricity retailer, and describe the optimal interruption strategy. We also demonstrate that structural models can be used to value contracts in competitive markets. Our numerical results indicate that, in a deregulated market, interruptible contracts can help alleviate supply problems associated with spikes of price and demand and that competition between retailers results in lower value and less frequent interruption.

Subject classifications: natural resources: energy, interruptible electricity; dynamic programming/optimal control.
Area of review: Environment, Energy, and Natural Resources.
History: Received December 2003; revision received June 2004; accepted June 2005.

1. Introduction

The market for one of the most important commodities in today's economic environment, electricity, has recently undergone significant changes. For most of its North American history, electricity in each geographical region was generated, transmitted, and distributed by one heavily regulated, vertically integrated company. The electricity industry is currently in transition towards a restructured market with many more market players in each region, most of which will provide only a part of the services provided by the original participants.[1]

In the regulated environment, risks to the market participants were mitigated by the mechanism of regulated cost recovery. However, under restructuring, and facing competition, such cost recovery is unlikely or limited, creating the need for the use of financial risk management tools and techniques. To mitigate financial risk, new financial products have been developed, and existing tools for management of supply and demand are being used. Among the latter is the interruptible contract, which allows one party to renege on its obligation to provide electricity to the other party a certain number of times over a certain period of time.[2] In this paper we provide a valuation framework for interruptible contracts from the point of view of an electricity retailer and study how these contracts may help such a retailer reduce its exposure to fluctuations in the demand and supply of electricity. In our setting, an electricity retailer has agreed to provide electricity to satisfy the demand of its customers. To serve this load, the electricity retailer either owns generating assets or has access to generating assets, for example, through bilateral agreements. The excess load has to be served through purchases in the spot electricity market. Examples of such retailers include Pacific Gas and Electric and Southern California Edison in California, and TXU and Reliant in the Electric Reliability Council of Texas.

A recent paper by Kamat and Oren (2002) presents a simple form of an interruptible contract in which one party can interrupt the other once over two possible interruption opportunities, and where it is assumed that interruption does not influence the spot electricity price. In our work, we extend and generalize the paper of Kamat and Oren (2002) in several directions. First, we allow for the possibility of multiple interruptions over many possible interruption dates, possibly with daily frequency, when there is a limit on the total number of interruptions.[3] Second, we allow for different types of interruptible contracts. Different types of contracts may generate differences in the optimal interruption policy because in some cases the cost of interruption may be sunk. Finally, the most important difference between our work and that of Kamat and Oren (2002), as well as other papers in the literature, involves the impact of the interruption on the spot price of electricity. While Kamat and Oren (2002) consider reduced-form models for electricity prices (either geometric Brownian motion, or a mean-reverting process with jumps), we construct a structural model in which the spot price of electricity is determined by supply and demand.

In our setting it is crucial to use a structural model that incorporates demand in determining electricity prices because much of the benefit to an electricity retailer from interrupting a load comes not from avoiding servicing the interrupted load, but instead from reducing the total load to the system, leading to systemwide lower prices. This feature is very valuable to an electricity retailer that needs to resort to the spot market to cover some of its demand because spot prices can spike to high levels when supply is tight.

Another contribution of this paper is the calibration of a structural model, based on the equilibrium between supply and demand of electricity. We present data that indicate that fluctuations in demand are mainly driven by temperature fluctuations, and proceed to model temperature using an autoregressive process, which is statistically estimated using over 50 years of temperature data. Supply, on the other hand, is modeled through the "supply curve," which orders electricity-generating plants based on their marginal generation cost. Due to differences in the generating technologies that are marginal at different levels of production, we model the supply curve using a two-regime model. In addition, supply is allowed to fluctuate due to outages and transmission constraints. The combination of the demand and supply models generates many of the observed characteristics of electricity prices, such as both mean reversion and short-lived spikes in electricity prices due to mean reversion in temperature and the two regimes of supply, respectively. Using this structural model, we are able to numerically value interruptible contracts and determine the optimal interruption policy from the point of view of the electricity retailer.

Our model allows us to study the impact of retailer competition on interruptible contracts, both in terms of value and in terms of changes in the optimal interruption policy. Specifically, because interruption is costly, when multiple electricity retailers serve the same area, competing retailers try to free-ride, resulting in less efficient use of interruptible contracts. In the case of identical electricity retailers, we find that, on the one hand, interruptions occur at higher systemwide loads, while, on the other hand, the value of interruptible contracts drops as the number of retailers increases. While competition lessens the incentive of any single retailer to introduce interruptible contracts, we find that the value of interruptible contracts remains high in situations where electricity retailers have limited generation available.

The rest of this paper is organized as follows: §2 describes the market setting as well as the different forms of interruptible contracts we consider. Section 3 describes the structural model for electricity prices that links electricity demand and generation supply. The model is calibrated with data from the Electric Reliability Council of Texas (ERCOT) System. In §4, we formulate a stochastic control problem for the valuation of interruptible contracts from the point of view of a risk-neutral electricity retailer, and describe the optimal interruption strategy as well as the value for the different forms of interruptible contracts. In §5, we discuss the case of multiple electricity retailers with interruptible contracts serving the same geographical area. Section 6 concludes. In the online appendix at http://or.pubs.informs.org/Pages/collect.html, we derive the technical results necessary for the solution of the problems formulated in §§4 and 5 and discuss the implementation and performance of the numerical algorithm.

2. Model

2.1. Market Description

We consider the case of a large retailer of electricity that contracts with retail customers in a specified geographic area to provide electricity to satisfy all of their electricity demand. The retailer charges a fixed retail price per unit of electricity, p_{retail}, to each of its customers (prices are typically differentiated by customer class, but we will ignore this issue here). The retailer has available a certain generation capacity, $L_{\text{generation}}$, either through the ownership of generators, through forward purchase agreements, or longer-term bilateral contracts. We assume that the cost of this electricity available to the retailer is fixed in advance at $p_{\text{generation}}$ and does not depend on the spot price of electricity. To the extent that it is hedged by long-term contracts, the retailer's exposure to the spot prices would be reduced. A typical retailer may not be completely hedged for all of its peak demand, however, and so would be exposed at the margin to the spot prices. When demand is higher than the generation capacity available to the electricity retailer, the retailer is forced to serve the demand through purchases in the spot electricity market. We assume that the retailer utilizes all power available from its own generators first, and then turns to the energy market. To be consistent, in the event that the generation available is greater than the load, the retailer can sell the surplus in the spot market. In the examples we consider, we focus on situations where the electricity retailer almost never has enough generation capacity available to serve the entire demand without resorting to the spot market. In practice, capacity may be purchased in advance and be truly sunk, while the generation of energy incurs additional costs that may be avoidable. With appropriate redefinition of prices, this case can be treated with the model we develop.[4]

Regarding the customers of the electricity retailer, we assume that they can only purchase electricity from the retailer and can only use electricity for consumption; i.e., they cannot resell it. The customers belong to one of two categories: They are either "residential," with fluctuating demand $L_{\text{residential},\,t}$; or, they are "industrial," with constant demand $L_{\text{industrial}}$. Under this specification, "industrial" customers may include both industrial and commercial users of electricity. In fact, industrial demand may also vary with time, complicating the design of the interruptible contracts because of the difficulty in setting a "baseline" for interruption of demand—see Borenstein (2005).

Total demand for day t is equal to $L_{\text{residential},\,t} + L_{\text{industrial}}$. We abstract from the intraday variation in demand by assuming that $L_{\text{residential},\,t} + L_{\text{industrial}}$ represents the average demand during on-peak hours in day t and that the average demand during on-peak hours is the main determinant of market price.

2.2. Interruptible Contracts

There are several variants of interruptible contracts offered by retailers of electricity. In its most general form, an interruptible contract between a retailer and a customer allows the retailer to interrupt part or all of the supply of electricity to the customer over some period of time in exchange for some form of pecuniary compensation. In most cases, the retailer does not physically interrupt the customer, but rather gives the customer an advance notice, typically between 30 minutes and 24 hours, requesting curtailment of the customer's load. Failure of the customer to curtail the load to the specified level can lead to severe penalties, effectively resulting in the interruption of the customer's load. We will assume, for the rest of the paper, that all loads are either served or interrupted. Interrupted loads are compensated for the interruption according to the provisions of the interruptible contract.

Although they possibly existed earlier, the earliest mention of interruptible contracts in the literature that we are familiar with refers to interruptible contracts for natural gas to industrial clients in the 1930s and 1940s (see Troxel 1949, p. 14 and Smith 1946, p. 421). Interruptible contracts in electricity are mentioned in Raver (1951, p. 293) and Lee (1953, p. 184) for industrial clients in the 1940s and 1950s in the Columbia River basin in the Pacific Northwest of the United States. To our knowledge, the first paper in the literature that provides a theoretical framework for studying interruptible contracts is the paper by Brown and Johnson (1969), who recognize that interruption of service is a natural consequence of an economic environment where resources are priced prior to the realization of uncertain demand. The paper by Tschirhart and Jen (1979) discusses the problem of segmenting the customers of a monopolistic retailer into service priority classes, with the objective of maximizing the monopolist's profit in a two-period setting. Chao and Wilson (1987) prove that introducing a few service priority classes together with an appropriate price menu results in overall efficiency gains and dominates random rationing. Chao et al. (1988) refine the implementation discussed in Chao and Wilson (1987) and describe the effect of interruptible contracts in monopolistic and oligopolistic market structures. Oren and Smith (1992) use interruptible contracts to design and implement a model to reduce annual peaks in electricity demand. Caves and Herriges (1992) use stochastic dynamic programming to maximize expected benefits from an interruptible program. In this paper, we use a similar formulation and extend the work of Caves and Herriges (1992) by quantifying the benefit of interruption based on a model of supply and demand of electricity. In addition, we allow for two types of interruptible contracts, flexibility in the amount of daily interrupted load, and interaction between the amount of interrupted load and the benefit to the electricity retailer.

While interruptible contracts existed in the regulated electricity industry as a way to prioritize interruption schedules in an emergency, they have become more prominent as a risk management tool after the two California electricity crises, in the summer of 1998 and the winter of 2001. During the 1990s and prior to 1998, while interruptible contracts provided the right to interruption by the utility, these rights were rarely exercised, leading to a skewed perception of their risk among customers. Because signing up for an interruptible contract provided a discount on the retail price of electricity, many customers that never intended to be interrupted, such as hospitals, schools, and nursing homes signed their electric load on interruptible contracts. Unsurprisingly, when called to interrupt, these customers refused to do so. We assume that interruptible contracts are between the retailer and "industrial" customers only, and that upon request, the customer always curtails the requested load. We focus on two particular types of interruptible contracts that appear to be among the most common. A detailed description of these contracts, as well as additional background, is available from the report of the Energy Division of the California Public Utilities Commission (2001).

The first form of an interruptible contract, which we call a *pay-in-advance* contract, allows the electricity retailer to interrupt a given percentage of an "industrial" customer's load a fixed number of times over the life of the contract. In exchange, the customer receives a discount on the retail price of electricity for the customer's entire load, $L_{\text{under\_contract}}$, and pays p_{reduced} per unit of electricity, rather than p_{retail}. Typical values for the parameters of this contract are a 15% discount on the retail price in exchange for 10 daily interruptions of 20% of the customer's load over the period of one year.

The second form of an interruptible contract, which we call a *pay-as-you-go* contract, allows the electricity retailer to interrupt part of a customer's load a fixed number of times in exchange for compensation, p_{fine}, per unit of load interrupted. This compensation is typically chosen to be considerably higher than the retail price p_{retail}. Typical values for the parameters of this contract allow for 10 interruptions with compensation, p_{fine}, ranging from \$150 per

MWh to $600 per MWh of interrupted electricity, depending in part on whether notice of interruption is given the day before interruption, or with shorter notice such as one hour in advance of interruption. In this paper, we will focus on interruptible contracts where notice of interruption is given the day before interruption.[5]

Besides constraints on the total number of interruptions, other constraints may also exist for both pay-in-advance and pay-as-you-go contracts. For example, the number of consecutive days of interruption may be limited, or no more than a certain number of interruptions may occur over a short period of time.

Assuming that the number of interruptible contracts signed between an electricity retailer and "industrial" customers is large, and that the load interrupted under each contract is comparatively small, the individual constraints are not binding on the electricity retailer's actions because the retailer can pool all the contracts together. For example, the number of times a particular customer may be interrupted is not relevant for the retailer, as long as the retailer is careful to rotate interruptions between all of its customers. From the retailer's point of view, pooling simplifies the management of the portfolio of interruptible contracts. For each type of interruptible contract, the retailer need only keep track of the maximum amount available for daily interruption and of the total remaining amount of interruption until the end of the year. We assume that the pooling approximation is valid, and that all interruptible contracts are effective over the same period (contractual limits on the exercise pattern can be used as a way to discriminate among customers with different cost profiles; in the context of a larger model that incorporates customer information, a retailer could minimize cost by designing interruptible contracts with different exercise patterns).

3. A Structural Model for Electricity Prices

While much of the literature on the stochastic process followed by electricity prices has focused on reduced-form models that mimic the observed price behavior (see Pilipovic 1997; Deng 1999, 2000; Kholodnyi 2004; and Geman and Roncoroni 2006), such models are of limited value for the problem we consider. Implicitly, in a reduced-form model one assumes that the price process is not influenced by the actions of market participants. However, in the case of a large retailer of electricity with interruptible contracts, the interruption has the effect of lowering demand as well as lowering the expected spot price. To account for this interaction between interruption and electricity price, we develop a structural model of the electricity market, where prices are determined by matching supply and demand, where we model supply and demand separately. In the literature, similar structural models were proposed by Skantze et al. (2000) (for the Pennsylvania-New Jersey-Maryland area), Barlow (2002), and Eydeland and Wolyniec (2003) (see also the references within). We calibrate the model with data from the ERCOT area during weekdays in the summer months because, in the case of ERCOT, summer weekdays is the period when electric loads are very high and when interruption is most likely to occur. (ERCOT covers almost all of Texas—for an introduction to the ERCOT electricity market, see Baldick and Niu 2005 and www.ercot.com). We note that the ERCOT market has several electricity retailers. We explicitly consider the case of competing electricity retailers in §5.

In our structural model, we try to reflect some of the characteristics of electricity markets. In particular, due to the fact that almost all consumers of electricity have fixed-price retail contracts, we assume that demand is inelastic; i.e., it does not depend on the spot electricity price. Given inelastic demand it is important that electricity generators do not collude nor exercise significant unilateral market power. To avoid consideration of market power, we assume a competitive market for the generation of electricity, but recognize that this is not always a reasonable assumption. (In principle, generation market power could be modeled as a shift in the supply curve—note that the electricity retailer is explicitly assumed to possess market power because it can influence price by adjusting its demand through interruptible contracts.)

Consistent with the assumption of a competitive generation market, we assume that generators are dispatched in order of marginal cost from lowest to highest. The total demand determines which of the generators are dispatched. We assume that the spot price of energy is equal to the marginal operating cost of the marginal dispatched generator. Sometimes there may be a violation of the strict merit order due to congestion of the transmission system. Moreover, start-up and minimum-load costs can affect the order of dispatching generation. We abstract from these issues by introducing random fluctuations to the supply curve.

In the rest of this section, we consider demand and supply in detail.

3.1. Demand

Stylistic facts concerning demand of electricity are that it is strongly seasonal (with daily, weekly, and annual patterns), strongly mean reverting, and highly predictable. Demand is influenced by environmental factors such as temperature and humidity, as well as population size and industrial activity. In this paper, we assume that demand has two components: one that is relatively stable due to "industrial" customers, and one that varies with time due to "residential" customers. We model demand fluctuations of the residential customers in terms of temperature fluctuations, which is the most important driving factor of demand in ERCOT during the summer, and limit our analysis to a single summer so that changes in the population size and industrial activity are negligible.

3.1.1. Temperature Model. We use a model for forecasting temperature similar to the one introduced by Cao and Wei (2000, 2004) (see also Campbell and Diebold 2005). In our model, the deviation of the actual temperature from the historical average of temperature over the next day, $t+1$, is a function of the deviation of the actual from the historical average of the temperature today, t, and the deviation of the actual from the historical average over the previous day, $t-1$. Cao and Wei (2000) used a model in which future temperature deviations, at time $t+1$, depend on temperature deviations over three previous dates, t, $t-1$, and $t-2$. We have found that for Texas the dependence on the temperature deviations for day $t-2$ is statistically insignificant, and we have not included this term in the model. The model allows for stochastic fluctuations around the historical average, with magnitudes that depend on the time of the year, and is described by the following equations:

$$\Delta_{t+1}^T = \rho_1^T \Delta_t^T + \rho_2^T \Delta_{t-1}^T + \sigma_{t+1}^T \varepsilon_{t+1}^T,$$
$$\sigma_t^T = \sigma_{(0)}^T - \sigma_{(1)}^T \left| \sin\left(\pi \frac{t+\phi}{365}\right) \right|, \quad (1)$$
$$\varepsilon_t^T \sim \text{i.i.d.}(N(0,1)),$$

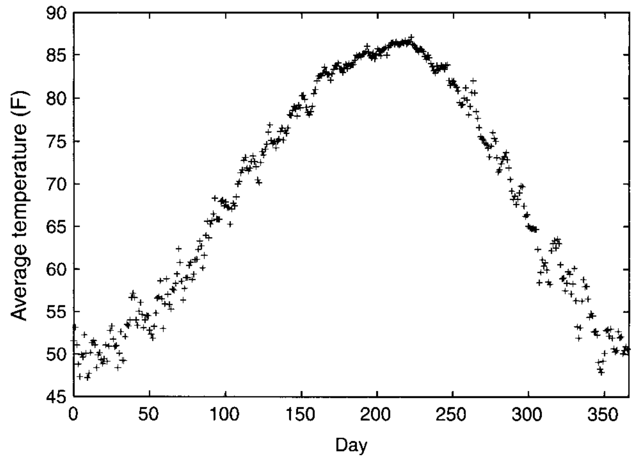
where $\Delta_t^T = T_t - \overline{T}_t$, T_t is the actual temperature for day t, $\overline{T}_t$ is the average temperature for day t, and ρ_1^T, ρ_2^T are the partial autocorrelation coefficients for deviations from average temperature. By substituting temperature forecasts rather than historical averages, the model can also incorporate information from short- and long-term meteorological forecasts. The magnitude of the random fluctuations is seasonal, with a fixed term $\sigma_{(0)}^T$ and a seasonal term of magnitude $\sigma_{(1)}^T$. The parameter ϕ corresponds to the date during the year when the fluctuations are the largest.

To calibrate the model for ERCOT, we use data available at the National Climatic Data Center website (www.ncdc.noaa.gov). We use daily data on average temperatures in central Texas from January 1948 through December 1999. Figure 1 presents the average daily temperatures. The variables $\overline{T}_t$ in Equation (1) are set to these averages.

After obtaining the values for the average temperatures, we calibrate the temperature model in two steps: First, we construct the variable $\Delta_t^T = T_t - \overline{T}_t$ for each day in the data set. Because the model is heteroskedastic, we use an iterative procedure, in which we start with a guess for $\sigma_{(0)}^T$, $\sigma_{(1)}^T$, ϕ. Using this guess for the heteroskedastic errors, we regress Δ_{t+1}^T on Δ_t^T and Δ_{t-1}^T to estimate the partial autocorrelation coefficients ρ_1^T, ρ_2^T. We then construct the deviations between the expected temperature deviations and the actual temperature deviations for each day, and use them to compute the deviations σ_t^T, from which we fit, using nonlinear regression (see Ratkowsky 1983), the parameters $\sigma_{(0)}^T$, $\sigma_{(1)}^T$, ϕ. We repeat the procedure until the values of the parameters $\sigma_{(0)}^T$, $\sigma_{(1)}^T$, ϕ converge. The estimated parameter values and their standard errors are reported in Table 1.

3.1.2. Demand vs. Temperature. To estimate the relationship between demand for electricity and temperature,

Figure 1. Average daily temperatures for central Texas, averaged over 1948–1999.



we use a data set of power loads for the summer 1999 period for ERCOT available at the ERCOT website (www.ercot.com). The data provide the average daily on-peak and off-peak load by region within ERCOT. We use average on-peak load, which includes load between 6 a.m. and 10 p.m. Monday through Friday. The reason for this choice is that night and weekend load is low enough that interruptions are not necessary. Figure 2 presents the relationship between average temperatures and on-peak load during weekdays for the period June 1 to August 31, 1999 in ERCOT. The lines in the graph represent the 10th percentile, median, and 90th percentile based on the estimated load-average temperature model. From Figure 2, it is clear that for the range of temperatures encountered during the summer months, there is a close-to-linear relationship between average on-peak load and average temperature. (Most variability of demand in Texas during the summer is driven by air-conditioning load, which is dependent on temperature—in a colder climate one may need to include additional terms in the load-temperature relationship.)

Based on Figure 2, we model the relationship between average temperature and average load by a linear function with additional random fluctuations:

$$L_t = \alpha_L + \beta_L T_t + \sigma_L \varepsilon_{L_t}, \quad \varepsilon_{L_t} \sim N(0,1), \quad (2)$$

where L_t is the load at time t, T_t the temperature, α_L the load intercept, β_L the expected marginal increase in load for

Table 1. Temperature model.

	Estimate	Standard error
Intercept	−0.0002	0.010
ρ_1^T	0.837	0.010
ρ_2^T	−0.188	0.010
$\sigma_{(0)}^T$ (Fahrenheit)	8.316	0.131
$\sigma_{(1)}^T$ (Fahrenheit)	5.747	0.185
ϕ (days)	−14.5	1.6

Figure 2. Average on-peak load vs. average temperature.

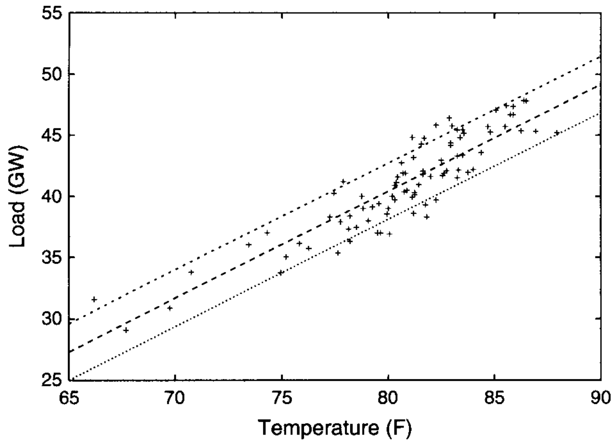
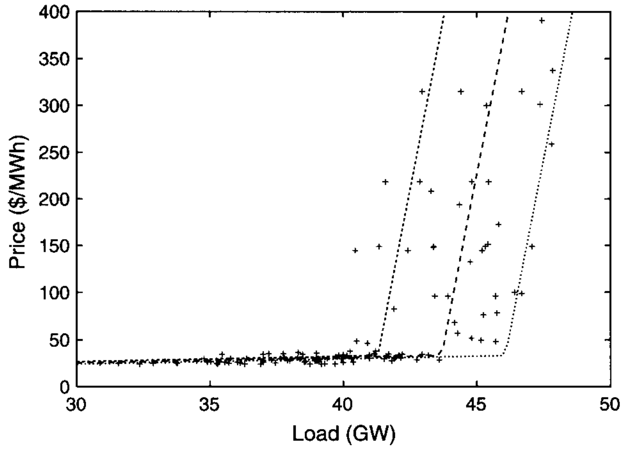


Figure 3. On-peak electricity price vs. average daily load.



a unit increase in temperature, and σ_L the magnitude of the random fluctuations around the linear relationship between load and temperature. Table 2 presents the ordinary least squares (OLS) regression estimates for the values of the parameters. The R^2 of the regression is 86%.

3.2. The Supply Curve

Most of the supply available in ERCOT is generated within ERCOT, due to limited transmission between ERCOT and surrounding areas. The generators that service the base load are coal-based or nuclear facilities, while intermediate and peaking plants include plants based on natural gas, oil, or hydroelectric power. Because we do not have access to the marginal costs of the available generators, we calibrate our model of the supply curve through the observed relationship between spot electricity price and electric load. To justify this approach, we note that because all ERCOT customers paid an essentially fixed retail price during the study period, we assume their demand to be inelastic with respect to the wholesale spot price. We note that market participants have additional, proprietary, information sources that can be used to improve the accuracy of the calibration. However, an advantage of our calibration procedure is that model prices incorporate past strategic decisions by market participants—see Eydeland and Wolyniec (2003) for a discussion of the difficulties of matching observed electricity prices using a structural model calibrated from marginal cost estimates.

In Figure 3, we present the relationship between the on-peak price per MWh of electricity and the average daily load, during weekdays for the period June 1 to August 31, 1999 in ERCOT. The lines in the graph represent the 10th percentile, median, and 90th percentile based on the estimated price-load model. From Figure 3, we notice that there appears to be two regimes for the supply curve: the low-demand regime, where load and prices are relatively low and price fluctuations are minor; and the high-demand regime, where load is high and price fluctuations are large. Based on these observations, we propose a two-regime model for the price/load relationship. We allow for random fluctuations in price to account for fluctuations in supply due to, for example, generator outages, transmission outages, transmission congestion, and possibly strategic behavior by market participants. For simplicity, we use a single random variable to represent the fluctuations for both regimes. This assumption is not critical for the valuation of interruptible contracts as long as the magnitude of the fluctuations is calibrated from the high-demand regime. The reason is that small errors in the calibration of the model parameters for the low-demand regime have only a minor impact on the value of interruptible contracts.

Table 2. Load model.

	Estimate	Standard error
Intercept α_L (GW)	−29.5	3.5
Slope β_L (GW/Fahrenheit)	0.874	0.044
σ_L (GW)	1.80	0.16

The model of the relationship between load and price is given by

$$P_t = \begin{cases} \beta_{S,l}(L_t + \sigma_S \varepsilon_{S_t}) + \alpha_{S,l} & \text{if } L_t + \sigma_S \varepsilon_{S_t} \leqslant S_b, \\ \beta_{S,h}(L_t + \sigma_S \varepsilon_{S_t}) + \alpha_{S,h} & \text{if } L_t + \sigma_S \varepsilon_{S_t} > S_b, \end{cases} \quad (3)$$

where P_t is the wholesale price at time t, L_t the demand at time t, ε_{S_t} is a standard, normally distributed random variable, and S_b the supply level that determines the break between the high-demand and low-demand regimes.

To calibrate the supply curve model, we use the days from the data in Figure 3 with prices above $60/MWh, assuming that they correspond to the high-demand regime. From these days, we estimate the parameters for the high-demand regime, as well as the magnitude of supply fluctuations σ_S. We estimate the parameters for the low-demand

Table 3. Supply curve model.

	Estimate	Standard error
$\beta_{S,l}$ (\$/GW)	0.554	0.281
$\beta_{S,h}$ (\$/GW)	146.0	78.6
$\alpha_{S,l}$ (\$)	8.86	10.41
$\alpha_{S,h}$ (\$)	−6,344.5	3,418.9
σ_S (GW)	1.863	0.16
S_b (GW)	43.68	

regime using days in which ERCOT load was below 39 GW. We use 39 GW to ensure that we have not crossed over to the high-demand regime. Alternatively, one could use a recursive procedure, where S_b is estimated and used as the cutoff for the estimation of the parameters for the low-demand regime. The break point S_b is calculated by requiring the expected price to be a continuous function of load; i.e.,

$$\beta_{S,l}S_b + \alpha_{S,l} = \beta_{S,h}S_b + \alpha_{S,h}.$$

The OLS estimates for the parameter values are presented in Table 3. The R^2 for the OLS regression for the low-demand regime is 23%, while the R^2 for the high-demand regime is 21%.

4. Valuation and Optimal Interruption Policy for Interruptible Contracts

In this section, we discuss the formulation of the stochastic optimal control problem that maximizes the expected value of the interruptible contracts from the point of view of the electricity retailer. We first solve the problem for two special cases: when there is no limit on yearly interruption and when there is no limit on daily interruption, respectively. We then present two particular base-case contracts with limits on both yearly and daily interruption and then solve for the optimal interruption policies for each base case. Finally, we discuss the value of the base-case interruptible contracts.

4.1. Stochastic Optimal Control Problem

The problem of determining the optimal interruption policy, as well as the value of interruptible contracts, can be formulated as a problem of optimal stochastic control, with the objective of maximizing the utility of the electricity retailer. We assume that the retailer is risk neutral with respect to gains and losses and has intertemporal preferences that can be quantified through a constant discount factor. Other choices for the risk aversion of the retailer are possible. However, choosing a risk-neutral retailer is sufficient to capture the factors that are important in determining the optimal interruption policy, as well as the value of an interruptible contract.

As we have already discussed, we assume that the electricity retailer can pool all the interruptible contracts, and therefore need only consider the load available for interruption the following day and the total load available for interruption during the remaining period. In addition, the retailer may think of all its customers in terms of three representative customers: the first customer has not signed an interruptible contract and pays p_{retail} on its load; the second customer has signed a pay-in-advance interruptible contract and pays a reduced price on its load, p_{reduced}, but does not receive any additional compensation upon interruption so that $p_{\text{fine}} = 0$; and the third customer has signed a pay-as-you-go contract, pays p_{retail} on its load, and receives compensation p_{fine} per unit of interruption, upon interruption.

The net profit, $\Delta\pi$, to the retailer during a day with 16 on-peak hours with:

- load of L prior to interruption,
- load signed under pay-in-advance-contracts of $L_{\text{under\_contract}}$,
- load interrupted from the pay-in-advance contracts of l_{advance},
- load interrupted from the pay-as-you-go contracts of l_{pago}, and
- spot price p_{spot}, which is a function of the expected load after interruption $L - l_{\text{advance}} - l_{\text{pago}}$, and of price fluctuations ε_S,

is given by

$$\begin{aligned}\frac{\Delta\pi(L, p_{\text{spot}}, l_{\text{advance}}, l_{\text{pago}})}{16} &= (L - L_{\text{under\_contract}} - l_{\text{pago}})p_{\text{retail}} \\ &\quad + (L_{\text{under\_contract}} - l_{\text{advance}})p_{\text{reduced}} - L_{\text{generation}}p_{\text{generation}} \\ &\quad - l_{\text{pago}}p_{\text{fine}} - (L - l_{\text{advance}} - l_{\text{pago}} - L_{\text{generation}})p_{\text{spot}}. \qquad (4)\end{aligned}$$

The net profit is made up of five different terms:

- $(L - L_{\text{under\_contract}} - l_{\text{pago}})p_{\text{retail}}$ corresponds to the revenue to the retailer from the customers that have not signed a pay-in-advance interruptible contract and were not interrupted under a pay-as-you-go contract,
- $(L_{\text{under\_contract}} - l_{\text{advance}})p_{\text{reduced}}$ corresponds to the revenue from customers that have signed a pay-in-advance interruptible contract but were not interrupted,
- $L_{\text{generation}}p_{\text{generation}}$ corresponds to the cost of procuring the generation available to the retailer at a fixed price,
- $l_{\text{pago}}p_{\text{fine}}$ corresponds to the cost to the retailer for interrupting customers under a pay-as-you-go contract, and
- $(L - l_{\text{advance}} - l_{\text{pago}} - L_{\text{generation}})p_{\text{spot}}$ corresponds to the cost of servicing the excess demand by buying electricity in the spot market.

Given our formulation of a structural model for electricity prices in §3, the load and the spot price of electricity at time t depend on the temperature deviations from the temperature historical averages, or forecasted values, at time t and $t-1$. Given the values of the state variables Δ_t^T,

Δ^T_{t-1} and the remaining interruptible loads, $L_{\text{advance, remaining}}$, $L_{\text{pago, remaining}}$, the value function for the retailer is given by

$$\begin{aligned}\pi_t(\Delta^T_t, \Delta^T_{t-1}, L_{\text{advance, remaining}}, L_{\text{pago, remaining}}) \\ = \max_{l_{\text{advance}}, l_{\text{pago}}} \beta\{\mathbb{E}[\Delta\pi(L_{t+1}, p_{\text{spot}, t+1}, l_{\text{advance}}, l_{\text{pago}}) \\ + \pi_{t+1}(\Delta^T_{t+1}, \Delta^T_t, L_{\text{advance, remaining}} - l_{\text{advance}}, \\ L_{\text{pago, remaining}} - l_{\text{pago}}) \mid \mathcal{F}_t]\}, \qquad (5)\end{aligned}$$

where the maximization is over

$$\begin{aligned}&0 \leqslant l_{\text{advance}} \leqslant \min(L_{\text{advance, daily}}, L_{\text{advance, remaining}}),\\ &0 \leqslant l_{\text{pago}} \leqslant \min(L_{\text{pago, daily}}, L_{\text{pago, remaining}}).\end{aligned} \qquad (6)$$

In Equation (5), β is the discount factor and $\mathcal{F}_t$ denotes the information available at time t. Note that the interruption amounts for the pay-in-advance and pay-as-you-go contracts, l_{advance} and l_{pago}, respectively, are chosen at time t, but interruption occurs over the next day, at time $t+1$. The expectation in Equation (5) is taken over the random variables $\varepsilon_{L_{t+1}}$, $\varepsilon_{S_{t+1}}$, ε^T_{t+1}.

Assuming a terminal date t_f for the interruptible contracts, we set

$$\pi_{t_f} = 0.$$

The maximization problem can be solved using stochastic dynamic programming with state variables Δ^T_t, Δ^T_{t-1}, $L_{\text{advance, remaining}}$, $L_{\text{pago, remaining}}$, and choice variables l_{advance}, l_{pago}. The stochastic dynamic programming algorithm is described in detail in the online appendix, and involves discretizing both Δ^T_t and Δ^T_{t-1} into N_T steps between $-D^T$ and D^T for D^T a suitable bound on temperature deviations. The state variables $L_{\text{advance, remaining}}$, $L_{\text{pago, remaining}}$ are discretized into N_L steps between 0 and the yearly amount available for interruption.

In our numerical experiments, we took $N_T = 21$, $N_L = 20$, and $D^T = 10$ (which corresponds to temperature steps of 1 Fahrenheit degree). The algorithm was programmed in C using the GNU Scientific Library for interpolations, integrations, and maximizations (see Galassi et al. 2003). Running on a 1.7 GHz Pentium 4 processor, the program computes the value of a 90-day contract in 90 seconds and performs 10,000 Monte Carlo simulations in three seconds.

4.2. Optimal Interruption Policy in Cases of No Limits

In this section, we consider the optimal interruption policy, first in the special case where there are no limits on yearly interruption, and second in the special case where there are no limits on daily interruption. In each case, the optimal interruption policy is determined by the first-order condition that, at the optimal policy, the marginal benefit to the retailer from additional interruption equals the marginal cost to the retailer.

4.2.1. No Limit on Yearly Interruption. We first consider the case when there is no limit in the total yearly amount available for interruption. Then, the value function in Equation (5) does not depend on $L_{\text{advance,remaining}}$, $L_{\text{pago, remaining}}$, and the maximization is myopic; i.e., on each day t, the optimal interruption policy maximizes expected net profit on day $t+1$ only. In this case, we can easily calculate the marginal cost and marginal benefit of interruption to the retailer. For the case of pay-in-advance contracts, the marginal cost is p_{reduced}, which corresponds to foregone revenue, while for pay-as-you-go contracts the marginal cost is $p_{\text{retail}} + p_{\text{fine}}$, which corresponds to foregone revenue and the fine paid per unit of interruption.

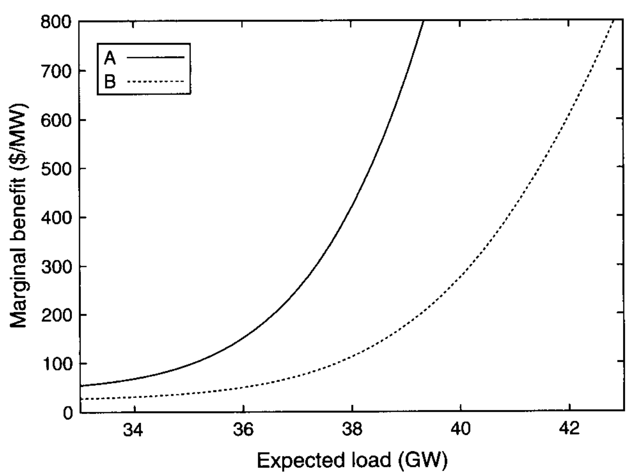
The marginal benefit is the same for both contract types, and is a function of the expected load. The marginal benefit has two components: One component corresponds to not servicing the interrupted load at high expected prices; the second component corresponds to lowering the overall demand, and therefore paying a smaller price to procure electricity for the entire serviced load. The value of the second component is measured in terms of the savings to the retailer and depends on the retail price p_{retail}.

The previous discussion leads to the following proposition.

Proposition 1. *In the case of no yearly limit for pay-in-advance and pay-as-you-go interruptible contracts, if $p_{reduced} \leqslant p_{retail} + p_{fine}$, then the optimal policy involves interrupting the pay-in-advance contracts up to their daily limit before interrupting any of the pay-as-you-go contracts.*

For the base-case contracts, the marginal benefit for different parameter values for the interruptible contracts is calculated in the online appendix. In Figure 4, we present the marginal benefit from interrupting a MW of electricity on August 31st in ERCOT, when there is an unlimited amount of yearly interruption available. Curve A corresponds to an

Figure 4. Marginal benefit from interrupting an MW of electricity.



electricity retailer that has zero generation available at a fixed cost; curve B corresponds to a retailer that has 35 GW of generation available at a fixed price.

The optimal policy can be determined from the figure in the following way: If the expected load is such that the marginal benefit is greater than the marginal cost, the electricity retailer interrupts an amount that is the lesser of:

- the maximum daily interruptible limit, and
- the amount for which the expected load is reduced to the point where the marginal benefit equals the marginal cost.

For example, from Figure 4, if the retailer has 35 GW of generation available, the retail price is \$60/MWh, and the fine per MWh of interruption is \$150, then the marginal cost is \$210/MWh, which matches marginal benefit at an expected load of 39.4 GW. If the daily interruptible limit is 2 GW and the expected load is 41 GW, the retailer will interrupt 1.6 GW; while if the expected load is 43 GW, the retailer will interrupt the entire 2 GW daily limit. In the case of a retailer without any generation available, with a retail price of \$60/MWh, and fine per MWh of interruption of \$600, the marginal cost is \$660/MWh, and interruption first occurs at an expected load of 38.8 GW. If the expected load is 40 GW, the retailer would interrupt 1.2 GW; while if the expected load is at or above 40.8 GW, the retailer would interrupt the entire 2 GW daily limit.

From Figure 4, we notice that the optimal interruption policy for interruptible contracts without yearly limits depends on several factors. In particular, the expected load at which interruption begins increases with the amount of generation available to the electricity retailer at a fixed price. The intuition for this result is that the marginal benefit of interruption for a given expected load decreases with increasing availability of fixed-price generation because a reduction in the expected spot price only affects the demand in excess of the capacity available from the fixed-price generation. In addition, we note that as the retail price of electricity paid to the retailer by its customers increases, the retailer interrupts at higher loads because the cost of interruption increases with the retail price. Finally, without yearly limits, the retailer interrupts at relatively low expected loads. In particular, interruption occurs at expected loads below the transition point between the two regimes in the supply curve. This aggressive behavior can be attributed to the large cost to the retailer of ending up in the high-demand regime because the electricity spot price applies to all the electricity procured from the spot market. Different assumptions on price formation, such as a "pay-as-bid" market, might produce qualitatively different results.

4.2.2. No Limit on Daily Interruption. In the case with no daily interruption limit, but with a yearly interruption limit, we can prove a proposition similar to Proposition 1.

PROPOSITION 2. *In the case with no daily interruption limits for pay-in-advance and pay-as-you-go interruptible contracts, if $p_{reduced} \leqslant p_{retail} + p_{fine}$, then the optimal policy involves interrupting the pay-in-advance contracts until their yearly limit is exhausted, before interrupting any of the pay-as-you-go contracts.*

PROOF. Assume that it is optimal to interrupt some amount from the pay-as-you-go contracts, along some price path, prior to exhausting the pay-in-advance contracts. Then, it is easy to see that the value function can be improved by following the strategy in which the interruption amount from the pay-as-you-go contracts is transferred to the pay-in-advance contracts, if possible. If, later in the price path, the yearly limit of the pay-in-advance contract is exhausted, an equal load from the pay-as-you-go contract is interrupted instead. Because following this alternative strategy results in lower cost for each price path where the priority of the interruption of the pay-in-advance contract is violated, we have a contradiction for the optimality of the original strategy. □

4.3. Base-Case Interruptible Contracts

To further study the optimal interruption policy and the value of interruptible contracts when there are limits on both daily and yearly interruption, we specify base-case contracts for the different types of interruptible contracts. The parameter values for these base-case contracts have been chosen with ERCOT in mind. For both types of contracts, we consider the possibility of interruption during weekdays over the months of June, July, and August only, which is the period when interruption is most likely in ERCOT.

4.3.1. Pay-in-Advance Contract. In the base-case pay-in-advance contract, the electricity retailer offers a 15% reduction to the retail price of electricity, $p_{\text{reduced}} = 0.85 \times p_{\text{retail}}$, to the entire load under contract, $L_{\text{under\_contract}}$. In exchange, the retailer may interrupt up to 20% of the load under contract daily,

$$L_{\text{advance, daily}} = 0.2 \times L_{\text{under\_contract}},$$

up to 10 times per year,

$$L_{\text{advance, yearly}} = 2 \times L_{\text{under\_contract}}.$$

Under this type of contract, there is no additional fine paid by the retailer upon interruption, so that $p_{\text{fine}} = 0$.

4.3.2. Pay-as-You-Go Contract. In the base case of the pay-as-you-go contract, the electricity retailer does not offer any reduction in the retail price; i.e., $p_{\text{reduced}} = p_{\text{retail}}$. In exchange for the right to interrupt customer load, the retailer pays a fine of either \$150/MWh or \$600/MWh of interrupted electricity. In addition, the customer may be interrupted up to 10 times per year, $L_{\text{pago, yearly}} = 10 \times L_{\text{pago, daily}}$.

4.4. Optimal Interruption Policy with Daily and Yearly Limits

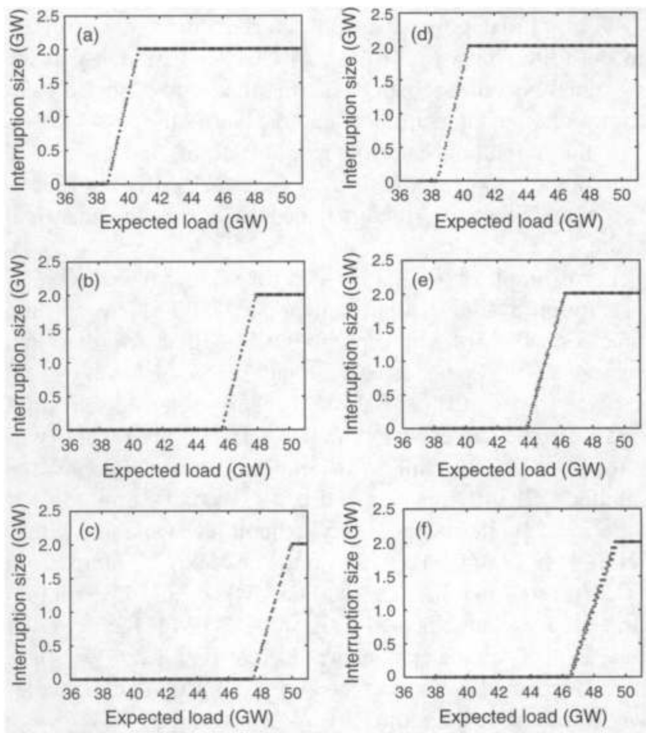
When there are both daily and yearly limits on both types of contracts, there is no generalization of Propositions 1 and 2 because one may want to avoid exhausting the pay-in-advance contracts to be able to interrupt larger amounts on some days. However, one can still say that pay-in-advance contracts will tend to be interrupted before pay-as-you-go contracts. The only violation to this order will occur when the remaining amount of interruption left in the pay-in-advance contracts is small, and when the daily limit of the pay-as-you-go contracts is small compared to the anticipated needs of daily interruption.

With both daily and yearly limits, the decision to interrupt becomes a choice between interrupting now versus waiting to interrupt later. This problem is similar to the problem of optimal early exercise of a financial option, with the additional complication that multiple exercises are possible, and that the amount exercised is an additional choice variable. This type of option is similar to the *swing option*, common in the natural gas and electricity markets. See Jaillet et al. (2004) for a valuation framework for the swing option.

Given the difficulty in solving a stochastic dynamic programming problem with many state and choice variables, we consider only one type of contract at a time. That is, we specialize to the situation where the electricity retailer has either pay-in-advance contracts or pay-as-you-go contracts, but not both. The numerical algorithm is described in detail in the online appendix. One of the difficulties in considering both types of contracts simultaneously lies in the fact that, under our framework, with a single contract type, we are able to reduce the problem to one with a single stochastic factor and one choice variable, as described in the online appendix. This reduction fails when both contract types coexist, increasing the stochastic factors to three and the choice variables to two. Such a high-dimensional problem can potentially be studied using methods similar to those discussed in Schultz (2003), where the possible random outcomes are approximated by a discrete set.

4.4.1. Pay-in-Advance Contracts. In Figure 5, we provide the optimal interruption strategy for pay-in-advance contracts with yearly and daily limits. The plots 5(a), 5(b), and 5(c) correspond to a pay-in-advance contract 60 days before the end of August, while the plots 5(d), 5(e), and 5(f) correspond to 30 days before the end of August. The figure shows the results for different amounts of interruption available. All plots are for an electricity retailer with 35 GW of generation available at a fixed price, who charges a retail price of \$60/MWh to its customers and with a daily limit on interruption of 2 GW. The plots 5(a) and 5(d) correspond to an unlimited amount of interruption remaining, the plots 5(b) and 5(e) to 20 GW of interruption remaining, and the plots 5(c) and 5(f) to 5 GW of interruption remaining. The pay-in-advance contract provides a discount of 15% to the entire load under contract.

Figure 5. Interruption strategy as a function of the expected load for pay-in-advance contracts.



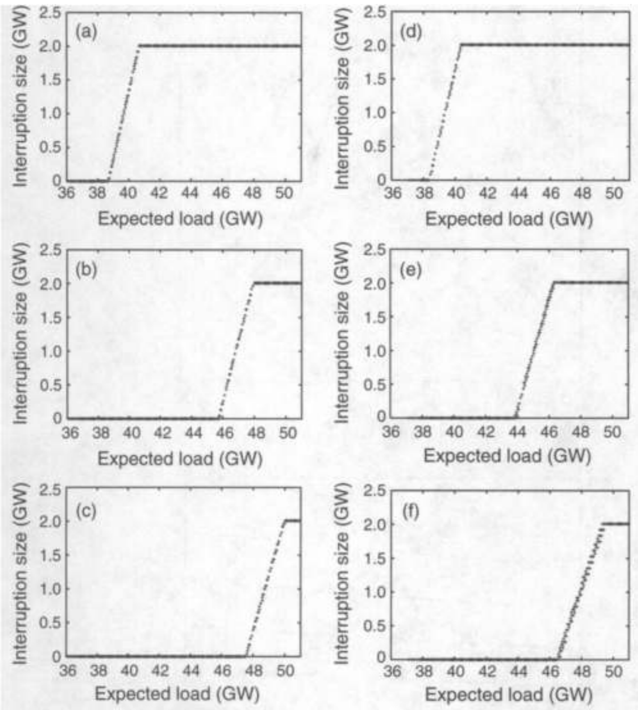
Notes. (a) 60 days prior to the end of August, unlimited amount of interruption remaining; (b) 60 days prior to the end of August, 20 GWs of interruption remaining; (c) 60 days prior to the end of August, 5 GWs of interruption remaining; (d) 60 days prior to the end of August, unlimited amount of interruption remaining; (e) 60 days prior to the end of August, 20 GWs of interruption remaining; (f) 60 days prior to the end of August, 5 GWs of interruption remaining.

From Figure 5, we notice that the most significant difference between the contract with yearly limits and the contract without yearly limits is that with yearly limits interruption occurs at higher expected loads. In particular, when the amount of interruptible load decreases, interruption occurs at higher expected loads. As expected, for the same level of remaining interruptible load, interruption occurs at lower expected loads closer to the end of the summer.

In addition, we note that the interruption policy is "fuzzy." The fuzziness is evident in plot 5(f) and is due to the fact that the optimal policy depends on two state variables, rather than just the expected load (these state variables are the deviation from historical temperatures at times t and $t-1$). Moreover, the slope of the interruption policy with respect to the expected load is increasing in the amount of remaining interruptible load. This is in line with the intuition that when smaller interruption amounts are available, the retailer waits longer before exhausting them, which in turn implies that the marginal value of interruption decreases as more interruptible load becomes available.

4.4.2. Pay-as-You-Go Contracts. In Figure 6, we provide the optimal interruption strategy for pay-as-you-go

Figure 6. Interruption strategy as a function of the expected load for pay-as-you-go contracts.



Notes. (a) 60 days prior to the end of August, unlimited amount of interruption remaining; (b) 60 days prior to the end of August, 20 GWs of interruption remaining; (c) 60 days prior to the end of August, 5 GWs of interruption remaining; (d) 60 days prior to the end of August, unlimited amount of interruption remaining; (e) 60 days prior to the end of August, 20 GWs of interruption remaining; (f) 60 days prior to the end of August, 5 GWs of interruption remaining.

contracts with yearly and daily limits. The figure provides the interruption strategy as a function of the expected load for the following day for pay-as-you-go contracts. The plots 6(a), 6(b), and 6(c) correspond to a pay-as-you-go contract 60 days before the end of August, while the plots 6(d), 6(e), and 6(f) correspond to the same contract 30 days before the end of August. The plots 6(a) and 6(d) correspond to an unlimited amount of interruption remaining; the plots 6(b) and 6(e) to 20 GW of interruption remaining; and the plots 6(c) and 6(f) to 5 GW of interruption remaining. For all contracts, the retailer has 35 GW of generation available, and the retail price is \$60/MWh. The daily amount that can be interrupted is 2 GW. The pay-as-you-go contract pays \$150/MWh of interruption.

In calculations we do not report, we verified that the interruption policy for the pay-as-you-go contracts is, both qualitatively and quantitatively, very similar to the interruption policy for the pay-in-advance contracts for reasonable parameter ranges. The result is not surprising because, intuitively, the two types of contracts are very similar, with the only difference being that the marginal cost of interrupting pay-as-you-go contracts is greater than the marginal cost of interrupting pay-in-advance contracts, due to the fine per unit of load interrupted under a pay-as-you-go contract.

4.5. Value of Interruptible Contracts

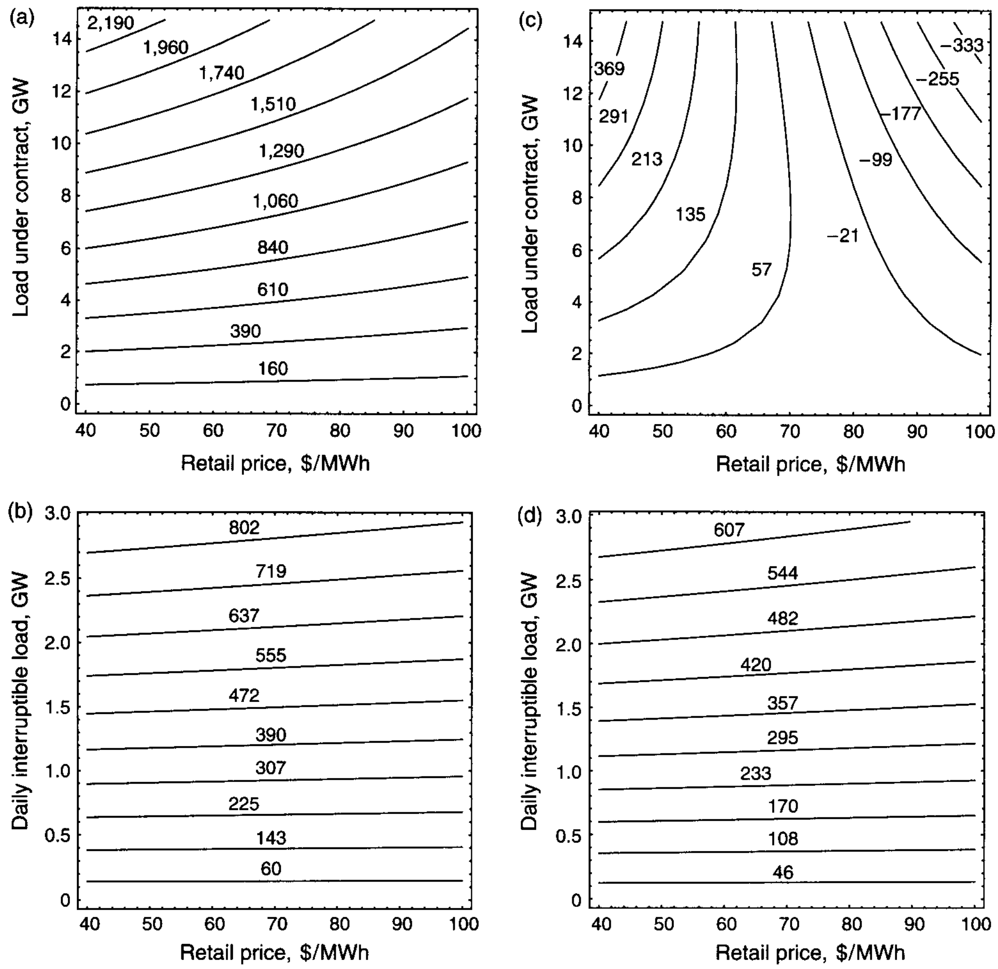
We define the value of an interruptible contract as the difference in the value function of the electricity retailer between having the interruptible contract and not having the interruptible contract. The value function for a retailer with no interruptible contracts can be easily calculated using Monte Carlo simulation because no choice variables are involved in that case.

In Figure 7, we present contour plots of the value of interruptible contracts. The plots 7(a) and 7(c) correspond to pay-in-advance interruptible contracts as the retail price that an electricity retailer charges and the total load under contract change. The discount provided to the entire load under contract is 15% from the retail price, the load available for interruption is equal to 20% of the load under contract, and interruption can occur up to 10 times. The plot 7(a) corresponds to a retailer that has no generation available at a fixed price and is forced to serve all the load from the spot market. The plot 7(c) corresponds to a retailer that has 35 GW of generation available. The plots 7(b) and 7(d) correspond to pay-as-you-go interruptible contracts, where interruption can occur up to 10 times and the retailer has 35 GW of generation available. The plot 7(b) corresponds to a contract with a fine of \$150/MWh of interrupted load, and the plot 7(d) to a contract with a fine of \$600/MWh. The value of the interruptible contracts is in millions of dollars.

From the figure, we notice that the amount of generation available to the retailer at a fixed cost is an important determinant of the price of an interruptible contract. In particular, when the retailer has no generation available, interruptible contracts are worth much more than when the retailer has 35 GW of generation available. The intuition behind this result is clear: If a retailer only has very small amounts of generation available, then interruption is very valuable, as it reduces both the amount of electricity bought in the spot market and the spot price itself. On the other hand, when the generation amount available is large, interruption is not as valuable because it occurs less often and the marginal amount bought in the spot market is smaller. The same intuition indicates that the marginal value of interruption decreases as a larger interruptible load is signed, i.e., interruptible contracts are more valuable when the retailer has little or no load available for interruption. In other words, generation and interruptible contracts are partial substitutes.

An additional factor that is important in the determination of the value of an interruptible contract is the fixed retail price charged by the electricity retailer to its customers. Intuitively, the higher the retail price, the higher the marginal cost of interruption, the fewer the interruptions, and the lower the value of an interruptible contract. This effect is seen in Figure 7, where, keeping the interruptible load fixed, the value of the interruptible contract decreases with the retail price.

Figure 7. Contour plots of the value of interruptible contracts (in millions of $).



Notes. (a) pay-in-advance contract, no generation available; (b) pay-as-you-go contract, \$150/MWh penalty; (c) pay-in-advance contract, 35 GWs of generation available; (d) pay-as-you-go contract, \$600/MWh penalty.

In addition to the dependence of the value of the interruptible contract to the amount of interruption available and the retail price, Figure 7 reveals that there is a big difference between the value of pay-in-advance and pay-as-you-go interruptible contracts. Pay-as-you-go interruptible contracts always have a positive value because payment and interruption are only made if interruption is to the benefit of the retailer. In contrast, it is possible that the value of pay-in-advance contracts is negative. Intuitively, because a large part of the cost of a pay-in-advance contract is provided upfront and is sunk, if the retailer signs up too large a load under a pay-in-advance interruptible contract, then the reduction in income due to the discount on the retail price is higher than the value added by the interruptible contract. For example, from Figure 7, we note that when the amount of generation available for a fixed price is 35 GW, the value of an interruptible pay-in-advance contract for an amount of interruption of 5 GW is positive for retail prices below \$80/MWh and negative for retail prices above \$80/MWh. The hyperbolic-looking level curves for the value of the contract are due to the fact that the value is identically zero when the load under contract is zero, and decreases as the retail price or the load under contract increases.

In the case of the pay-as-you-go contract, on the other hand, the value of the contract is positive because payment is made only after it is optimal to interrupt. This result, together with the intuition developed in §4.2, suggests that a retailer prefers interrupting pay-in-advance contracts before pay-as-you-go contracts, ceteris paribus.

5. Symmetric Equilibrium with Multiple Electricity Retailers

We have so far considered the case of a single electricity retailer who is able to use interruptible contracts to lower demand. This situation corresponds closely to partly regulated electricity markets, such as the one in Mexico.[6] However, in markets in the United States, Europe (United Kingdom, Norway), and the Pacific (Australia, New Zealand), there are often several retailers exposed to the same spot prices and each retailer may have separate interruptible contracts. In such a situation, each retailer would like the other retailers to exercise their interruptible contracts to lower overall demand without paying the costs associated with interruption. In a competitive market, coordination failure results, with each retailer interrupting amounts that, overall, are smaller than the amounts that would be interrupted by a single retailer, or by colluding retailers.

5.1. Framework

We illustrate and quantify the coordination failure in the simple case where all retailers are identical and each retailer has daily limits on the amount of interruption, but does not have a limit on the amount interrupted over the entire period. Under this scenario, the interruption decision does not depend on past behavior, and the problem is reduced to determining the optimal interruption strategy in a single day, given the daily interruption limits.

We consider the case of n identical electricity retailers, where each retailer has the same amount of generation available, receives the same retail price on electricity sold to consumers, and has signed identical pay-as-you-go interruptible contracts, each with its own customers. In addition, we assume that demand is equally divided between retailers. The profit function $\pi^{(i)}$ on a single day for retailer i is given by

$$\frac{\pi^{(i)}(L, l^{(1)}, \dots, l^{(n)})}{16} = \left(\frac{L}{n} - l^{(i)}\right) p_{\text{retail}} - l^{(i)} p_{\text{fine}} - L^{(i)}_{\text{generation}} p_{\text{generation}} - \left(\frac{L}{n} - l^{(i)} - L^{(i)}_{\text{generation}}\right) p_{\text{spot}}, \tag{7}$$

where p_{spot} is the spot price of electricity, which depends on the load L and the entire amount of interruption $\sum_{i=1}^{n} l^{(i)}$. The load served by each retailer is L/n. The term $(L/n - l^{(i)}) p_{\text{retail}}$ corresponds to the revenue to retailer i from selling electricity to its consumers. The term $l^{(i)} p_{\text{fine}}$ corresponds to the cost to retailer i of interrupting an amount $l^{(i)}$. The term $L^{(i)}_{\text{generation}} p_{\text{generation}}$ corresponds to the cost of procuring the generation available to the ith retailer at a fixed price. The term $(L/n - l^{(i)} - L^{(i)}_{\text{generation}}) p_{\text{spot}}$ corresponds to the cost of purchasing the excess electricity in the spot market.

Each retailer maximizes its expected profit by choosing the amount of load to interrupt $l^{(i)}$. The Nash equilibrium can be found by each retailer assuming that every other retailer interrupts an amount $l^{(j)^*}$, $i \neq j$, and then choosing the amount it interrupts, to maximize its own profit. The first-order condition is given by

$$\frac{\partial}{\partial l^{(i)}} \mathbb{E}(\pi^{(i)}) = 16 \frac{\partial}{\partial l^{(i)}} \mathbb{E}\Bigg(\left(\frac{L}{n} - l^{(i)}\right) p_{\text{retail}} - l^{(i)} p_{\text{fine}} - L^{(i)}_{\text{generation}} p_{\text{generation}} - \left(\frac{L}{n} - l^{(i)} - L^{(i)}_{\text{generation}}\right) p_{\text{spot}} \Bigg) = 0 \tag{8}$$

or,

$$p_{\text{retail}} + p_{\text{fine}} = -\frac{\partial}{\partial l_i} \mathbb{E}\left(\left(\frac{L}{n} - l^{(i)} - L^{(i)}_{\text{generation}}\right) p_{\text{spot}} \right). \tag{9}$$

Because each retailer faces an identical problem, there is a symmetric equilibrium, where each retailer interrupts an amount $l^{(i)} = l^*$ satisfying

$$p_{\text{retail}} + p_{\text{fine}} = -\frac{\partial}{\partial l} \mathbb{E}\left(\left(\frac{L}{n} - l^{(i)} - L^{(i)}_{\text{generation}}\right) p_{\text{spot}} \right) \Bigg|_{l^{(i)} = l^*}, \tag{10}$$

where p_{spot} is a function of the total load after interruption, $L - l^{(i)} - \sum_{i=1,\, i\neq j}^{n} l^* = L - l^{(i)} - (n-1) l^*$.

In the case of symmetric retailers, the solution of Equation (10) is relatively simple because each retailer faces the same problem. The objective function of each retailer is similar to that of the monopsonistic retailer with two following modifications: Each retailer serves an equal fraction of the total load, and each retailer's interruption has a smaller effect on the spot price as the number of retailers increase. Because the equilibrium is symmetric, the problem can be solved using a representative agent, which formally reduces the problem to the case of a monopsonistic retailer facing no limits on the interruption, studied in §4.2. Further details are provided in the online appendix.

5.2. Interruption Policy and Value of Interruptible Contracts

To compare cases with a different number of retailers, n, we set the total generation available to all the retailers at a fixed price, and the total daily interruptible load, equal to constants

$$nL^{(i)}_{\text{generation}} = L_{\text{generation}},$$

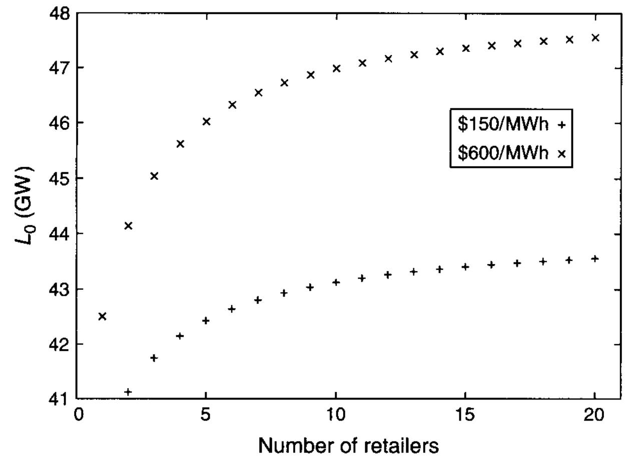
$$nl^{(i)} = \bar{l}.$$

Numerical results for the optimal amount of interruption, as well as the value of the interruptible contracts to each electricity retailer, for the parameter values calibrated from our model, for different numbers of identical retailers, and different amounts of generation available to each retailer at a fixed price, are presented in Figure 8 and Table 4.

In Figure 8, we present the load at which an electricity retailer starts to interrupt, L_0, as a function of the number of electricity retailers that serve the same area. The interruptible contracts are of the pay-as-you-go type. The total amount of generation available, $L_{\text{generation}}$, is 35 GW, and the retail price is \$60/MWh. The total daily amount that can be interrupted, $\bar{l}$, is 2 GW, and there is no global limit. The pay-as-you-go contract pays either \$150/MWh or \$600/MWh of interruption.

From Figure 8, we note that as the number of competitors increases, interruption occurs at higher expected loads.

Figure 8. Load at which an electricity retailer starts to interrupt as a function of the number of electricity retailers.



The effect is more pronounced at higher values of the fine per unit of interrupted load. For example, with five competitors, interruption occurs at an expected load of 42 GW when the fine is $150/MWh, but at an expected load of 46 GW when the fine is $600/MWh.

This behavior is also reflected in the results reported for the value of the interruptible contracts in Table 4. The values presented in the table are expressed as a percentage of the value of the interruptible contract when there is a single retailer. The total amount of generation available is either zero or 35 GW, and the retail price is $60/MWh. The total daily amount that can be interrupted is 2 GW, and there is no global limit. The pay-as-you-go contract pays either $150/MWh or $600/MWh of interruption. We notice that increased competition decreases the value of interruptible contracts, and that this decrease is bigger when the generation available to the electricity retailer is higher, as well as when the cost per unit of interruption increases. The decrease in value is significant, and with just five competitors, for a fine of $600/MWh of interrupted load, the value of an interruptible contract drops up to 46% in the case where the total amount of generation available is 35 GW. We note, however, that when there is no generation available to the retailers, interruptible contracts remain very valuable.

Table 4 also presents results in the limit of infinitely many identical electricity retailers, each one of infinitesimal size. In this limit, each retailer effectively acts as a price taker because interruption by any one retailer does not impact the spot price. The combined value of all the interruptible contracts is significantly lower than in the case of strategic behavior by a few large retailers. This result confirms that, in the case of a few large retailers, it is important to consider the impact of each retailer's actions on the spot electricity price. Acting as a price taker in such a situation would result in significant errors in both the choice of the interruption policy and the valuation of interruptible contracts.

6. Conclusions

We have presented a structural model of electricity prices and a framework for valuing interruptible contracts. In our structural model, supply and demand are stochastic processes whose parameters are statistically estimated to obtain a model for the spot electricity price. The advantage of a structural over a reduced-form model is to allow interaction between decisions of market participants and spot electricity prices. In the context of our paper, this interaction is crucial, as optimal interruption reduces both the demand for electricity and the spot electricity price. The use of a structural model has also enabled us to study the impact of competition on both the value of interruptible contracts and on the optimal interruption policy.

We valued interruptible contracts from the point of view of retailers of electricity. Our analysis suggests that in the absence of forward or bilateral contracts, or ownership of generation assets, the interruptible contracts are quite valuable and the retailer interrupts aggressively. As more generation is available at a fixed price, or as the number of competing retailers increases, the value of interruptible contracts diminishes, and interruption occurs at higher expected loads. This result has important implications for electricity retailers and sheds some light on the reason for

Table 4. Value of interruptible contracts under competition.

Competitors	$L_{\text{generation}}, p_{\text{Fine}}$ 0 GW, $150	$L_{\text{generation}}, p_{\text{Fine}}$ 0 GW, $600	$L_{\text{generation}}, p_{\text{Fine}}$ 35 GW, $150	$L_{\text{generation}}, p_{\text{Fine}}$ 35 GW, $600
1	$3,940 MM	$3,120 MM	$920 MM	$440 MM
2	98%	93%	95%	85%
3	97%	87%	92%	71%
4	95%	80%	89%	61%
5	93%	74%	87%	54%
10	85%	50%	80%	39%
40	69%	20%	73%	25%
∞	63%	14%	71%	23%

Notes. Values for multiple retailers are expressed as a percentage of the value for a single retailer. Values for a single retailer in millions of dollars.

the use of interruptible contracts in California, where, after deregulation, retailers had only limited generating resources available.

We studied two types of contracts: the pay-in-advance contract, in which the retailer agrees to a discount for the entire load of a customer in exchange for the right to interrupt part of the load a certain number of times; and the pay-as-you-go contract, where the retailer compensates the customer for the interrupted load upon interruption. Given a choice between different types of interruptible contracts, pay-as-you-go contracts are preferable to the retailer because, due to the advance payment of the pay-in-advance contracts, it is possible in cases where the retailer signs up too large an interruptible load that the value of the interruptible pay-in-advance contract is negative, while, on the other hand, the value is always positive for the pay-as-you-go contracts. Our methodology can be combined with information on customer preferences regarding types of interruptible contracts to decide the optimal design and mix of different contract types.

Other than valuing interruptible contracts, the structural model we have presented can be useful in the optimal asset allocation problem for an electricity retailer that can choose among generation plants, forward contracts, bilateral contracts, options, and interruptible contracts, as well as in the optimal design of new types of contracts.[7] As well as valuing interruptible contracts, our model can be used as a small part in a larger optimal allocation problem, where the electricity retailer determines the optimal mix of generation assets and interruptible contracts. We plan to explore these problems in future research.

Endnotes

1. For a general introduction to restructured electricity markets, see Stoft (2002). For a description of "Standard Market Design" of restructured electricity markets as envisaged by the U.S. Federal Energy Regulatory Commission, see United States of America Federal Energy Regulatory Commission (2002).

2. Rassenti et al. (2002) provide experimental evidence demonstrating that the use of interruptible contracts is an effective way of reducing or even eliminating strategic behavior on the part of electricity generators.

3. The work of Kamat and Oren (2002) can be extended to accommodate multiple interruptions when there is no limit on the total number of interruptions. In addition, Kamat and Oren (2002) allow for multiple notification times and provide closed-form solutions.

4. This market setting is very similar to the one faced by Pacific Gas and Electric and Southern California Edison shortly after electricity deregulation in California. It is also similar to the situation faced by those retailers in ERCOT who choose to meet their demand through purchases from the "balancing market." One such ERCOT retailer, Texas Commercial Energy, which relied primarily on balancing market purchases and did not have any interruptible contracts, went bankrupt after being exposed to high balancing market prices in February 2003.

5. Kamat and Oren (2002) discuss the valuation of interruptible contracts with multiple notification times, rather than multiple interruptions, in the context of a reduced-form model of electricity prices.

6. The Mexican market is regulated on the distribution side, where all electricity retailers are owned by the Mexican government. These retailers also own significant amounts of generation. However, there are additional private generators that typically have long-term contracts with the retailers, and which are called upon at times of high demand.

7. Fahrioglu and Alvarado (2000, 2001) discuss methods for an electricity retailer to estimate the demand among its customers for interruptible contracts and describe an incentive structure that encourages customers to reveal their true value of power.

Acknowledgments

The authors acknowledge the helpful comments and suggestions of seminar participants at the University of Texas at Austin, King's College, the conference on modeling, optimization and risk management in finance at the University of Florida, the 7th Annual Conference on Real Options in Washington, DC, the 2003 INFORMS annual meeting in Atlanta, the Mexican Electricity Regulatory Commission (Comisión Reguladora de Energía), and the workshop on energy risk and energy derivatives in Paris. They also thank Andres Almazan, Aydoğan Altı, and Tridib Sharma for discussions. The authors are thankful for financial support from NSF grant 0082506. Earlier versions of this paper were titled "Valuation and Optimal Interruption for Interruptible Electricity Contracts."

References

Baldick, Ross, Hui Niu. 2005. Lessons learned: The Texas experience. James Griffin, Steven Puller, eds. *Electricity Deregulation: Where to from Here?* University of Chicago Press, Chicago, IL, 182–224.

Barlow, Martin T. 2002. A diffusion model for electricity prices. *Math. Finance* **12** 67–70.

Borenstein, Severin. 2005. Time-varying retail electricity prices: Theory and practice. James Griffin, Steven Puller, eds. *Electricity Deregulation: Where to from Here?* University of Chicago Press, Chicago, IL, 317–357.

Brown, Gardner, Jr., M. Bruce Johnson. 1969. Public utility pricing and output under risk. *Amer. Econom. Rev.* **59** 119–128.

California Public Utilities Commission. 2001. Energy division's report on interruptible programs and rotating outages. Retrieved June 13, 2006. http://www.cpuc.ca.gov/PUBLISHED/REPORT/5119.htm.

Campbell, Sean D., Francis X. Diebold. 2005. Weather forecasting for weather derivatives. *J. Amer. Statist. Assoc.* **100** 6–16.

Cao, Melanie, Jason Wei. 2000. Pricing weather derivatives: An intuitive and practical approach. *Risk* 67–70.

Cao, Melanie, Jason Wei. 2004. Weather derivatives valuation and market price of weather risk. *J. Futures Markets* **24** 1065–1089.

Caves, Douglas W., Joseph A. Herriges. 1992. Optimal dispatch of interruptible and curtailable service options. *Oper. Res.* **40** 104–112.

Chao, Hung-po, Robert Wilson. 1987. Priority service: Pricing, investment, and market organization. *Amer. Econom. Rev.* **77** 899–916.

Chao, Hung-po, Shmuel S. Oren, Stephen A. Smith, Robert B. Wilson. 1988. Priority service: Market structure and competition. *Energy J.* **19** 77–104.

Deng, Shijie. 1999. Stochastic models of energy commodity prices: Mean-reversion with jumps and spikes. Working paper, Georgia Institute of Technology, Atlanta, GA.

Deng, Shijie. 2000. Pricing electricity derivatives under alternative spot price models. *Proc. 33rd Hawaii Internat. Conf. System Sci.* University of Hawaii, Honolulu, HI.

Eydeland, Alexander, Krzysztof Wolyniec. 2003. *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. John Wiley and Sons, Hoboken, NJ.

Fahrioglu, Murat, Fernando L. Alvarado. 2000. Designing incentive compatible contracts for effective demand management. *IEEE Trans. Power Systems* **15** 1255–1260.

Fahrioglu, Murat, Fernando L. Alvarado. 2001. Using utility information to calibrate customer demand management behavior models. *IEEE Trans. Power Systems* **16** 317–322.

Galassi, Mark, Jim Davies, James Theiler, Brian Gough, Gerard Jungman, Michael Booth, Fabrice Rossi. 2003. *Gnu Scientific Library: Reference Manual*, 2nd ed. Network Theory Ltd., Bristol, UK.

Geman, Hélyette, Andrea Roncoroni. 2006. Understanding the fine structure of electricity prices. *J. Bus.* **79**(3) 1225–1262.

Jaillet, Patrick, Ehud I. Ronn, Stathis Tompaidis. 2004. Valuation of commodity-based swing options. *Management Sci.* **50** 909–921.

Kamat, Rajnish, Shmuel S. Oren. 2002. Exotic options for interruptible electricity supply contracts. *Oper. Res.* **50** 835–850.

Kholodnyi, Valery A. 2004. Valuation and hedging of European contingent claims on power with spikes: A non-Markovian approach. *J. Engrg. Math.* **49** 233–252.

Lee, Maurice W. 1953. Hydroelectric power in the Columbia basin. *J. Bus. Univ. Chicago* **26** 173–189.

Oren, Shmuel S., Stephen A. Smith. 1992. Design and management of curtailable electricity service to reduce annual peaks. *Oper. Res.* **40** 213–228.

Pilipovic, Dragana. 1997. *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill, New York.

Rassenti, Stephen J., Vernon L. Smith, Bart J. Wilson. 2002. Using experiments to inform the privatization/deregulation movement in electricity. *Cato J.* **21** 515–544.

Ratkowsky, David A. 1983. *Nonlinear Regression Modeling*. Marcel Dekker, New York.

Raver, Paul J. 1951. A public administrator's view. *Amer. Econom. Rev.* **41** 289–298.

Schultz, Rüdiger. 2003. Stochastic programming with integer variables. *Math. Programming Ser. B* **97** 285–309.

Skantze, Petter, Andrej Gubina, Marija Ilic. 2000. Bid-based stochastic model for electricity prices: The impact of fundamental drivers on market dynamics. MIT Energy Laboratory Report EL 00-004, Massachusetts Institute of Technology, Cambridge, MA.

Smith, Nelson Lee. 1946. Rate regulation by the Federal Power Commission. *Amer. Econom. Rev.* **36** 405–425.

Stoft, Steven. 2002. *Power System Economics: Designing Markets for Electricity*. IEEE Press, Wiley Interscience, John Wiley and Sons, Piscataway, NJ.

Troxel, Emery. 1949. Inflation in price-regulated industries. *J. Bus. Univ. Chicago* **22** 1–16.

Tschirhart, John, Frank Jen. 1979. Behavior of a monopoly offering interruptible service. *Bell J. Econom.* 244–258.

United States of America Federal Energy Regulatory Commission. 2002. Standard market design and structure notice of proposed rulemaking. 18 CFR part 35, docket RM01-12-000.