# Structural Estimation of Switching Costs for Peaking Power Plants

Stein-Erik Fleten[*], Erik Haugom[†], Alois Pichler[‡], Carl J. Ullrich[§]

November 3, 2017

**Abstract**

This paper estimates costs associated with mothballing, restarting, abandoning and maintaining peaking power plants. The paper develops a real options model to explain switching and maintenance behavior of plant managers. The constrained optimization approach to estimate crucial costs accommodates non-parametric dynamics for the expectations of the plant managers regarding future profitability.

The empirical analysis is based on a database of the annually reported status of power plants to the United States Energy Information Administration (EIA) during 2001-2009. We arrive at economically meaningful estimates of maintenance costs and switching costs, and discuss these in light of rates used in the Pennsylvania-New Jersey-Maryland capacity market.

**Keywords: Dynamic discrete choice models, Real options, Dynamic programming, Irreversible investment, Electricity markets.**
**JEL Codes: C14, C61, D22, D92, G13, G31, H41, Q40**

## 1 Introduction

We use nonparametric structural estimation to estimate the costs associated with shutting down, starting up, and abandoning peaking power plants, specifically simple cycle combustion turbines (CTs). Estimates of switching costs are surprisingly difficult to obtain in practice.

Our case study is made possible by the availability of detailed power plant data from the United States. Each year the owners of existing power plants must file Form 860 with the Energy Information Administration (EIA). From these data it is possible to determine whether an existing plant was shutdown, started up, or abandoned. Our sample includes 8189 plant-year observations from the period 2001–2009. These data are augmented with time series of electricity prices and fuel prices, available from electricity market operators and the EIA.

We use the discrete decision process framework of Rust (1987). The papers by Gamba and Tesser (2009) and Su and Judd (2012) also consider parametric stochastic processes – processes involving exponential distributions or geometric Brownian motions – to model the underlying state variable. We employ the estimator developed by Su and Judd (2012). Intrinsic to this framework is the use of a shock process, reflecting the unobserved heterogeneity across plants and over time.

[*]Norwegian University of Science and Technology, NO–7491, Trondheim, Norway. Corresponding author. Contact: stein-erik.fleten@iot.ntnu.no

[†]Inland Norway University of Applied Sciences, Lillehammer, Norway; Norwegian University of Science and Technology, NO–7491, Trondheim, Norway. Contact: erik.haugom@inn.no

[‡]Norwegian University of Science and Technology, NO–7491, Trondheim, Norway. Contact: alois.pichler@mathematik.tu-chemnitz.de

[§]James Madison University, Department of Finance and Business Law, Harrisonburg, VA, 22807. Contact: ullriccj@jmu.edu

We modify the Su and Judd (2012) approach by capturing the dynamics of the exogenous state variable using a nonparametric kernel density estimator.[1] The nonparametric kernel density estimator does not require unrealistic assumptions about the data generating process. Instead, the time series of observed state variable transitions are used directly to estimate managers' expectations regarding future profitability.

In contrast to the existing literature on structural estimation, we do not normalize to unit value the scaling parameter of the unobserved payoff shock process intrinsic to structural estimation. Instead, we present an estimator for the scaling parameter.

Of key concern in empirical analysis of irreversible investment is the endogeneity of plant manager's decisions. The firm-specific value of the stream of future revenues and costs associated with being in an operating state and possibly switching to another is not observable, and the decision to switch (or not) must be regarded as endogenous. Another concern is that plant managers have a better view on the decisions at hand than the analysts. They know more in detail the status of the profitability of the plant in question, such as the technical condition, the behavior and efficiency of nearby competitors, relevant regulatory or local market conditions, and firm policies.[2]

Understanding shutdown, startup, and abadonment decisions is important for designing efficient mechanisms in electricity capacity markets. In an effort to provide incentive for firms to build and maintain sufficient peaking capacity, Independent System Operators in the United States recently have introduced capacity markets such as the Reliability Pricing Model (RPM) in the Pennsylvania–New Jersey–Maryland (PJM) system. Capacity markets provide revenue to plants for maintaining availability and therefore help to ensure system reliability.

Participants in RPM bid an Avoidable Cost Rate (ACR). Avoidable costs are the incremental costs of being a capacity resource, i.e., the costs which could be avoided if a particular plant were shut down for a year. Owners of power plants may either develop estimates of these costs for each individual plant or use default rates provided by the market. From our switching and maintenance costs we estimate ACRs. Our estimates of ACRs are less than the default values used in PJM's capacity market, implying that consumers may be paying too much for system reliability.

**Outline of the paper.** Section 2 provides motivation and institutional background. Section 3 describes our data sample. Section 4 addresses our empirical strategy, including the nonparametric approach to model the transitions. Resulting cost estimates are discussed in Section 5, with a comparison to standard cost rates. In Section 6 we provide policy implications.

## 2 Motivation and background

Because electricity is not yet storable in meaningful quantities, supply and demand must balance in real-time. For storable commodities such as gasoline, supply and demand shocks are at least partially absorbed by inventory. Lack of inventory in electricity markets leads to high price volatility as shocks flow directly through to prices. (Ullrich, 2012)

Non-storability also has important consequences for electric system reliability. System operators must maintain reserve generators, i.e., peaking plants, which are able to produce electrical energy on short notice. Non-storability combined with price inelasticity means that system operators also need generators

[1] An early attempt to establish nonparametric models in the context of structural estimation is Bansal et al. (1995), who nonetheless approximate a nonparametric density by a parametric family. Newey et al. (1999) and Guerre et al. (2000) estimate a univariate density nonparametrically; the density then serves as an input parameter for the structural model. Bontemps et al. (2000) and Li et al. (2002) extend this approach to multivariate densities and provide convergence rates. In contrast Musalem et al. (2010) choose a Bayesian approach.

[2] See Appendix B in Fleten et al. (2017) for a discussion of the real world problems associated with determining the costs of shutting down and starting up a peaking power plant.

which can vary output in real-time in order to match changing demand, i.e., generators which can follow load. CTs are well suited for both roles and are in use worldwide.

**Renewable Energy** The penetration of renewable electricity sources such as solar and wind plants has increased as many jurisdictions pursue policies to decrease reliance on fossil fuels. Solar and wind power are intermittent, i.e., their availability and electrical output vary with ambient conditions. Real-time variations in the output of intermittent plants can make it harder to maintain reliability of the electrical grid. Smith and Cook (2015) explore the problems currently facing Hawaii, a state with a relatively large amount of renewable generation. According to the authors, "... *sudden swings in the output of solar and wind ... force the state's main utility to scramble to try to keep overall supply of power steady*."

As the amount of solar and wind capacity continues to increase, this problem is likely to worsen. PJM addressed this issue in a recent white paper ( PJM Interconnection LLC (2017), p.5).

> A marked decrease in operational reliability was observed for portfolios with significantly increased amounts of wind and solar capacity ... Nevertheless, PJM could maintain reliability with unprecedented levels of wind and solar resources, assuming a portfolio of other resources that provides a sufficient amount of reliability services.

CTs can provide the reliability services needed at high levels of solar and wind penetration. But, increased penetration of solar and wind plants also tends to reduce the value of CTs. Because solar and wind plants have zero or near-zero variable operating costs, they displace plants with higher variable costs.[3] As the value of a CT falls, the owner may exercise her option to shutdown the plant, i.e., put the plant into standby mode, in order to reduce maintenance costs.[4] The costs of maintaining a plant in standby mode are less than the costs of maintaining an operational plant. The owner of a plant which was previously shutdown holds the options to either restart the plant (if market conditions improve) or abandon the plant.[5]

**Switching Costs** Shutdown, startup, and abandonment entail one-time switching costs. These switching costs influence switching decisions. All else equal, higher switching costs lead to less switching. In practice, shutdown costs for CTs are near zero. Abandonment costs are negative as the owner can sell a used CT in the secondary market. Startup costs are difficult to estimate even for industry professionals.[6] There are few if any publicly available estimates of startup costs. Therefore one of the main goals of this paper is to estimate switching costs for CTs.

## 2.1 Definition of Shutdown, Startup, and Abandonment

The owners of power plants in the United States must each year file Form 860 with the EIA. Included in Form 860 is the status of each plant. For our purposes, the relevant statuses are Operating (OP), Standby (SB), and, Retired (RE). A plant in state OP is available for operation. A plant in state SB has been shut down and cannot be made ready for operation in the short term.[7] A plant in state RE has been abandoned

[3] As detailed in Caldecott and McDaniels (2014), the value of CT-based power plants in Europe has dropped significantly.

[4] Note that we use the term *shutdown* to refer to what is sometimes called mothballing or laying up in the real options literature. We do not use *shutdown* to refer to overnight cycling of plants.

[5] Brennan and Schwartz (1985) provides the basic framework for the real options to switch operating modes. Moel and Tufano (2002) use the Brennan and Schwartz model to examine operational switching of gold mines within a real options framework. Fleten, Haugom, and Ullrich (2017) use a reduced form model and find that regulatory uncertainty reduces the likelihood of shutdown and startup for CTs. Fleten and Näsäkkälä (2010) examine the value of switching options for peaking plants and find, among other things, that the option to abandon has relatively little value.

[6] We thank Steve Marshall of Lakeland Electric and Paul D. Clark II of the City of Tallahassee for sharing their insights and experience.

[7] The EIA provides variable definitions in a *Layout* file accompanying the EIA 860 data. The 2000 *Layout* file defines SB as "*Cold Standby (Reserve): deactivated (mothballed), in long-term storage and cannot be made available for service in a short period of time, usually requires three to six months to reactivate.*"

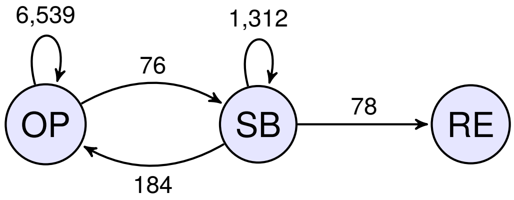


Figure 1: Transitions between the three states *operating* (OP), *standby* (SB) and *retired* (RE). Included is the number of observed transitions.

and cannot return to service. These data allow us to infer status changes, as follows.

- We define a *shutdown* to occur when a plant moves from state OP to state SB and label this transition $OP \rightarrow SB$.
- We define a *startup* to occur when a plant moves from state SB to state OP and label this transition $SB \rightarrow OP$.
- We define an *abandonment* to occur when a plant moves from state SB to state RE and label this transition $SB \rightarrow RE$.

Other possible (non)transitions include $OP \rightarrow OP$ and $SB \rightarrow SB$. Figure 1 summarizes the status changes in our dataset.

# 3 Data

The sample includes simple cycle combustion turbines in the northeastern part of the United States. The sample period is 2001–2009. See Fleten, Haugom, and Ullrich (2017) for further details about the choice of plants and geographic locations.

The cash flow for a power plant is determined by the spark spread, the difference between the price of electricity and the cost of fuel used to produce it. A peaking plant consists of a series of daily call options on the spark spread. The *heat rate* of a power plant is the amount of fuel, measured in millions of British thermal units (MMBtu), required to generate one megawatt hour (MW h) of electricity. A lower number indicates greater efficiency.

Our primary source of heat rate data is the Continuous Emissions Monitoring Systems (CEMS) data from the U.S. Environmental Protection Agency. The CEMS include both generation (MWh) and fuel use (MMBtu) data for individual generators, allowing us to calculate heat rates for individual generators.[8]

We use daily spot prices for New York Harbor No. 2 Oil and NYMEX Henry Hub natural gas. These data are available from the EIA website. Electricity prices come from the Independent System Operator (ISO) websites. Consistent with our focus on peaking plants, we use electricity prices for the peak period of the day, defined as the industry standard 16 hour period beginning at 06:00 and ending at 22:00.

Each year the EIA publishes its Annual Energy Outlook document. The accompanying assumptions document includes performance and cost estimates for new electric generating capacity. (See, for example, Cost and Performance Characteristics of New Central Station Electricity Generating Technologies in Energy Information Administration, page 89.) We use these data to estimate variable non-fuel generation costs. Later we use these data as a point of comparison for our results.

[8]Heat rate data were included in Form 860 for 1990-1995. We use these data when CEMS data are unavailable. For plants which neither appear in CEMS, nor have heat rate data in Form 860, we estimate heat rate based upon the age and size of the plant. See Appendix A in Fleten, Haugom, and Ullrich (2017) for the details.

| Transition | $OP \to OP$ | $OP \to SB$ | $SB \to OP$ | $SB \to SB$ | $SB \to RE$ | Total |
|---|---|---|---|---|---|---|
| Average ($/kW/year) | 12.6 | 5.4 | 17.7 | 12.3 | 2.1 | 12.5 |
| Standard deviation | 14.0 | 9.9 | 16.4 | 14.1 | 5.0 | 14.0 |
| Observations | 6539 | 76 | 184 | 1312 | 78 | 8189 |

Table 1: Summary statistics by transition for the profitability state variable $X$, in units of $/kW/year.

### 3.1 Profitability

Consider a plant which has heat rate $H$ in units of $\frac{\text{MMBtu}}{\text{MWh}}$. We calculate the plant-specific spark spread ($S_n$) expressed in units of dollars per megawatt hour ($/MWh), for day $n$ as

$$S_n = P_n^e - H * P_n^f - V,$$

where $P_n^e$ is the day $n$ electricity price ($/MWh), $P_n^f$ is the day $n$ fuel price ($/MMBtu), and $V$ ($/MWh) is the variable non-fuel generation cost.

Profitability per unit of capacity per year ($/kW/year) is the state variable $X$ in our optimization. The profitability for year $t$ is given by

$$X_t = \sum_{n=1}^{T_t} \max\left(S_n, 0\right) * \left(\frac{16}{1000\,\text{kW}\,\text{MW}^{-1}}\right),$$

where 16 is the number of peak hours in a day and $T_t$ is the number of days in year $t$. The max function captures the optionality of the plant. On days for which the spread is negative, the plant is assumed not operate and the profit is zero.

**Profitability Summary Statistics** We calculate profitability for all plants in the sample, operational or otherwise. For those plants which have status $SB$, the profitability is hypothetical. In this case $X_t$ is the profitability which would have obtained if the plant had been in state OP in year $t$. This counter-factual is a signal of potential profitability.

Table 1 presents summary statistics for profitability. Plants which shut down ($OP \to SB$) have relatively low profitability. Plants which start up ($SB \to OP$) have relatively high profitability. Plants which remain shut down ($SB \to SB$) have similar profitability to plants which remain in the operating mode ($OP \to OP$). Status changes – shut downs and start ups – happen only when there are large differences in profitability. Only very profitable plants are started up and only very unprofitable plants are shut down. Plants which are abandoned ($SB \to RE$) have lesser profitability still.

# 4 Empirical framework: Structural estimation

Structural estimation is a technique to uncover parameters hidden in a model of economic decision making. The parameters we want to estimate in our case study are switching costs and maintenance costs of peaking power plants, which are selected by the maximum likelihood approach. We refer to Table 2 for the economic interpretation of the variables.

[9] The index $i$ refers to the observation ($i = 1, \ldots,\ 8189$) while the index $t$ refers to the year. In what follows, $X_t$ is the profitability in year $t$ and is a random variable. $X_i$ is the realization of the random variable and is one component of a plant-year observation. The other two components each observation are $s_i$, the state of the plant in the current year, and $u_i$, the state of the plant in the upcoming year. The state of the plant in the upcoming year, $u \equiv s_{t+1}$, therefore represents the *decision* of the plant manager.

| Symbol | Description |
|---|---|
| $t$ | Time index; the unit time period is a year. |
| $X_t$ | The state process; in our specific case the state process is an indicator of profitability per unit of capacity per unit of time expressed in units of dollars per kilowatt-year, \$/kW/year |
| $(X_t, \varepsilon_t)$ | The augmented state process; the second process, $\varepsilon_t$, is not accessible to observation. |
| $s,\ u \in S$ | $S :=$ {operating, standby, retired} are operating states of the power plants and decided by the plant manager. |
| $(X_i, s_i, u_i)$ | An observation consists of a profitability $X_i$ during the current year, the state $s_i$ of the system in the current year, and $u_i$, the state of the system in the following year after the managers decision.[9] |
| $g(x, s;\ u)$ | The payoff during a single period. The payoff function $g(\cdot)$ comprises the expected cash flow for the next period and the costs associated with the transition from $s$ to $u$. |
| $V(x, s)$ | Value function – the accumulated discounted future payoffs achieved from an optimal policy. |
| $v(x, s)$ | Expected (or $s$-alternative-specific) value function. The function $v$ is the average of the different value functions $V$ among all agents operating a power plant in the market. |
| $\beta \in (0, 1)$ | Discount factor $\beta = \dfrac{1}{1 + \text{interest rate}}$. |

Table 2: Definition of variables

In what follows we develop the Bellman equation first, which we then extend to the s-alternative-specific value function. This is essential, as we use these results to develop non-parametric structural estimation (Section 4.3 below).

## 4.1 The real options problem

The value function of a dynamic decision problem satisfies a fixed-point equation, also known as Bellman equation. Bellman's equation is used to uncover the optimal policy, which depends on the current state $s_t$ solely. The Bellman equation for the value function $V$ is

$$\begin{aligned} V(x,s) &= \max_{s_t = s_t(X_t)} \mathbb{E}\left( \sum_{t=0}^{\infty} \beta^t g\left(X_t, s_t;\ s_{t+1}\right) \middle| X_0 = x \right) \\ &= \max_{s_t = s_t(X_t)} \mathbb{E}\left( g\left(X_0, s_0;\ s_1\right) + \beta \cdot \sum_{t=0}^{\infty} \beta^t g\left(X_{t+1}, s_{t+1};\ s_{t+2}\right) \middle| X_0 = x \right) \\ &= \max_{s_t = s_t(X_t)} \mathbb{E}\left( g\left(X_0, s_0;\ s_1\right) + \beta \cdot \mathbb{E}\left( \sum_{t=0}^{\infty} \beta^t g\left(X_{t+1}, s_{t+1};\ s_{t+2}\right) \middle| X_1 = x_1 \right) \middle| X_0 = x \right) \\ &= \max_{s_t = s_t(X_t)} \mathbb{E}\left( g\left(X_0, s_0;\ s_1\right) + \beta \cdot V\left(X_1, s_1\right) \middle| X_0 = x \right) \\ &= \max_{u \in S}\ g\left(x, s;\ u\right) + \beta \cdot \mathbb{E}\left( V\left(X_{t+1}, u\right) \middle| X_t = x \right), \end{aligned} \tag{1}$$

where the maximum is among all control sequences $s_t$ with $s_0 = s$ and which depend on $X_t$ only in the present Markovian setting; this dependency is denoted $s_t = s_t(X_t)$ in the maximum in the previous formula (1).

## 4.2 The s-alternative-specific value function

Every plant manager has additional understanding of his own operation, which is not known to the analyst. Structural estimation models this additional knowledge by augmenting the state space process $(X_t)$ with the parallel process $(\varepsilon_t)$ of additional information, a payoff shock. Every $\varepsilon_t = (\varepsilon_{t,u})_{u\in S}$ is a vector carrying the additional information which is associated with the actions $u \in S$. This information is hidden from the researcher who observes only the state $X_t$.

In analogy to (1), the augmented value function is

$$V(x, \varepsilon, s) = \max_{u\in S} g(x, \varepsilon, s;\, u) + \beta \cdot \mathbb{E}\left( \int V(X_1, \varepsilon_1, u)\mathcal{E}\left(\mathrm{d}\varepsilon_1|X_1\right) \middle| X_0 = x \right), \tag{2}$$

and its Bellman function is similar to (1).

The transition of the payoff shock $\varepsilon$, which is described by the distribution $\mathcal{E}$, is *independent* from $X$ by assumption and one may integrate on every fiber $\{X_0 = x\}$ separately (conditional independence, cf. Rust (1987)).

The *expected value function*, or *s-alternative-specific value function* is defined as

$$v(x, s) := \mathbb{E}\left( \int V\left(X_1, \varepsilon_1, s\right) \mathcal{E}\left(\mathrm{d}\varepsilon_1\right) \middle| X_0 = x \right), \tag{3}$$

it is the conditional expectation of the individual value functions. To apply structural estimation we deduce a fixed-point equation for the expected value function $v$. To do so we rewrite the Bellman equation (2)

$$V(x, \varepsilon, s) = \max_{u\in S}\ g(x, \varepsilon, s;\, u) + \beta \cdot v\left(x, u\right) \tag{4}$$

and, by taking expectations of (4), we get

$$\begin{aligned} v(x, s) &= \mathbb{E}\left( \int V\left(X_1, \varepsilon_1, s\right) \mathcal{E}\left(\mathrm{d}\varepsilon_1\right) \middle| X_0 = x \right) \\ &= \mathbb{E}\left( \int \max_{u\in S}\ g(X_1, \varepsilon_1, s;\, u) + \beta \cdot v(X_1, u)\mathcal{E}\left(\mathrm{d}\varepsilon_1\right) \middle| X_0 = x \right). \end{aligned} \tag{5}$$

The latter equation is a fixed-point equation for $v$, but in contrast to (1) the maximization and expectation are interchanged.

To manage the inner integral of a maximum one may specify the payoff function $g$ by accounting for the payoff shock in a linear way according to

$$g(x, \varepsilon, s, u) = g(x, s, u) + \varepsilon_u, \tag{6}$$

and by specifying the distribution of $\mathcal{E}$. Rust (1987) uses the term *additive separability* for the particular decomposition (6).

The integrand in (5) is a maximum, and it is well-known from extreme value theory that the normalized maximum, which is an extreme value, converges to an extreme value distribution (the Fisher–Tippett–Gnedenko theorem identifies three types of limiting extreme value distributions, cf. Embrechts et al. (1997)). The Gumbel variable is the only extreme value distribution with two-sided support. Given the maximizations in (4) and (5) the process $\varepsilon$ thus is a process of mutually independent Gumbel variables, which is independent from $X$. This observation allows an important simplification of formula (5). Indeed, the Gumbel distribution is closed under maximization, and in this case a closed form formula for expectations is available and given by

$$\int \max_{u \in S} (\varepsilon_u + c_u) \mathcal{E}(\mathrm{d}\varepsilon_u) = b \cdot \log \left( \sum_{u \in S} \exp \frac{c_u}{b} \right), \tag{7}$$

where $b$ is a scale parameter that controls the variance of the payoff shocks (cf. Remark 2); Equation (7) is detailed in Proposition 7 in the Appendix A.

Specifying $c_u := g(X_1, s; u) + \beta \cdot v(X_1, u)$ and applying (7) to

$$v(x, s) = \mathbb{E} \left( \int \max_{u \in S} \left( g(X_1, s;\ u) + \varepsilon_{1,u} + \beta \cdot v\left(X_1, u\right) \right) \mathcal{E}\left(\mathrm{d}\varepsilon_{1,u}\right) \middle| X_0 = x \right) \tag{8}$$

reduces the inner integral. The fixed point equation for the s-alternative specific value function (8) simplifies to

$$v(x, s) = \mathbb{E} \left( b \cdot \log \left( \sum_{u \in S} \exp \left( \frac{g(X_1, s;\ u) + \beta \cdot v(X_1, u)}{b} \right) \right) \middle| X_0 = x \right). \tag{9}$$

*Remark* 1. The Gumbel distribution was first incorporated in the model by Rust, while the Generalized Extreme Value Models with conditional logit choice date back to a series of papers by McFadden (cf., for example, McFadden (1973)). A particular advantage of the Gumbel distribution is the simple expression (9) and the explicit formula for the conditional choice probability (Proposition 8 in Appendix A below), exploited in the likelihood function.

*Remark* 2. The scale parameter $b$ can be interpreted as a degree of uncertainty, as the standard deviation of a Gumbel distribution is $b\frac{\pi}{\sqrt{6}} \simeq 1.28\, b$. In particular, the choice $b = 0$ represents decisions without deviations: this degenerate case describes the classical situation in which all managers decide in the same way for a given state.

## 4.3 The estimation problem

Estimates are selected by a maximum likelihood approach (cf. Rust (1987), Su and Judd (2012)), that is, by solving the problem

$$\begin{aligned} \text{maximize} \quad & \mathcal{L}\left(g, v_g, (X_i, s_i, u_i)_{i=1}^N\right) \\ \text{subject to} \quad & v_g = t_g\left(v_g\right), \\ & g \in \mathcal{G}, \end{aligned} \tag{10}$$

where $N$ is the number of observations and $\mathcal{L}$ is the likelihood of observing data $(X_i, s_i, u_i)_{i=1}^N$ conditional on the payoff function $g(\cdot) \in \mathcal{G} = \{g_\theta(\cdot) \colon \theta \in \Theta\}$. The payoff function $g \in \mathcal{G}$ is chosen from a set $\mathcal{G}$ of potential candidate functions. The constraints in equation (10) reflect optimality of the economic model. The operator $t_g$ is for the specified payoff function $g(\cdot)$ and $v_g$ is the expected value function corresponding to the payoff $g(\cdot)$ satisfying the constraint $v_g = t_g(v_g)$.

The constraint

$$v_g = t_g(v_g) \tag{11}$$

in (10) is the fixed-point equation for the expected, or $s$-alternative-specific value function $v(\cdot)$ derived from the Bellman equation of the value function $V(\cdot)$ derived in Section 4.2 above, where $t_g$ is the operator

$$t_g(v)(x, s) := \mathbb{E} \left( b \cdot \log \sum_{u \in S} \exp \frac{g(X_1, s;\ u) + \beta \cdot v(X_1, u)}{b} \middle| X_0 = x \right). \tag{12}$$

$t_g$ is a contraction with Lipschitz constant $\beta < 1$ and Banach's fixed-point theorem ensures that (11) has a unique solution (which we call $v_g$) in the proper space.

Since the payoff shock is not observed by the analyst, the decision chosen for a given state $(x, s)$ is not deterministic but is given by the choice probability

$$P_v(u|\, x, s) = \frac{\exp\left(\frac{g(x,s;u)+\beta\, v(x,u)}{b}\right)}{\sum_{u'\in D} \exp\left(\frac{g(x,s;u')+\beta\, v(x,u')}{b}\right)}, \tag{13}$$

which follows from the fact that the process $\varepsilon$ follows a Gumbel distribution. Equation (13) is discussed, justified and detailed in Proposition 8 in Appendix A. The likelihood function $\mathcal{L}$ is thus

$$\mathcal{L}\left(g, v, (X_i, s_i, u_i)_{i=1}^n\right) = \prod_{i=1}^N P_v(u_i|\, X_i, s_i),$$

which is well-known to be a consistent and efficient estimator.

*Remark* 3. The first approach to solving (10) is the nested fixed point (NFXP) algorithm (cf. Rust (1987)). In the NFXP, for every choice $g \in \mathcal{G}$ the fixed point equation $v_g = t_g\left(v_g\right)$ has to be solved, as the function $v_g$ enters the objective in the maximization (10) (or its approximation (17) below). This is the most expensive part of the computational problem in NFXP. In contrast, in the Su and Judd (2012) approach, the solution $v_g$ maximizing (10) is found in a single optimization formulation.

Our approach described below introduces a direct estimator for $t_g$ which is free of parameters. The method recovers $v_g$ as a by-product of the optimization. Moreover, the approach described ensures convergence to the continuous solution $v_g$. Approximations of the solution are constructed by fixing a grid of supporting points on the positive real line for every $s \in S$ and by linear interpolation of the functions $v(\cdot, s)$, $s \in S$ in between. The supporting points are refined successively to a dense set in $\mathbb{R}_{\geq 0}$ for every $s \in S$, which ensures pointwise convergence of the approximations to $v_g$.

## Estimation of conditional expectation

The probability in the maximum likelihood estimator (10) involves the operator $t_g$ which is an expectation, conditional on $\{X_0 = x\}$. To evaluate $t_g(v)(x, \cdot)$ at a specified point $x$ (cf. (12)) it is necessary to evaluate a conditional expectation.

**The nonparametric approach: employing kernel estimators.** To estimate the conditional expectation of $f(X_{t+1})$ relative to $X_t$, that is $\mathbb{E}(f(X_{t+1}) \mid X_t)$ without an explicit assumption on the underlying process, we pair subsequent observations and consider

$$(X_i, X_{i+1}) \qquad \text{for } i = 1, 2, \dots, N-1. \tag{14}$$

Then the Nadaraya–Watson estimator for the operator

$$t_g(v)(x, s) = \mathbb{E}\left(b \cdot \log \sum_{u\in S} \exp \frac{g(X_{t+1}, s;\, u) + \beta \cdot v(X_{t+1}, u)}{b} \,\middle|\, X_t = x\right)$$

is

$$\hat{t}_g(v)(x, s) := \sum_{i=1}^{N-1} \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{i'=1}^{N-1} K\left(\frac{x-X_{i'}}{h}\right)} \cdot b \cdot \log \sum_{u\in S} \exp \frac{g(X_{i+1}, s;\, u) + \beta \cdot v(X_{i+1}, u)}{b}, \tag{15}$$

where $K(\cdot)$ is an appropriate kernel function and $h > 0$ a suitable bandwidth. Uniform consistency of this estimator is justified in Atuncar et al. (2008).

The estimator $\hat{t}_g$ maintains all properties of the original operator $t_g$, as the following Lemma reveals. We provide a proof of the Lemma in Appendix A.

**Lemma 4.** *For the choice $\beta < 1$ the mapping $v \mapsto \hat{t}_g(v)$ is a contraction on $\ell^\infty([0,\infty) \times S)$ (the linear space of bounded function on $[0,\infty) \times S$), and $v = \hat{t}_g(v)$ has a unique fixed point.*

The estimator $\hat{t}_g$ maintains essential properties on functions which are piecewise linear. This observation is important for numerical treatments as it allows us to consider linear spline functions in implementations. The following corollary is immediate.

**Corollary 5** (Interpolation)**.** *For $d+1$ fixed numbers $x_0 < x_1 < \cdots < x_d$ in $\mathbb{R}$ let $I$ denote the linear interpolation operator, such that*

$$I(v_0, \dots v_d)(x) = \begin{cases} v_0 & \text{if } x \leq x_0, \\ v_j \frac{x_{j+1}-x}{x_{j+1}-x_j} + v_{j+1} \frac{x-x_j}{x_{j+1}-x_j} & \text{if } x_j \leq x \leq x_{j+1}, \\ v_d & \text{if } x \geq x_d. \end{cases}$$

*Then*

$$\hat{\hat{t}}_g\left(\left(v_0^s, \dots v_d^s\right)_{s\in S}\right) := \left( \sum_{i=1}^{N-1} \frac{K\left(\frac{x_j - X_i}{h}\right)}{\sum_{i'=1}^{N-1} K\left(\frac{x_j - X_{i'}}{h}\right)} \cdot b \log \sum_{u \in S} \exp \frac{g(X_{i+1}, s;\, u) + \beta I\left(v_0^u, \dots v_d^u\right)(X_{i+1})}{b} \right)_{j=0}^{d} \tag{16}$$

*is a contraction on $\mathbb{R}^{(d+1)\cdot|S|}$ with a unique fixed point. ($|S|$ is the cardinality of different state modes; $|S| = 3$ in our case, cf. Table 2).*

**The choice of the kernel and bandwidth.** For our set of data and our particular purposes we find the logistic kernel

$$K(x) = \frac{1}{4} \frac{1}{\left(\cosh \frac{x}{2}\right)^2} = \frac{1}{e^x + 2 + e^{-x}}$$

suitable because it allows for all moments and its tails are fat enough to include more distant observations as well.

The choice of this particular logistic kernel is not restrictive. For the bandwidth $h$ we chose Silverman's rule of thumb (cf. Silverman (1998)), that is

$$h_N = \operatorname{std}(X_i) \cdot \left( \frac{4}{(m+2)\,N} \right)^{\frac{1}{m+4}} \approx \operatorname{std}(X_i) \cdot N^{-\frac{1}{m+4}},$$

where $\operatorname{std}(\cdot)$ is the standard deviation of the sample, $m = 2$ is the dimension of each individual pair of the samples (cf. (14)) and $N$ the sample size.

**Augmented likelihood function.** Following the literature (cf., for example, King and Zeng (2001)) it is natural to adjust the likelihood to reflect the different sample sizes of the groups. This is accomplished by the *augmented likelihood*

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_{i=1}^{N} \frac{1}{N_i} \log P_{v_g}(u_i | X_i, s_i) \\ \text{subject to} \quad & v_g = t_g(v_g), \\ & g \in \mathcal{G}. \end{aligned} \tag{17}$$

Here,

$$N_i \in \{6539;\ 76;\ 184;\ 1312;\ 78\}$$

is the sample size of the group to which the observation $(X_i, s_i, u_i)$ belongs to as in Figure 1 (instead of $N = 8189$ in (10)) ($i \in \{OP \to OP, OP \to SB, SB \to OP, SB \to SB, SB \to RE\}$).

**Observation** An observation in our estimation exercise is a triple $(X_i, s_i, u_i)$ consisting of the following components:

- the profitability $X_i$ for the current year,
- the operating state of the power plant $s_i \in S$ in the current year, and,
- the decision of the manager regarding the operating state $u_i \in S$ of the power plant in the upcoming year.

Note that this structure applies to switching problems more generally.

The nonparametric estimation exercise relies on pairs of observations $(X_t, X_{t+1})$, the profitability in the current year $t$ and in the coming year $t+1$. At the time of the decision, the profitability for the coming year is not yet known. It is reasonable to assume that decision makers have estimates of profitability for the coming year. In practice the plant manager is likely to rely upon production costing software (e.g., PROSYM, UPLAN, EGEAS) to simulate the operation of the regional electric system and therefrom derive an estimate of profitability for the upcoming year $t+1$.

Figure 2 presents the evolution of profitability from one year to the next. The density in Figure 2 is estimated based on the pairs $(X_t, X_{t+1})$ (cf. (14)), which are available from the observations: one ordinate represents the profitability indicator of this year, $X_t$, the other ordinate the profitability indicator in the subsequent year, $X_{t+1}$. The density thus describes the Markov kernel, which is used in the expectation to compute the value function, for example in (12).

*Remark* 6. Figure 2 indicates that there is no common pattern of transitions of the profitability from a year to the next. If the transitions were described by a geometric Brownian motion (GBM), then a slice of the density plot would be lognormal.

## 4.4 Unobserved heterogeneity

Our model captures the difference between the (within-model) expected immediate payoff and the real payoff observed by the decision maker, i.e., the payoff shock, by Gumbel variables with a common scale parameter $b$. Corollary 9 in Appendix A (cf. particularly (25)) states that the *difference of Gumbel variables* follows a logistic distribution.

To estimate unobserved heterogeneity we employ a logistic regression for binary classification with parameters $\theta = (\alpha, \eta)$ and the probabilities

$$P(OP \mid X) = \frac{1}{1 + \exp\left(-\frac{X-\alpha}{\eta}\right)} \quad \text{and} \quad P(SB \mid X) = \frac{1}{1 + \exp\left(\frac{X-\alpha}{\eta}\right)} \tag{18}$$

(note, that $P(OP \mid X) + P(SB \mid X) = 1$). The purpose of this regression is to identify the managerial decisions $\{OP, SB\}$ given the status of the variable $X$, that is the profitability indicator.

The parameters $\alpha$ and $\eta$ in (18) allow a natural interpretation. The location parameter $\alpha$, the median of the logistic distribution, divides the managers' decisions into the two groups *operating*, that is $\{X > \alpha\}$ and *standby* ($\{X < \alpha\}$, respectively). The scale parameter $\eta$ describes the standard deviation of these decisions:

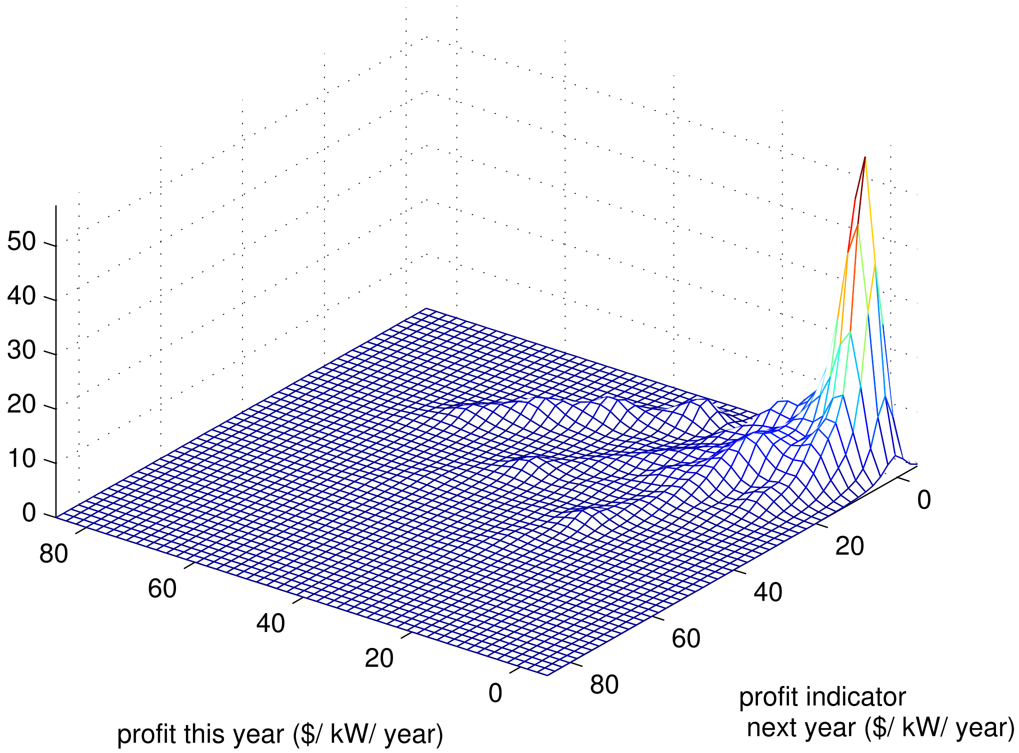


Figure 2: Bivariate density of the observed transition $(X_t,\ X_{t+1})$ of the annual profitability indicator.

that is the uncertainty within these decisions, or the respective heterogeneity of all managerial decisions. A comparison with (25) in Appendix A identifies the parameters $b = \eta$. Numerical computations of the parameter $\eta$ give evidence that $b \approx \$5.3$ /kW/year for our sample data. This parameter typically is set to unity in the structural estimation literature. To the best of our knowledge we are the first to estimate this parameter.

The parameter $b$ captures the degree to which managers make different decisions even when the observable (to the researcher) data are the same or similar. The magnitude of $b$ determines the importance of unobserved shocks in the decision making process. In our data, this effect is relatively large as the estimate of $b \approx \$5.3$ /kW/year is approximately 42% of the mean value of the profitability state variable (\$12.5 /kW/year from Table 1).

## 4.5 The payoff function $g(\cdot)$

The payoff function $g(\cdot)$ describes the expected cash flow for the year. We include not only the profitability indicator $X$ but also two other costs, (i) the costs of continuing maintenance $M_u$ given that the generator is in state $u$, and, (ii) the switching costs associated with the transitions themselves, $K_{s\to u}$.

In addition to the (limited) heterogeneity induced by the payoff shock we capture persistent unobservable heterogeneity by allowing some of the payoff function parameters to be random variables (cf. Train (2002)). Specifically, we let the maintenance cost in the standby state ($M_{SB}$) and the start up cost ($K_{SB\to OP}$) be (discretized versions of) Gaussian variables.[10] We estimate the means ($\mu_{SB}$, $\mu_{SB\to OP}$) and the standard deviations ($\sigma_{SB}$, $\sigma_{SB\to OP}$,) of these random variables. If start up costs are similar across plants, the estimated standard deviation should be low.

The payoff function is given by

$$g_{j,\theta}(X,s;\,u)=\begin{cases} X - M_{OP} & s=\text{OP},\ u=\text{OP},\\ \frac{1}{2}\big(X - M_{OP} - (\mu_{SB} + y_j\cdot\sigma_{SB})\big) - K_{OP\to SB} & s=\text{OP},\ u=\text{SB},\\ \frac{1}{2}\big(X - (\mu_{SB} + y_j\cdot\sigma_{SB}) - M_{OP}\big) - (\mu_{SB\to OP} - y_j\cdot\sigma_{SB\to OP}) & s=\text{SB},\ u=\text{OP},\\ -\mu_{SB} - y_j\cdot\sigma_{SB} & s=\text{SB},\ u=\text{SB},\\ -\frac{1}{2}(\mu_{SB} + y_j\cdot\sigma_{SB}) - K_{SB\to RE} & s=\text{SB},\ u=\text{RE},\\ -\infty & \text{otherwise}, \end{cases} \tag{19}$$

where the parameter $\theta$ carries the parameters to be estimated, i.e.,

$$\theta = (M_{OP}, \mu_{SB}, \sigma_{SB}, K_{OP\to SB}, \mu_{SB\to OP}, \sigma_{SB\to OP}, K_{SB\to RE}),$$

and $y_j$ are discretization points of a Gaussian random variable reflecting persistent unobservable types among the plant owners. The discretization chosen employs the representative points (quantizers) of the distribution with the corresponding optimal weights, such that the distance to the genuine distribution is minimized (cf. Graf and Luschgy (2000, Section 5) who use the Wasserstein distance).

The value $-\infty$ is included in the payoff function to exclude other transitions. Notice that for plants which are either shut down or started up, we include only half of the profit, as well as half of the maintenance cost for both the operational ($OP$) and shutdown ($SB$) states, $\frac{1}{2}\big(X - M_{OP} - (\mu_{SB} + y_j\cdot\sigma_{SB})\big)$. Because our data are observed at annual frequency, we do not know when status changes occur. We assume that shut downs and start ups happen mid-year so that in both cases the plant is assumed operational for half of the year. Similarly we include half of the maintenance cost for the $SB$ state. The parameters of interest are the switching costs $K_{OP\to SB}$, $\mu_{SB\to OP}$ and $K_{SB\to RE}$.

[10] Computations were too burdensome when increasing the number of parameters affected by such heterogeneity from two to three. We tried most combinations of parameters and settled upon cross-plant heterogeneity in $M_{SB}$ and $K_{SB\to OP}$.

A generator which has been retired has no value beyond any potential salvage value as described above, that is

$$v(\cdot, \text{retired}) = 0.$$

What remains to be computed is

$$v(\cdot, s), \qquad \text{for } s \in \{\text{operating, standby}\}.$$

As justified in Corollary 5, it is possible to employ linear interpolation by fixing supporting points $x_0 < x_1 < \ldots x_d$. The problem is

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{N} \frac{1}{N_i} \log \sum_{j=1}^{n} p_j P_{I(v_{g_j})}(u_i | X_i, s_i) \\ \text{subject to} \quad & v_{g_j} = \hat{\hat{t}}_{g_j}\left(v_{g_j}\right), \\ & g_j \in \mathcal{G}, \end{aligned} \tag{20}$$

where $g_j \in \mathcal{G}$, in view of (19), means that $\theta$ contains the variables in the optimization procedure (20). We impose the constraints $M_{OP} \geq 0$, $K_{OP\to SB} \geq 0$, $\mu_{SB\to OP} \geq 0$ and $\mu_{SB} \geq 0$ on our optimization procedure, and this is reflected in the functions $g \in \mathcal{G}$ as well.

Irrespective of the supporting points $x_0 < \cdots < x_d$ this problem is always feasible for every choice of $\theta$. By augmenting the sequence $x_0 < \cdots < x_d$ by additional points a net is obtained, which converges finally to the value function $v$, the solution of (10).

# 5 Results

Table 3 presents our results. We begin by discussing the estimated maintenance costs. Unlike switching costs, estimates of ongoing maintenance costs are available from the EIA. In 2010 the EIA commissioned a consultant to develop new estimates, see Energy Information Administration (2010).[11] According to that document, the annual fixed O&M costs for combustion turbines range from \$6.70 /kW/year to \$6.98 /kW/year.

The first column of results Table 3 presents our estimated maintenance costs ($M_{OP}$) for a plant which is in the operating state. The estimates, approximately \$8 /kW/year to \$10 /kW/year, are greater than EIA estimates. The EIA numbers are for *new* plants. Our estimates are based on existing plants which vary in age and size. It is reasonable that the annual costs of maintenance for an old combustion turbine should be greater than for new technology.

The second and third columns of Table 3 present estimated mean ($\mu_{SB}$) and standard deviation ($\sigma_{SB}$) of maintenance costs in the shutdown state ($M_{SB}$). The estimated maintenance costs in the shutdown state are approximately 20–40 % of the maintenance costs in the operating state. This reduction in maintenance costs relative to those in the operational state provides incentive for plant owners to shut down peaking plants.

The fourth column contains estimates of shut down costs ($K_{OP\to SB}$). In every case these costs are estimated to be zero. Near zero shut down costs are consistent with the reality that the owner of a peaking plant can essentially turn the machine off and ignore it.

The fifth and sixth column present estimated mean ($\mu_{SB\to OP}$) standard deviation ($\sigma_{SB\to OP}$) of start up costs ($K_{SB\to OP}$). These numbers are relatively small in magnitude. During the optimization exercise

[11]These estimates are in as much as possible based on real world experience. From Energy Information Administration (2010, p. 2):

> Where possible, costs estimates were based on information regarding actual or planned projects available to the consultant. When this information was not available, project costs were estimated by using costing models that account for current labor and material rates that would be necessary to complete the construction of a generic facility.

| | $M_{OP}$ | $\mu_{SB}$ | $\sigma_{SB}$ | $K_{OP\to SB}$ | $\mu_{SB\to OP}$ | $\sigma_{SB\to OP}$ | $K_{SB\to RE}$ |
|---|---|---|---|---|---|---|---|
| $\beta = 0.91$ | 8.50 | 2.45 | 0.16 | 0.00 | 0.79 | 0.46 | −31.3 |
| | (1.22) | (1.03) | (0.18) | (0.00) | (1.32) | (0.77) | (11.0) |
| $\beta = 0.95$ | 9.32 | 3.23 | 0.05 | 0.00 | 0.56 | 0.32 | −49.0 |
| | (1.28) | (1.06) | (0.10) | (0.00) | (1.36) | (0.79) | (22.5) |
| $\beta = 0.97$ | 10.0 | 3.87 | 0.02 | 0.00 | 0.46 | 0.27 | −62.7 |
| | (1.15) | (1.13) | (0.08) | (0.00) | (1.41) | (0.82) | (39.5) |

Table 3: Maintenance and switching cost estimates for peaking power plants, in units of \$/kW/year. $M_{OP}$ is the ongoing cost of maintenance for an operating plant. $M_{SB}$ is the ongoing cost of maintenance for a plant which has been shut down. $K_{OP\to SB}$ is the one-time cost of shut down. $K_{SB\to OP}$ is the one-time cost of start up. $K_{SB\to RE}$ is the one-time cost of abandonment. The numbers in parentheses are standard deviations of the estimates, calculated by (parametric) bootstrapping. Results are shown for three different discount factors $\beta = 0.91$, $\beta = 0.95$, and $\beta = 0.97$.

we found that the algorithm essentially traded off (i) maintenance costs in the shutdown state, and, (ii) start up costs. This trade off is also consistent with the real world. The owner of a peaking plant may chose to spend very little on maintenance for a plant which is shut down. In that case, the cost of starting up will be greater. However, if the plant owner invests in maintenance when the plant is shut down, starting up the plant is less costly.

Finally, the switching costs associated with an abandonment ($K_{SB\to RE}$) are negative. This result is consistent with the existence of a secondary market for used CTs. Our estimated salvage values range from approximately 3.2 % to 7.4 % of the cost of a brand new combustion turbine (Energy Information Administration (2010)).[12]

## 5.1 Avoidable Cost Rates

From our estimates of switching and maintenance costs we can infer avoidable cost rates (ACR). In our setting, avoidable costs are equal to the maintenance cost in the operating state less three things: (i) the cost to shut down, (ii) the maintenance cost in the shutdown state, and (iii) the cost to start up one year later.

$$ACR = M_{OP} - K_{OP\to SB} - \mu_{SB} - \mu_{SB\to OP}. \tag{21}$$

We find $ACR$ = \$14.41/MW-day, \$15.15/MW-day, and \$16.55/MW-day for $\beta = 0.91$, 0.95 and 0.97, respectively. These estimates are less than RPM default rates which range from \$17/MW-day to \$30/MW-day and even higher in more recent years. We interpret this result to mean that the actual cash flow differences associated with shutting down peaking plants are less than the defaults values used in PJM.

If default ACR values are too high, then RPM prices may be higher than is necessary to ensure system reliability. The American Public Power Association and the Electric Power Research Institute published a report in 2010 (American Public Power Association (2010)) which referred to RPM as "... *a market that is costing consumers more than needed to ensure reliability ...*" Our results are consistent with this conclusion.

The procedure for estimating avoidable costs, available from PJM, calls for a bottom-up approach in which labor, materials, administrative and general costs, etc., are estimated separately and then summed. While the bottom-up method is useful for attributing common costs amongst individual plants, it likely does

[12]Common practice in the literature is to model the exogenous state variable using a parametric approach. For comparison we develop estimates based on the autoregressive model described in Appendix B. In general the estimates using the autoregressive setup are similar to our nonparametric approach, though the standard deviations tend to be smaller. We prefer the nonparametric approach as we believe that an autoregressive model is too simple to capture the data generation process.

not capture the firm's incremental cash flow caused by shutting down a particular plant. In contrast our estimates of avoidable costs are based upon firm's actual decisions, along with market prices of electricity and fuel.

## 6 Conclusion and policy implications

We use structural estimation to obtain estimates of switching costs for peaking power plants, which are difficult to obtain in practice. These estimates are implied from a time series of observed switches across many plants and from assuming rational behavior on the part of the plant managers. That is, the maintenance and switching costs are obtained by maximising the likelihood of observing the data, subject to the assumption that the switches follow a real options model.

We develop nonparametric techniques for structural estimation, as the process observed does not follow a known parametric process. As additional contribution to structural estimation we present an estimator for the scaling parameter of the unobserved payoff shock process.

Our cost estimates have policy implications for electricity market regulators, who should note that a real options lense is a useful way of viewing the availability of peaking power plants. For the PJM market in particular, our estimates imply avoided costs which can be as low as half the default values used in PJM's RPM capacity market, consistent with the interpretation that consumers are paying more than necessary for system reliability.

## A Extreme value distributions – the Gumbel variable

It is shown that the Gumbel distribution is closed under maximization (indeed, this is the essential property of any extreme value distribution). Further, a closed form formula for the probability of choice is provided for the Gumbel distribution. A very comprehensive discussion of extreme value distributions can be found in Embrechts et al. (1997).

The cumulative distribution function (cdf) of a Gumbel distribution is $F(z) = \exp\left(-e^{-\frac{z-\mu}{b}-\gamma}\right)$, where $\gamma = 0.577\,215\,66\ldots$ is the Euler–Mascheroni constant. Its mean is $\mu$, and the variance is $b^2\frac{\pi^2}{6}$.

**Proposition 7** (The extreme value distribution is closed under maximization)**.** *Let* $(\varepsilon_i)_{i=1}^n$ *be independent random variables which are Gumbel distributed with mean* $\mu_i$ *and common scale parameter* $b > 0$. *Then the maximum* $\varepsilon := \max\{\varepsilon_i + c_i : i = 1, \ldots n\}$ *of the shifted variables is again Gumbel distributed with mean*

$$\mathbb{E}(\varepsilon) = \mu := b \cdot \log\left(\sum_{i=1}^n \exp\left(\frac{\mu_i + c_i}{b}\right)\right)$$

*and the same scale parameter* $b$*, where* $c_i \in \mathbb{R}$ *are arbitrary constants.*

*Proof of Proposition 7.* From the cumulative distribution function of the Gumbel distributions with respective means it follows that

$$\begin{aligned}
P\left(\max_{i\in\{1,\dots n\}} \varepsilon_i + c_i \le z\right) &= P\Big(\varepsilon_1 + c_1 \le z,\, \varepsilon_2 + c_2 \le z, \dots, \varepsilon_n + c_n \le z\Big)\\
&= \prod_{i=1}^n P\left(\varepsilon_i \le z - c_i\right) = \prod_{i=1}^n \exp\left(-e^{-\frac{z-c_i-\mu_i}{b}-\gamma}\right)\\
&= \exp\left(-\sum_{i=1}^n e^{-\frac{z-c_i-\mu_i}{b}-\gamma}\right) = \exp\left(-e^{-\frac{z}{b}-\gamma}\cdot \sum_{i=1}^n e^{\frac{\mu_i+c_i}{b}}\right)\\
&= \exp\left(-e^{-\frac{z}{b}-\gamma}\cdot e^{\frac{\mu}{b}}\right) = \exp\left(-e^{-\frac{z-\mu}{b}-\gamma}\right),
\end{aligned}$$

because $\sum_{i=1}^n e^{\frac{\mu_i+c_i}{b}} = e^{\frac{\mu}{b}}$. This reveals the assertion. □

The following proposition addresses the probability of choice. Again, an explicit formula is available for shifted Gumbel variables.

**Proposition 8** (Choice probabilities for shifted Gumbel variables)**.** *Let* $(\varepsilon_i)_{i=1}^n$ *be independent Gumbel distributed random variables with individual mean* $\mu_i$ *and common scale parameter* $b > 0$*. Then the probability of choice for the variables shifted by* $c_i$ *is*

$$P\left(\varepsilon_1 + c_1 = \max_{i\in\{1,2,\dots n\}} \varepsilon_i + c_i\right) = \frac{\exp\left(\frac{c_1+\mu_1}{b}\right)}{\exp\left(\frac{c_1+\mu_1}{b}\right) + \cdots + \exp\left(\frac{c_n+\mu_n}{b}\right)}. \tag{22}$$

*Proof of Proposition 8.* Without loss of generality one may consider a pair $(\varepsilon_1, \varepsilon_2)$ of independent Gumbel variables with location parameter 0, because the maximum in (22) itself is Gumbel distributed by Proposition 7.

Thus

$$\begin{aligned}
P\left(\varepsilon_1 + c_1 \ge \varepsilon_2 + c_2\right) &= P\left(\varepsilon_2 \le \varepsilon_1 + c_1 - c_2\right)\\
&= \int_{-\infty}^{\infty} f\left(x_1\right) \int_{-\infty}^{x_1+c_1-c_2} f\left(x_2\right) \mathrm{d}x_2 \mathrm{d}x_1\\
&= \int_{-\infty}^{\infty} f\left(x_1\right) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) \mathrm{d}x_1,
\end{aligned} \tag{23}$$

where the cdf of the Gumbel distribution has been substituted. By substituting the probability density function (pdf) $f$, (23) continues as

$$\begin{aligned}
P\left(\varepsilon_1 + c_1 \ge \varepsilon_2 + c_2\right) &= \int_{-\infty}^{\infty} \frac{1}{b} \exp\left(-\frac{x_1}{b} - e^{-\frac{x_1}{b}}\right) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) \mathrm{d}x_1\\
&= \int_{-\infty}^{\infty} \frac{1}{b} e^{-\frac{x_1}{b}} \exp\left(-e^{-\frac{x_1}{b}}\left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right) \mathrm{d}x_1\\
&= \left[\frac{\exp\left(-e^{-\frac{x_1}{b}}\left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right)}{1 + e^{-\frac{c_1-c_2}{b}}}\right]_{x_1=-\infty}^{\infty}\\
&= \frac{1}{1 + e^{-\frac{c_1-c_2}{b}}} = \frac{e^{\frac{c_1}{b}}}{e^{\frac{c_1}{b}} + e^{\frac{c_2}{b}}}.
\end{aligned} \tag{24}$$

This completes the proof. □

Finally we give provide a proof that the difference of Gumbel variables enjoys a logistic distribution (cf. Nadarajah (2007)).

**Corollary 9.** *If $\varepsilon_1$ and $\varepsilon_2$ are Gumbel distributed with mean $\mu_1$ and $\mu_2$ and common scale parameter $b > 0$. Then the difference $\delta := \varepsilon_2 - \varepsilon_1$ follows a logistic distribution with mean $\mu = \mu_2 - \mu_1$ and cumulative distribution function*

$$F_\delta(z) = \frac{1}{1 + \exp\left(-\frac{z-\mu}{b}\right)}, \tag{25}$$

*which is the distribution function of a logistic variable.*

*Proof of Corollary 9.* If follows from (24) in the proof of the preceding theorem that

$$F_\varepsilon(z) = P(\varepsilon_2 - \varepsilon_1 \le z) = P(\varepsilon_1 + z \ge \varepsilon_2) = \frac{1}{1 + e^{-\frac{z-(\mu_2-\mu_1)}{b}}},$$

which completes the proof. □

We include a proof of Lemma 4.

*Proof.* Observe first that the sample $(X_i)_{i=1}^N$ and $S$ are finite, such that $g(X_{i+1}, s; u)$ is uniformly bounded. Moreover the mapping

$$v \mapsto b \cdot \log \sum_{u \in S} \exp \frac{g_u + \beta \cdot v}{b} = b \cdot \log \left( \exp\left(\frac{\beta \cdot v}{b}\right) \cdot \sum_{u \in S} \exp\left(\frac{g_u}{b}\right) \right) = \beta \cdot v + b \cdot \log \sum_{u \in S} \exp\left(\frac{g_u}{b}\right)$$

is an affine linear function in $v$ with slope $\beta < 1$. Due to the construction of the operator $\hat{t}_g$ in (15) with nonnegative weights summing to 1 it follows that $\hat{t}_g$ is a contraction with Lipschitz constant $\beta < 1$. Hence Banach's fixed-point theorem applies and guarantees a unique fixed-point $v_g$ in $\ell^\infty([0,\infty) \times S)$, that is, $v_g = \hat{t}_g(v_g)$. □

# B The parametric approach: employing a time series model.

For the purposes of comparison we estimate switching and maintenance costs using the parametric approach. Assume that the time series of the state variable $X$ follows an autoregressive scheme (AR)

$$X'_{t+1} = \mu + \rho X'_t + \sigma \mathcal{N}_t \tag{26}$$

with error term $\mathcal{N}_t$, where $\mathcal{N}_t$ are independent random variables. In this situation the conditional expectation is given by the explicit formula

$$\mathbb{E}\left(f(X'_{t+1}) | X'_t = x\right) = \mathbb{E} f\left(\mu + \rho x + \sigma \mathcal{N}_t\right), \tag{27}$$

reducing thus the conditional expectation to a simple expectation.

We estimate the parameters $\mu$, $\rho$ and $\sigma$ of the underlying time process in (26) for normally distributed random variables $\mathcal{N}$. We then apply the expression (27) to the censored process $X_t := \max(0, X'_t)$ to evaluate expressions as (12) explicitly, conditionally on $\{X_t = x\}$. The censored process $X_t$ reflects the fact that the process we observe is zero with positive probability, but never negative.

The parameters for the time series (26) are estimated by a maximum likelihood approach, such that

$$X_{t+1} = 2.3 + 0.45 \cdot X_t + 9.2 \cdot \mathcal{N}_t,$$

where $\mathcal{N}_t$ are independent standard normal variables.

## Acknowledgments

We recognize the Norwegian research centre CenSES, Centre for Sustainable Energy Studies, and acknowledge support from the Research Council of Norway (RCN) through project 228811. A. Pichler acknowledges RCN project 207690, and thanks Chemnitz University of Technology, Faculty of Mathematics, for hosting during the latter part of this research.

## References

American Public Power Association. A Review of PJM's Reliability Pricing Model. September 2010. 15

G. S. Atuncar, C. C. Y. Dorea, and C. R. Gonçalves. Strong consistency of Kernel density estimates for Markov chains failure rates. *Statistical Inference for Stochastic Processes*, 11(1):1–10, 2008. ISSN 1387-0874. doi:10.1007/s11203-006-9003-7. 10

R. Bansal, A. R. Gallant, R. Hussey, and G. Tauchen. Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics*, 66(1):251–287, 1995. doi:10.1016/0304-4076(94)01618-A. 2

C. Bontemps, J.-M. Robin, and G. J. Van den Berg. Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review*, 41(2):305–358, 2000. doi:10.1111/1468-2354.00066. 2

M. Brennan and E. Schwartz. Evaluating natural resource investments. *The Journal of Business*, 58(2): 135–157, 1985. 3

B. Caldecott and J. McDaniels. Stranded generation assets: Implications for European capacity mechanisms, energy markets and climate policy. Working paper, University of Oxford, 2014. 3

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*, volume 33 of *Stochastic Modelling and Applied Probability*. Springer, 1997. URL `http://books.google.com/books?id=BXOI2pICfJUC`. 7, 16

Energy Information Administration. Assumptions to the Annual Energy Outlook 2009. (DOE/EIA-0554(2009)), March 2009. 4

Energy Information Administration. Updated capital cost estimates for electricity generation plants. November 2010. 14, 15

S.-E. Fleten and E. Näsäkkälä. Gas-fired power plants: Investment timing, operating flexibility and CO2 capture. *Energy Economics*, 32(4):805–816, 2010. 3

S.-E. Fleten, E. Haugom, and C. J. Ullrich. The real options to shutdown, startup, and abandon: U.S. electricity industry evidence. *Energy Economics*, 63:1–12, 2017. 2, 3, 4

A. Gamba and M. Tesser. Structural estimation of real options models. *Journal of Economic Dynamics & Control*, 33:798–816, 2009. 1

S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg, 2000. doi:10.1007/BFb0103945. 13

E. Guerre, I. Perrigne, and Q. Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000. doi:10.1111/1468-0262.00123. 2

G. King and L. Zeng. Logistic regression in rare events data. *Political Analysis*, 9(2):137–163, 2001. 10

T. Li, I. Perrigne, and Q. Vuong. Structural estimation of the affiliated private value auction model. *RAND Journal of Economics*, pages 171–193, 2002. URL http://www.jstor.org/stable/3087429. 2

D. McFadden. *Conditional logit analysis and qualitative choice behavior*, chapter 4, pages 105–142. 1973. 8

A. Moel and P. Tufano. When are real options exercised? an empirical study of mine closings. *Review of Financial Studies*, 15(1):35–64, 2002. 3

A. Musalem, M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010. doi:10.1287/mnsc.1100.1170. 2

S. Nadarajah. Linear combination of gumbel random variables. *Stochastic Environmental Research and Risk Assessment*, 21(3):283–286, 2007. ISSN 1436-3240. doi:10.1007/s00477-006-0063-4. 18

W. K. Newey, J. L. Powell, and F. Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603, 1999. doi:10.1111/1468-0262.00037. 2

PJM Interconnection LLC. PJM's evolving resource mix and system reliability. 2017. 3

J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033, Sept. 1987. doi:10.2307/1911259. 1, 7, 8, 9

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRC, 1998. 10

R. Smith and L. Cook. Hawaii wrestles with vagaries of a green grid. *The Wall Street Journal*, June 29, 2015:B1, 2015. 3

C.-L. Su and K. L. Judd. Constrained optimization approaches to estimation of structural models. *Econometrica*, 80:2213–2230, 2012. 1, 2, 8, 9

K. Train. *Discrete choice methods with simulation*. Cambridge University Press, 2002. 13

C. J. Ullrich. Realized volatility and price spikes in electricity markets: The importance of observation frequency. *Energy Economics*, 34(6):1809–1818, 2012. 2