Nonparametric Structural Estimation of Switching Options in Power Plants

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

July 24, 2014

Abstract

We develop a method to estimate irreversible switching costs associated with economic state changes. Necessary input include observations of state changes over several facilities and over time, and a time series of a profit indicator. The dynamic of the latter exogenous state variable is modelled nonparametrically. This extends recent contributions in structural estimation, combining nonparametric statistics with nonlinear programming. We implement the method on a unique dataset of U.S. power plants. In the wake of increased penetration of electricity generation from renewable sources into energy systems, regulators are particularly concerned about temporary and permanent shutdowns of conventional plants. The results indicate that our method is able to arrive at economically meaningful estimates of maintenance cost and switching costs.

Keywords: Structural Estimation, Hamilton–Jacobi–Bellman equations, Value Function, Stochastic Optimization, Switching Options
JEL classification: C14, C61, D92, G13, G31, Q40

1 Introduction

Structural estimation of dynamic discrete choice models was introduced in the paper [25] by Rust, in which he studies replacing or repairing bus engines in fleet vehicles. The fleet operator bases his decisions on detailed inspections and on experience, while only a fraction of this comprehensive information is available to the observing economist.

An economist who observes several decision makers is confronted with a similar situation of incomplete information. Every decision maker makes individual, presumably optimal decisions based on information which is at least partially hidden from the observing economist. From the economist's perspective the decision makers act heterogeneously.

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

†Lillehammer University College, Lillehammer, Norway, Norwegian University of Science and Technology, NO-7491, Trondheim, Norway, Contact: erik.haugom@hil.no.

‡Norwegian University of Science and Technology, Actuary of the Actuarial Association of Austria, Contact: aloisp@ntnu.no.

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

By employing structural estimation, which combines a dynamic decision model with empirical data, it is nonetheless possible to estimate economic primitives such as cost parameters in the assumed decision model. Structural estimation for dynamic discrete choice builds on dynamic programming. It considers a function (the value function), which is often the expected value of all discounted, weighted future realizations of a profit measure (the net present value). An essential assumption is that the system evolves according an underlying Markovian process. In an economic environment, however, only a fraction of this underlying process is available to the observer (cf. Rust's initial example).

An optimal control policy is characterized by Bellman equations, which derive from the dynamic programming principle. In the framework of structural estimation the unobserved process is assumed to be conditionally independent of the observed process, and assumed to follow a very special distribution (an extreme value distribution). By accepting this assumption the equations simplify significantly, as a closed form formula for the associated expectation is available.

The papers by Rust [25], Gamba and Tesser [13] and Su and Judd [28] consider parametric stochastic processes (exponential distributions, or geometric Brownian motions) to model the underlying, exogenous state variable. Our paper generalizes this approach and considers nonparametric, general transitions. The transition operator in the case study is estimated from observations by employing kernel functions, which are well-known from nonparametric density estimation and nonparametric regression.

A further improvement of this paper concerns the handling of unobserved heterogeneity, which is described by a parameter. This parameter is typically given externally or set to one. Here, we estimate the heterogeneity of different decision makers from the data available and thus get a more accurate statistical model. It turns out that the parameter describing heterogeneity smoothes the results of Bellman's equations. We outline further that structural estimation reduces to the classical Bellman theory, if heterogeneity disappears, i.e., if all decision makers act homogeneously. Structural estimation thus is a general framework comprising the Hamilton–Jacobi–Bellman theory.

We focus on dynamic decision processes in which the decision maker can make costly mode switches (economic state changes) under uncertainty. Brekke and Øksendal [5] offer an analysis of the optimal solution of such switching problems quite generally, noting that it belongs to the class of generalized impulse control problems studied, e.g., by Bensoussan and Lions [3]. Brennan and Schwartz [6], Dixit [9] and Triantis and Hodder [29] note that the combination of uncertainty, irreversible switching costs, and discretion over timing of the switches has widespread implications in dynamic economics. In contrast to these theoretical contributions, we develop a method for the empirical analysis of such problems in which the economist observes several decision makers over time, the relevant economic states and the actual switches, but not the switching costs. The method can also derive other primitives associated with the decision problem such as preference parameters or cost determinants.

Previous work that uses non-parametric methods in structural estimation of dynamic models include Bansal et al. [2], who nevertheless specify an autoregressive structure for the exogenous state variables. Newey et al. [22], Bontemps et al. [4], Guerre et al. [14] and Li et al. [17] employ non-parametric estimation for structural estimation of static models within simultaneous equations, auctions, and labor search.

Case study. We apply the new method to the estimation of switching costs associated with mothballing (which we refer to as shutting down), restarting, and/ or abandoning existing power plants.[1] We focus our analysis on peak plants, i.e., those plants whose function it is to generate electricity during times of peak demand. Peak plants are crucial for the integrity of the electricity grid. Given the advent of renewable resources with zero or near-zero variable costs, the profitability of peak plants is likely to be greatly reduced. This is the situation in Europe today, and a similar problem is likely to appear in U.S. markets as well. It is therefore important to understand thoroughly the drivers of shutdown, start up, and abandonment decisions for peak plants.

In addition to advancing a new estimation method, we contribute by providing reasonable estimates of the costs involved in shutdown, start up, and abandonment. We believe this is a significant contribution in its own right.

Our case study is made possible by the availability of detailed data. Our sample includes 8,189 observations for peak plants located in the U.S. Our switching cost estimates are reasonable when compared to initial and maintenance costs for new capacity. These switching cost estimates themselves should be valuable to utility system planners, regulators, and other parties concerned with electric system planning.

Outline of the paper. Section 2 describes structural estimation and provides motivation. Section 3 addresses the new non-parametric approach. Section 4 describes the case study and details the data. Results are presented in Section 5. In Section 6 we discuss the relevance of the results and conclude.

2 Theoretical Framework

2.1 The framework for structural estimation

In this section we define the nomenclature and introduce the structural estimation method. The notation, as well as the outline closely follow the literature, as for example in Gamba and Tesser [13].

Nomenclature

k — Time index; the unit time period is a year.

X_k — The state process; in our specific case the state process is an indicator of profitability per unit of capacity expressed in units of dollars per kilowatt, $\$/kW$.

(X_k, ε_k) — The augmented state process; the second process, ε_k, is not accessible to observation.

$s_k,\ u_k \in S := \{\text{operating, standby, retired}\}$ are operating states of the power plants and decided by the plant manager.

[1]Structural estimation applications to electricity related problems include Rust and Rothwell [26] and Rothwell and Rust [24], who study the effect of a regulatory shift and plant optimal lifetime, respectively, for nuclear power plants. In energy economics more broadly, recent work on structural estimation include analysis of timing; Rapson [23] study the timing of appliance investment, Kellogg [15] well drilling, Muehlenbachs [20] well decommissioning, Burr [7] solar PV investment and Lin and Thome [18] corn-ethanol plant investment.

(X_i, s_i, u_i) An observation consists of a profitability X_i during the time period, 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.

$g(x, s;\, u)$ The payoff during a single period. The payoff function $g(\cdot)$ comprises the expected cash flow for the next period and all 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 value function, 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 $\beta = {}^1\!/_{(1+\text{interest rate})}$.

2.2 Bellman equation of an individual decision maker

The state process is observed over a sequence of years, $X_0, X_1, \ldots$. A decision s_k is allowed at every stage k ($k = 0,\, 1, \ldots$). Assume the development of the process depends only on its history, that is the process is nonanticipative and the decision is based on the history of the system. Employing a payoff function $g(\cdot)$ which describes the profitability in each period, the investor will thus maximize the function

$$V(x, s) := \max_{s_k \lhd \sigma(X_k)} \mathbb{E}\left(\sum_{k=0}^{\infty} \beta^k g\left(X_k, s_k;\, s_{k+1}\right) \middle|\, X_0 = x\right);$$

the maximum here is among all decision processes $(s_0, s_1, \ldots)$, with $s_0 = s$. It is natural to assume that the process X is Markovian and independent of time. That is $s_k = s_k(X_k)$ for some measurable function s_k (cf. Shiryaev [27, Theorem II.4.3]), which is made explicit by writing $s_k \lhd \sigma\,(X_k)$. It follows that

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

which is the usual Bellman equation.[2] Note that X_{k+1} is distributed, conditional on X_k, in the same way as X_1 is distributed conditional on X_0. For this reason we shall express the conditional expectations based on X_0 in what follows.

[2]Cf. Fleming and Soner [11, Chapter IX].

It follows from (1) that the value function V is a fixed point for

$$T(V)\,(x,s) := \max_{u\in S}\, g\,(x,s;\,u) + \beta\cdot \mathbb{E}\left(V\,(X_1,u)\right|X_0=x)\,, \tag{2}$$

that is,

$$V = T(V). \tag{3}$$

The operator T is a contraction by Blackwell's sufficient conditions (for every individual $s\in S$) on the Banach space $\ell^\infty\left([0,\infty)\times S\right)$ or $C([0,\infty)\times S)$. T is Lipschitz continuous (the discount factor $\beta<1$ is the Lipschitz constant), if equipped with the sup-norm $\|f\|_\infty := \sup_{x\geq 0,\, s\in S}|f(x,s)|$. The Banach fixed point theorem thus ensures that (3) has a unique solution, and further that $|V|\leq \frac{\|g\|_\infty}{1-\beta}$. The value function is thus uniformly bounded, provided that $g(\cdot)$ is uniformly bounded, and V is continuous, provided that g is continuous.

2.3 The Bellman equation for a decision maker in an uncertain economic environment

In an economic environment several decision makers are observed, each of whom makes decisions individually. We assume that each individual decision maker acts rationally, but he has more information than the observing economist. To model this situation we consider the augmented process $(X_k,\,\varepsilon_k)_{k=0}$, with the process ε independent of X. Every $\varepsilon_k = (\varepsilon_{k,u})_{u\in S}$ is a vector carrying the additional information which is associated with the actions $u\in S$. This information is hidden to the observing economist, just the state X_k is available to him.

Every individual decision maker bases his decision on the augmented state space process (X_k,ε_k) and the payoff function $g\,(x,\varepsilon,s;\,u)$. In analogy to (1) it follows that

$$V(x,\varepsilon,s) = \max_{u\in S}\, g\,(x,\varepsilon,s;\,u) + \beta\cdot\mathbb{E}\left(\int V(X_1,\varepsilon_1,u)\mathcal{E}\left(\mathrm{d}\varepsilon_1|X_1\right)\middle|\, X_0=x\right) \tag{4}$$

has to hold for the value function. The transition ε of the unobserved parameter, 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; V is the value function, and $\mathcal{E}$ is the probability measure for ε (conditional independence, cf. Rust [25]).

Define now the *expected value function*, or s-alternative-specific value function

$$v(x,s) := \mathbb{E}\left(\int V\left(X_1,\varepsilon_1,s\right)\mathcal{E}\left(\mathrm{d}\varepsilon_1\right)\middle|\, X_0=x\right). \tag{5}$$

Then the Bellman equation (4) becomes

$$V(x,\varepsilon,s) = \max_{u\in S}\, g\,(x,\varepsilon,s;\,u) + \beta\cdot v\,(x,u) \tag{6}$$

and, by taking expectations of (6),

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

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 error term in a linear way (additive separability, cf. Rust [25]) according to $g(x, \varepsilon, s, u) = g(x, s, u) + \varepsilon_u$, and by specifying the distribution of $\mathcal{E}$. The Gumbel distribution is closed under maximization, and in this case a closed form formula for expectations is available and given by

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

as is detailed in Proposition 8 in the Appendix. Specifying $c_u := g(X_1, s; u) + \beta \cdot v(X_1, u)$ and applying (7) to

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

reduces the inner integral. The fixed point equation (8) simplifies to

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

Remark 1. Notice, that by assuming a Gumbel distribution the inner integral in (7) simplifies to a simple expression involving the logarithm of a sum of exponentials, but the integrals with respect to ε_1 disappear. The Gumbel distribution is an *extreme value type I* distribution. A proof that Gumbel's distribution is closed under maximization is provided in Proposition 8 in the Appendix, such that (9) is justified.

Assuming that $\mathcal{E}$ is a Gumbel distribution is, of course, restrictive. The choice of a Gumbel distribution is due to Rust, while the Generalized Extreme Value Models with conditional logit choice dates back to a series of papers by McFadden (cf., for example, McFadden [19]). The advantage of choosing a Gumbel distribution is the simple expression (9), and the explicit formula for the conditional choice probability (Proposition 9 in the Appendix), which will be exploited as well in what follows.

Remark 2 (The special case $b = 0$). It should be mentioned that the parameter b smoothes the kink in the function $(g_u)_{u\in Z} \mapsto \max\{g_u : u \in S\}$. The approximation quality increases, whenever b decreases to 0. In particular it holds that

$$\max\left\{g_u :\ u \in U\right\} \le b \cdot \log \sum_{u\in U} e^{g_u/b} \xrightarrow[b\to 0]{} \max\left\{g_u :\ u \in U\right\},$$

such that (9) results in

$$v\left(x, s\right) = \mathbb{E}\left(\max_{u\in S}\ g\left(X_1, s;\, u\right) + \beta \cdot v\left(X_1, u\right) \middle| X_0 = x\right)$$

in the limit whenever $b = 0$. This is a recursion for $v(x, s) = \mathbb{E}\left(V\left(X_1, s\right) \middle| X_0 = x\right)$, where V is the usual value function in Section 2.2. This recovers the classical Bellman theory, but involving the Gumbel distribution leads to a more general, still tractable model. In this situation ($b = 0$) the genuine value function can be recovered, by (2), as $V(x, s) = \max_{u\in S}\ g\left(x, s;\, u\right) + \beta \cdot v\left(x, u\right)$.

The additional parameter b can be interpreted as a degree of uncertainty, as the standard deviation of a Gumbel distribution is $b\frac{\pi}{\sqrt{6}} \simeq 1.28\, b$, where b is the scale parameter. 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.

It is convenient to introduce the operator

$$t_g(v)(x,s) := \mathbb{E}\left(b \cdot \log \sum_{u \in S} \exp \frac{g\left(X_1, s;\, u\right) + \beta \cdot v\left(X_1, u\right)}{b} \middle| X_0 = x \right). \tag{10}$$

Eq. (9) then rewrites as a fixed point equation on $\ell^\infty\left([0,\infty) \times S\right)$ as

$$v = t_g\left(v\right),$$

which is consistent with (3). t_g is again a contraction with Lipschitz constant $\beta < 1$ and Banach's fixed point theorem ensures that (10) has a unique solution (which we call v_g) in the proper space.

Remark 3 (The optimal policy). Eq. (2) provides the optimal strategy for the individual manager by maximizing $u \in S$. This is not the case for the expected value function (9). The reason for this difference is because an optimal decision cannot be specified in a random environment. The function v is an average over all decision makers, per (5). The concept of a single *optimal decision* does not make sense in the present context, as v is the average over all decision makers.

3 Structural estimation

The goal of structural estimation is to uncover switching costs associated with economic state changes. Our non-parametric approach assures that the switching costs we seek to estimate appear only in the payoff function $g(\cdot)$. The best model can be selected by a maximum likelihood approach, that is by solving the problem (cf. Su and Judd [28])

$$\begin{array}{ll} \text{maximize} & \mathcal{L}\left(g, v_g, (X_i, s_i, u_i)_{i=1}^N\right) \\ \text{subject to} & v_g = t_g\left(v_g\right), \\ & g \in \mathcal{G}, \end{array} \tag{11}$$

where $\mathcal{L}$ is the likelihood of observing data $(X_i, s_i, u_i)_{i=1}^N$ conditional on the payoff function $g \in \mathcal{G}$ and N is the number of observations.

In the lingo of Su and Judd [28], $\mathcal{L}(g, v, X)$ is an *augmented* likelihood function, because $\mathcal{L}$ involves the function v as an auxiliary variable. The payoff function $g \in \mathcal{G}$ is chosen from a set $\mathcal{G}$ of potential candidate functions. t_g is the operator (10) for this function $g(\cdot)$. v_g is the expected value function corresponding to the payoff $g(\cdot)$ satisfying the constraint $v_g = t_g(v_g)$.

To make a maximum likelihood estimator available we need an expression for the probability of choice. We explicitly use

$$P_v\left(u \middle| x, s\right) = \frac{\exp\left(\frac{g(x,s;\, u) + \beta\, v(x,u)}{b}\right)}{\sum_{u' \in D} \exp\left(\frac{g(x,s;\, u') + \beta\, v(x,u')}{b}\right)}, \tag{12}$$

which follows from the assumption that the process ε follows a Gumbel distribution. Formula (12) is detailed in Proposition 9 in the Appendix. The likelihood function $\mathcal{L}$ is thus

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

Remark 4. 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 (11) or (16) below. This is the most expensive part of the computational problem. The solution v_g maximizing (11) is a by-product of the maximum likelihood process and has the statistical interpretation of a nuisance parameter.

The approach described below solves the problem. It introduces a direct estimator for t_g, which is free of parameters. Moreover, the approach described ensures convergence to the continous 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 (11) involves the operator t_g. To evaluate $t_g(v)$ at a specified point x (cf. (10)) it is necessary to evaluate a conditional expectation: t_g is an expectation, conditional on $\{X_0 = x\}$. To estimate the conditional expectation of $f\left(X_1\right)$ relative to X_0, that is $\mathbb{E}\left(f\left(X_{k+1}\right) \mid X_k\right)$, we pair subsequent observations and consider

$$\left(X_i, X_{i+1}\right) \qquad i = 1, 2, \dots N-1. \tag{13}$$

Then the Nadaraya–Watson estimator[3] for the operator

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

is

$$\hat{t}_g(v)\left(x, s\right) := \sum_{i=1}^{N-1} \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{i'=1}^{N-1} K\left(\frac{x - X_{i'}}{h}\right)} \cdot b \cdot \log \sum_{u \in S} \exp \frac{g\left(X_{i+1}, s;\, u\right) + \beta \cdot v\left(X_{i+1}, u\right)}{b}, \tag{14}$$

where $K(\cdot)$ is an appropriate kernel function and $h > 0$ a suitable bandwidth. Consistency of this estimator, in an even broader context, is justified in Atuncar et al. [1].

The estimator $\hat{t}_g$ maintains all properties of the original operator t_g, as the following lemma reveals.

Lemma 5. *For the choice $\beta < 1$ the mapping $v \mapsto \hat{t}_g(v)$ is a contraction on $\ell^\infty\left([0, \infty) \times S\right)$, and $v = \hat{t}_g(v)$ has a unique fixed point.*

[3]The Nadaraya–Watson operator is well-known from kernel regression.

Proof.

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

is an affine linear function in v with slope $\beta < 1$, from which the assertion that $\hat{t}_g$ is a contraction is immediate. To insure the existence of a fixed point, observe that the function $g(\cdot)$ is evaluated at only finitely many points $(X_{i+1})_{i=1}^{N-1}$, such that $g(\cdot)$ is uniformly bounded. Hence Banach's fixed point theorem applies and uniqueness is ensured. □

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. The following corollary is immediate.

Corollary 6 (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*

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

Then

$$\hat{\hat{t}}_g\left((v_0^s, \ldots v_d^s)_{s \in S}\right) := \left(\sum_{i=1}^{N-1} \frac{K\left(\frac{x_j - X_i}{h}\right)}{\sum_{i'=1}^{N-1} K\left(\frac{x_j - X_{i'}}{h}\right)} \cdot b \log \sum_{u \in S} \exp \frac{g(X_{i+1}, s;\, u) + \beta \cdot I(v_0^u, \ldots v_d^u)(X_{i+1})}{b} \right)_{j=0}^{d} \tag{15}$$

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).

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

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

convenient, 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, other kernels provide reasonable results as well. For the bandwidth we chose a number (4.13) by trial and error.

4 Case Study

The specific application we consider is the case of shutting down, restarting, and abandoning peak power plants. We use a unique data set of plants located in the United States. We limit our sample to simple cycle combustion turbine power plants (hereafter CTs) located in the northeastern part of the United States.[4]

The value of CTs lies in their ability to respond to high prices typical of electricity wholesale markets. We choose to focus on peaking plants because we think they are the plants most susceptible to shutdown, start up, and/or abandonment due to economics reasons.[5] Economic considerations determine the threshold values for state changes. These threshold values ought to account for the one-time cost to change states, that is switching costs. It is these switching costs we seek to estimate.

In this section we describe the data and define the payoff function $g(\cdot)$ used in the optimization exercise.

4.1 Data

The owners of power plants in the United States must each year file Form 860 with the Energy Information Administration, hereafter EIA. Included in Form 860 is the status of each power plant. For our purposes, the relevant statuses are as follows.

- OP — operating,
- SB — standby, and,
- RE — retired.

A plant in state OP is available for operation. A plant in state SB has been shutdown and cannot be made ready for operation in the short term.[6] A plant which in state RE has been abandoned and cannot return to service.

- We define a *shutdown* to occur when a plant moves from state OP in year i to state SB in year $i+1$ and label this transition $OP \to SB$.[7]
- We define a *start up* to occur when plant moves from state SB in year i to state OP in year $i+1$ and label this transition $SB \to OP$.
- We define an *abandonment* to occur when a plant moves from state SB in year i to state RE in year $i+1$ and label this transition $SB \to RE$.

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

[4] See Fleten, Haugom, and Ullrich [12] for further details about the choice of plants and geographic locations.

[5] On the other hand base load plants are designed to run at full output for extended periods of time. Base load plants are profitable in nearly all market conditions. Shutdown, start up, and abandonment decisions for base load plants are more likely to be determined by other issues, e.g., mechanical failure or regulatory pressure.

[6] 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*

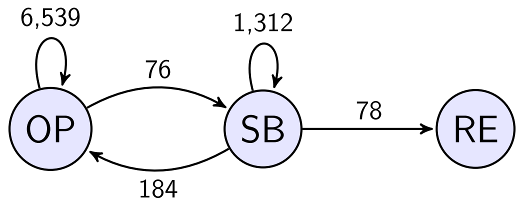

Figure 1: Transitions between the three states *operating* (OP), *standby* (SB) and *retired* (RE). Included is the number of transitions.

Augmented likelihood function. Our data consist of distinct groups. There are 6,539 occurrences an operating plants continuing to operate ($OP \to OP$), but only 76 occurrences of shutdown ($OP \to SB$). This is not accounted for in the log-likelihood as (11) treats each observation in the same way.

Following the literature (cf., for example, King and Zeng [16]) it is natural to augment the likelihood to reflect the different sample sizes of the groups. This is accomplished by the augmented likelihood

$$\begin{array}{rl} \text{maximize} & \sum_{i=1}^{N} \frac{1}{N_i} \log P_{v_g}\left(u_i \middle| X_i, s_i\right) \\ \text{subject to} & v_g = t_g\left(v_g\right), \\ & g \in \mathcal{G}. \end{array} \tag{16}$$

Here,

$$N_i \in \{6,539;\ 76;\ 184;\ 1,312;\ 78\}$$

is the sample size of the group to which the observation (u_i, s_i, X_i) belongs to according Figure 1 (instead of $N = 8,189$ in (11)) ($i \in \{OP \to OP,\ OP \to SB,\ SB \to OP,\ SB \to SB,\ SB \to RE\}$). As a consequence the five groups provided in Figure 1 are equally weighted, and the weights within the group are chosen to reflect the individual importance of each group.

4.2 Observation

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

(i) the operating state of the power plant $s_i \in S$ in the current year,

(ii) the profitability X_i during the year, and,

(iii) the decision of the manager regarding the operating state $u_i \in S$ of the power plant in the upcoming year.

for service in a short period of time, usually requires three to six months to reactivate."

[7] Intermediate and peaking power plants can be shutdown overnight. This short term shutdown option (cycling) is not what we consider in this paper. Our focus is on the decision to shutdown a plant for an extended period of time, sometimes referred to as laying-up or mothballing in the real options literature.

4.3 Profitability

Spread Options. 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 can be viewed as a collection of daily European call options on the spark spread. Consider a plant which has heat rate HR in units of $MMBtu/MWh$.[8] We calculate the plant-specific spark spread ($SPRD_n$) expressed in units of dollars per megawatt hour ($\$/MWh$), for day n as

$$SPRD_n = P_n^{elec} - HR * P_n^{fuel} - VOM,$$

where P_n^{elec} is the day n electricity price ($\$/MWh$), P_n^{fuel} is the day n fuel price ($\$/MMBtu$), and VOM ($\$/MWh$) is the variable non-fuel generation cost.[9]

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

$$X_i = \sum_{n=1}^{T_i} \max\left(SPRD_n, 0\right) * \left(\frac{16}{1000\, kW/MW}\right),$$

where 16 is the number of peak hours[10] in a day and T_i is the number of days in year i. The max function captures the optionality of the plant. On days for which the spread is negative, the plant does not operate and the profit is zero.

Observation Pairs. As discussed in (13), the optimization relies on pairs of observations (X_i, X_{i+1}), that is the profitability in the current year i and in the upcoming year $i+1$. At the time of the decision, the profitability for the upcoming year is not yet known.[11]

We calculate profitability in both the current year i and the upcoming year $i+1$ using actual electricity prices and fuel prices. Because we use actual fuel and electricity prices, together with plant-specific heat rate information, we can 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_i is the profitability which would have obtained if the plant had been in state OP in yr i.[12]

Table 1 presents summary statistics for profitability.[13] Figure 2 presents the evolution of profitability from one year to the next.[14] The density in Figure 2 is estimated based on the pairs

[8] The heat rate of a power plant is the amount of fuel required, measured in millions of British thermal units ($MMBtu$), required to generate one megawatt hour (MWh) of electricity. A lower number indicates greater efficiency.

[9] Daily spot prices for New York Harbor No. 2 Oil and NYMEX Henry Hub natural gas are taken from the EIA website. Electricity prices come from the PJM, ISO-NE, and NYISO websites.

[10] 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 from 06:00 ("hour ending" 7, or HE7 in industry parlance) through 22:00 (HE22). We obtain daily peak prices by taking the simple average of the hourly spot prices during the peak period.

[11] In practice the plant manager would rely upon sophisticated software to simulate the operation of the regional electric system and therefrom derive an estimate of profitability for the upcoming year $i+1$. So-called production costing software (e.g., PROSYM, UPLAN, EGEAS, PowrSym) is common in the industry.

[12] We assume that if the plant had been operating the effect on electricity prices and fuel prices would have been negligible. This is a reasonable assumption for the peaking plants in our sample.

[13] Notice that the total average profitability for the current year ($12.5/kW, from the final column) is roughly one third higher than the average profitability for the upcoming year ($9.0/kW). This is because the first year of our sample, 2001, was the single most profitable year during the sample period, with average profitability $37.7/kW, not shown in the table. Because 2001 is the first year of our sample, it is not included in the averages for the upcoming year.

[14] The volume under the surface sums to 1.

Transition	$OP \to OP$	$OP \to SB$	$SB \to OP$	$SB \to SB$	$SB \to RE$	Total
Observations	6,539	76	184	1,312	78	8,189
State variable current year						
Average in \$/ kW	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
State variable upcoming year						
Average \$/ kW	9.5	4.4	6.3	7.1	N/A	9.0
Standard deviation	10.5	7.6	8.1	7.6	N/A	10.0

Table 1: Summary statistics (average, and standard deviation in \$/kW) of the state variable

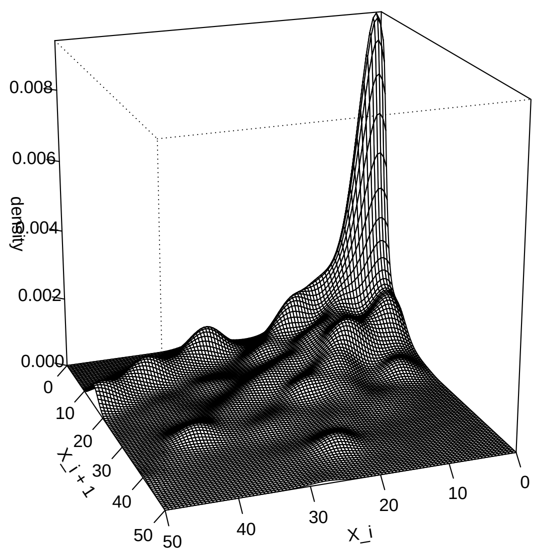

Figure 2: Bivariate density of the observed transition (X_i, X_{i+1}) of the profitability indicator in $\$/kW$.

(X_i, X_{i+1}) (cf. (13)), which are available from the observations: one ordinate represents the profitability indicator of this year, X_i, the other ordinate the profitability indicator in the subsequent year, X_{i+1}. The density thus describes the Markov kernel, which is used in the expectation to compute the value function, for example in (9).

Remark 7 (Lack of evidence for a parametrized model). Figure 2 gives evidence that there is no distinguishable pattern of transitions of the profitability from a year to the next. This is a clear indicator that the model cannot be treated by employing standard tools, e.g., by assuming a (geometric) Brownian motion.[15]

[15] If the transitions were described by a GBM, then a slice of the density plot should look lognormal.

4.4 Unobserved heterogeneity

Heterogeneity means that different decision makers will react differently, even though the state variables are the same. This reflects different cost functions, strategic positions and switching opportunities, which are unobserved by the economist. We assume that the immediate payoff from each decision has a Gumbel distribution.[16] As discussed above, we model unobserved heterogeneity by the parameter b. It follows from (9) that the dimension of b is $\$/kW$. To estimate unobserved heterogeneity we employ a logistic regression for binary classification with parameters $\theta = (\alpha, \eta)$ and probabilities

$$P(OP \,|\, X) = \frac{1}{1 + \exp\left(-\frac{X-\alpha}{\eta}\right)} \qquad \text{and} \qquad P(SB \,|\, X) = \frac{1}{1 + \exp\left(\frac{X-\alpha}{\eta}\right)}. \tag{17}$$

(Note, that $P(OP \,|\, X) + P(SB \,|\, X) = 1$). 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 α and η in (17) allow a natural interpretation. α is a location parameter, the median of the distribution dividing the managers' decisions into the two groups *operating*, that is $\{X > \alpha\}$ and *standby* ($\{X < \alpha\}$, respectively). The scale parameter η with dimension $\$/kW$ describes the standard deviation of these decisions; that is the uncertainty within these decisions, or the respective heterogeneity of all managerial decisions.

Numerical computations of the parameter η give evidence that $\eta \approx \$5.3/kW$ for the data in our database. We choose $b = \$3.3/kW$, $b = \$5.3/kW$, and $b = \$7.3/kW$ in the presentation of numerical results below to demonstrate the sensitivity of this particular parameter.

4.5 The payoff function $g(\cdot)$

The payoff function $g(\cdot)$ describes the expected cash flow for the year. The profitability indicator x for the year is a calculated value, as discussed above.

We also must include two other costs, (i) the costs of continuing maintenance M_u given that the generator is in state u in the subsequent year, and, (ii) the costs associated with the transitions themselves, $K_{s \to u}$. The payoff function is given by

$$g(x, s;\, u) \quad = \quad \begin{cases} x - M_{OP} & \text{if } s = \text{operating and } u = \text{operating}, \\ x/2 - K_{OP \to SB} & \text{if } s = \text{operating and } u = \text{standby}, \\ x/2 - K_{SB \to OP} & \text{if } s = \text{standby and } u = \text{operating}, \\ -M_{SB} & \text{if } s = \text{standby and } u = \text{standby}, \\ -K_{SB \to RE} & \text{if } s = \text{standby and } u = \text{retired}, \\ -\infty & \text{else}. \end{cases} \tag{18}$$

To exclude other transitions the value $-\infty$ is included in the payoff function.[17] The parameters of

[16] Corollary 10 in the Appendix elaborates that the difference of Gumbel variables has logistic distribution, such that assuming a logistic distribution with parameters as in (17) is consistent with (23) in the Appendix, and thus consistent with the general approach to structural estimation.

[17] Notice that for plants which are either shutdown or started up, we include only half of the profit, that is $x/2$. Because our data is at the annual frequency, we do not know when during the year the state changes. We assume that shutdowns and start-ups happen mid-year so that in both cases the plant is assumed operation for half of the year. Also, for the year in which a state change occurred, we attribute all costs to switching costs rather than to costs of continuing maintenance.

interest are the switching costs $K_{OP\to SB}$, $K_{SB\to OP}$ and $K_{SB\to RE}$.[18]

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 6, it is possible to employ linear interpolation by fixing supporting points $x_0 < x_1 < \dots x_d$. The problem is

$$\begin{array}{ll} \text{maximize} & \frac{1}{N}\sum_{i=1}^{N} \log P_{I(v_g)}\left(u_i \mid X_i, s_i\right) \\ \text{subject to} & v_g = \hat{\hat{t}}_g(v_g), \\ & g \in \mathcal{G}, \end{array} \tag{19}$$

where $g \in \mathcal{G}$, in view of (18), means that

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

are the variables in the optimization procedure (19). We impose the constraints $M_{OP} \geq 0$, $K_{OP\to SB} \geq 0$, $K_{SB\to OP} \geq 0$ and $M_{SB} \geq 0$ on our optimization procedure, and this is reflected in the functions $g \in \mathcal{G}$ as well.

Irrespective of the supporting points $x_0 < \dots < x_d$ this problem has a solution, the problem is always feasible for every choice of

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

By augmenting the sequence $x_0 < \dots < x_d$ by additional points a net is obtained, which converges finally to the value function v, the solution of (11).

5 Results

The first row of results in Table 2 presents the estimated maintenance costs (M_{OP}) for a plant which is in the operating state. The estimates range from \$6.4/kW to \$10.4/kW. According to the assumptions in the 2009 Annual Energy Outlook prepared by the EIA, the annual fixed maintenance costs for a conventional combustion turbine are \$12.11/kW.[19] The close correspondence of our results to industry estimates gives us confidence that the estimation process is effective.

The second and third rows of results in Table 2 indicate that both the switching costs incurred for shutdown ($K_{OP\to SB}$) and the maintenance cost for a plant which is in the shutdown state (M_{SB}) are low for most combinations of interest rates and the heterogeneity parameter b. It is reasonable that the maintenance cost of a plant in the shutdown state is lower than that in the operating state;

[18]While start ups ($OP \to SB$) and shutdowns ($SB \to OP$) involve cash outflows (switching costs), retirements ($SB \to RE$) may result in a cash inflow. There is an active secondary market for used combustion turbines. That is, it may be that $K_{SB\to RE} < 0$. A negative value for $K_{SB\to RE}$ would then be interpreted as a resale (or salvage) value.

[19]See Table 8.2 - Cost and Performance Characteristics of New Central Station Electricity Generating Technologies in EIA [10], on page 89. The 2002 AEO has a corresponding fixed O&M cost of \$6.45/kW.

Interest rate	5 %			10 %			15 %		
Level of heterogeneity, b	3.3	5.3	7.3	3.3	5.3	7.3	3.3	5.3	7.3
M_{OP}	8.4	9.3	10.4	6.4	7.2	8.2	6.3	6.7	6.4
$K_{OP \to SB}$	3.9	6.2	7.7	2.6	4.6	5.7	1.7	2.1	3.9
M_{SB}	2.1	3.3	5.1	0.0	1.2	2.8	0.1	0.9	0.9
$K_{SB \to OP}$	6.8	7.1	9.0	4.1	4.7	6.5	5.1	6.4	4.8
$K_{SB \to RE}$	-38.5	-46.2	-47.2	-40.0	-44.4	-46.3	-25.8	-31.2	-43.8

Table 2: Costs for changing the operational mode, depending on different interest rates and heterogeneity imposed (quantities in \$/kW)

otherwise it would never pay off to shut down at all. Once the generator is shutdown management often chooses to invest little to nothing in its upkeep.[20]

The fourth row of results in Table 2 provides the estimated switching costs incurred for start up ($K_{SB \to OP}$). These costs range from \$4.1/kW to \$9.0/kW, less than the continuing maintenance costs for an operating plant.

Finally, the switching costs associated with an abandonment ($K_{SB \to RE}$) are negative, that is the owner of the plant can expect to receive a cash inflow upon retirement of the plant. This result is consistent with the existence of a secondary market for used CTs, or with the value of freeing the space for a replacement plant. The salvage values reported in Table 2 range from \$25.8/kW to \$47.2/kW. According the the 2009 AEO, the cost for a brand new (conventional) CT is \$638/kW. Therefore our estimated salvage values range from 4% to 7% of the cost of a brand new CT, a result which seems reasonable.

6 Discussion

Current projections from NERC indicate that solar and wind power generating technologies will increase by an order of magnitude or more in the United States over the next ten years. In Europe the advent of large supplies of renewable energy, particularly wind and solar, has reduced the market value of gas-fired combustion turbine-based plants.[21] It is likely that the similar issues will arise in the U.S. – renewable plants such as solar and wind will reduce energy prices during peak periods.

The flexibility provided by gas-fired peak plants is an important part of the electricity system. Peak plants perform two important functions in electricity markets by providing quick-start and load-following capabilities. Both of of these functions are essential for maintaining the integrity of the electric grid.[22] As a peak plant becomes less profitable, the owner must consider exercising the real option to either (i) shutdown, or (ii) permanently abandon the plant. Large scale shutdown and/or abandonment of peak plants could endanger system reliability.

[20]We thank Paul Clark of the City of Tallahassee, Florida and Steve Marshall of Lakeland Electric for their expertise and guidance in these matters.

[21]See for example Caldecott and McDaniels [8].

[22]Because electricity cannot be stored, a shock which reduces the supply of electricity relative to demand (e.g., if an operating generator were forced out of service) may result in the need quickly start up a peak plant. Nonstorability means there can be no inventory of electricity, therefore supply and demand must balance in real time. Many peak plants are capable of varying output in real time in order to match demand, or load-following.

Decreasing reserve margins in the UK have prompted the proposal of a supplemental balancing reserve – a payment to compensate the owners of existing plants in order to forestall early shutdowns of gas-fired plants. The calculation of the exact value of any such payment should account for the costs incurred in a shutdown. The decision to continue operation of a plant which would have otherwise been shutdown means that the owner will not have to incur a shutdown cost.

The existence of supplemental balancing reserve payments changes the economics for owners of mothballed plants. The owners of these plants will have a new incentive to restart. Hence, policy makers should consider restart costs when calculating the amount of the payment.

Our results contribute by providing realistic estimates of these switching costs which can be accounted for by both plant owners and for regulators concerned with ensuring system reliability.

Conclusions

In this paper we describe a new method for structural estimation and apply it to the estimation of switching costs for a sample of peak power plants in the U.S. The approach combines a non-parametric regression for capturing transitions in the exogenous profitability state variable, with a one-step nonlinear optimization for the structural estimation. This allows for efficiency and few assumptions; e.g. we do not need to represent the expected value function with polynomial basis functions. The data we use are particularly well-suited to test the new method because the state variable is not described by any known stochastic process. Understanding how switching is performed by merchant power plant owners is important because the functions served by peak plants are vital for a smooth and reliable operation of the electric system, reducing the risk of load shedding and blackouts. Our estimates of switching costs are reasonable when compared with industry estimates for the costs of new turbines. The results suggest that the proposed non-parametric method is useful for analyzing data for which it is important to relax assumptions about the state process.

7 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.

References

[1] G. Atuncar, C. Dorea, and C. 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. URL http://dx.doi.org/10.1007/s11203-006-9003-7.

[2] 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.

[3] A. Bensoussan and J. L. Lions. *Impulse control and quasi-variational inequalities*. Gauthier-Villars, 1984.

[4] 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.

[5] K. Brekke and B. Øksendal. Optimal switching in an economic activity under uncertainty. *SIAM Journal on Control and Optimization*, 32(4):1021–1036, 1994. doi: 10.1137/S0363012992229835. URL http://epubs.siam.org/doi/abs/10.1137/S0363012992229835.

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

[7] C. Burr. Subsidies, tariffs and investments in the solar power market. *Working Paper, University of Arizona*, 2012.

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

[9] A. Dixit. Entry and exit decisions under uncertainty. *Journal of Political Economy*, 97:620–638, 1989.

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

[11] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, 2006.

[12] S.-E. Fleten, E. Haugom, and C. J. Ullrich. Keeping the lights on until the regulator makes up his mind. Working paper, Available at SSRN: http://ssrn.com/abstract=192861, 2013.

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

[14] E. Guerre, I. Perrigne, and Q. Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000.

[15] R. Kellogg. The effect of uncertainty on investment: Evidence from Texas oil drilling. National Bureau of Economic Research, 2010.

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

[17] T. Li, I. Perrigne, and Q. Vuong. Structural estimation of the affiliated private value auction model. *RAND Journal of Economics*, pages 171–193, 2002.

[18] C.-Y. C. Lin and K. Thome. Investment in corn-ethanol plants in the Midwestern United States: An analysis using reduced-form and structural models. Working Paper, University of California at Davis, Davis, 2013.

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

[20] L. Muehlenbachs. Testing for avoidance of environmental obligations. *Resources for the Future Discussion Paper 12–12*, 2012.

[21] 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. URL http://dx.doi.org/10.1007/s00477-006-0063-4.

[22] W. K. Newey, J. L. Powell, and F. Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603, 1999.

[23] D. Rapson. Durable goods and long-run electricity demand: A case study of air conditioner purchase behavior. *Working Paper, University of California, Davis*, 2010.

[24] G. Rothwell and J. Rust. On the optimal lifetime of nuclear power plants. *Journal of Business & Economic Statistics*, 15(2):195–208, 1997.

[25] J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033, Sept. 1987. doi: http://www.jstor.org/stable/1911259.

[26] J. Rust and G. Rothwell. Optimal response to a shift in regulatory regime: The case of the US nuclear power industry. *Journal of Applied Econometrics*, 10:S75–S118, 1995.

[27] A. N. Shiryaev. *Probability.* Springer, New York, 1996.

[28] C.-L. Su and K. L. Judd. Constrained optimization approaches to estimation of structural models. *Econometrica*, 80:2213–2230, 2012.

[29] A. J. Triantis and J. E. Hodder. Valuing flexibility as a complex option. *Journal of Finance*, 45(2):549–565, 1990.

Appendix

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. The cumulative distribution function (cdf) of a Gumbel distribution is $F(z) = \exp\left(-e^{-\frac{z-\mu}{b}-\gamma}\right)$, where $\gamma = 0.57721566\ldots$ is the Euler–Mascheroni constant. Its mean is μ, and the variance is $b^2\frac{\pi^2}{6}$.

Proposition 8 (The extreme value distribution is closed under maximization)**.** *Let* $(\varepsilon_i)_{i=1}^n$ *be independent random variables which are Gumbel distributed with mean* μ_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. 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 9 (Choice probabilities for shifted Gumbel variables)**.** *Let* $(\varepsilon_i)_{i=1}^{n}$ *be independent Gumbel distributed random variables with individual mean* μ_i *and common scale parameter* $b > 0$*. Then the probability of choice for the variables shifted by* c_i *is*

$$P\left(\varepsilon_1 + c_1 = \max_{i\in\{1,2,\dots n\}} \varepsilon_i + c_i\right) = \frac{\exp\left(\frac{c_1+\mu_1}{b}\right)}{\exp\left(\frac{c_1+\mu_1}{b}\right) + \dots + \exp\left(\frac{c_n+\mu_n}{b}\right)}. \tag{20}$$

Proof. 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 (20) itself is Gumbel distributed by Proposition 8.

Thus

$$\begin{aligned} P\left(\varepsilon_1 + c_1 \ge \varepsilon_2 + c_2\right) &= P\left(\varepsilon_2 \le \varepsilon_1 + c_1 - c_2\right) \\ &= \int_{-\infty}^{\infty} f\left(x_1\right) \int_{-\infty}^{x_1+c_1-c_2} f\left(x_2\right) \mathrm{d}x_2 \mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} f\left(x_1\right) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) \mathrm{d}x_1, \end{aligned} \tag{21}$$

where the cdf of the Gumbel distribution has been substituted. By substituting the probability density function (pdf) f, (21) continues as

$$\begin{aligned} P\left(\varepsilon_1 + c_1 \ge \varepsilon_2 + c_2\right) &= \int_{-\infty}^{\infty} \frac{1}{b} \exp\left(-\frac{x_1}{b} - e^{-\frac{x_1}{b}}\right) \exp\left(-e^{-\frac{x_1+c_1-c_2}{b}}\right) \mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{b} e^{-\frac{x_1}{b}} \exp\left(-e^{-\frac{x_1}{b}}\left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right) \mathrm{d}x_1 \\ &= \left[\frac{\exp\left(-e^{-\frac{x_1}{b}}\left(1 + e^{-\frac{c_1-c_2}{b}}\right)\right)}{1 + e^{-\frac{c_1-c_2}{b}}}\right]_{x_1=-\infty}^{\infty} \\ &= \frac{1}{1 + e^{-\frac{c_1-c_2}{b}}} = \frac{e^{\frac{c_1}{b}}}{e^{\frac{c_1}{b}} + e^{\frac{c_2}{b}}}. \end{aligned} \tag{22}$$

This completes the proof. □

Finally we give provide a proof that the difference of Gumbel variables enjoys a logistic distribution (cf. Nadarajah [21]).

Corollary 10. *If ε_1 and ε_2 are Gumbel distributed with mean μ_1 and μ_2 and common scale parameter $b > 0$. Then the difference $\varepsilon := \varepsilon_2 - \varepsilon_1$ follows a logistic distribution with mean $\mu = \mu_2 - \mu_1$ and cumulative distribution function*

$$F_\varepsilon(z) = \frac{1}{1 + \exp\left(-\frac{z-\mu}{b}\right)}, \tag{23}$$

which is the distribution function of a logistic variable.

Proof. If follows from (22) in the proof of the preceding theorem that

$$F_\varepsilon(z) = P\left(\varepsilon_2 - \varepsilon_1 \leq z\right) = P\left(\varepsilon_1 + z \geq \varepsilon_2\right) = \frac{1}{1 + e^{-\frac{z-(\mu_2-\mu_1)}{b}}},$$

which completes the proof. □