Estimation of market distribution functions in electricity pool markets

Golbon Zakeri, Geoffrey Pritchard and Andrew Philpott
University of Auckland

Abstract

The market distribution function is a probability distribution that can be used to model the randomness in generation level and clearing price that generators in electricity pool markets must take account of when they submit their offers. We discuss estimation techniques for this distribution, and discuss ways of measuring the quality of these estimators, in particular using classical statistical approaches as well as expected excess cost.

1 Introduction

Since the advent of the wholesale electricity market in New Zealand in 1996, generators have been competing to supply power to consumers as part of a wholesale electricity pool market. In electricity pool markets generators submit an offer of generation to a central system operator who dispatches power from the offered generation so as to minimize the total cost of generating power. Each generation offer takes the form of a supply function or *offer stack*, which specifies the price at which a generator is prepared to offer a certain amount of power to the market. Although the exact form of the offer stack varies with the market design, typically these can be represented by piecewise constant functions, consisting of a finite number of tranches of power.

The prices at which the generated power is offered are at the discretion of the generator. If the price asked is too high then they are unlikely to be dispatched and so will earn no revenue. On the other hand, the price should not under normal circumstances be chosen to be less than the marginal cost of generation. In a perfectly competitive environment, in which the choice of offer stack from an individual generator has no effect on the price, a stack that offers power at its marginal cost maximizes the profit of the generator However most electricity markets have a small number of large generating companies whose offers have an effect on the clearing price and quantity they are dispatched. In this setting, generators are faced with the question of what offer stack to submit in order to maximize their profit.

With perfect information on the demand for electricity, and the offers of other generators, it is possible for a generator to compute a stack that maximizes its profit. For some simple cases it is possible to compute Nash equilibria for the one-shot game where generators each choose an optimal supply function to submit assuming that the others do not change their (optimal) offers. Although of interest in a theoretical sense, supply-function equilibria are extremely difficult to compute for all but the simplest models, and are of limited use in practical hour-by-hour trading operations.

The optimal stacks to submit to the market on an hour-by-hour basis will be computed with little information on the offers of the other generators, and with imperfect forecasts of demand. This makes the profit R associated with an offer stack a function of the random quantity Q that the generator is dispatched as well as the (random) clearing price P. In this circumstance the generator might seek an offer stack to maximize the expectation of R with respect to some probability distribution.

An approach to this problem is described in (Anderson and Philpott 2002), who define a *market distribution function* $\Psi(q,p)$ as being the probability of a generator not being fully dispatched if they were to offer a single quantity q at price p. Anderson and Philpott show that the expected profit from offering a stack defined by a curve $\mathbf{s}$ is

$$E[R] = \int_s R(q,p)d\Psi(q,p).$$

In this paper we develop a methodology based on maximum likelihood estimation for constructing estimates of $\Psi(q,p)$ from historical dispatch data. We begin in the next section by providing an intuitive derivation of the market distribution function, using an approach that makes the estimation procedure straightforward. In section 3 we describe the estimation technique, and in section 4 we study the classical statistical properties of this estimator. In particular we show that our estimator is consistent, asymptotically unbiassed, and obeys a central limit theorem. Finally in section 5 we conclude by discussing the performance of our estimator when being used in optimization.

2 Market distribution functions revisited

In this section we shall review the concept of a market distribution function. Although these have been formally defined in (Anderson and Philpott 2002), here we adopt a different approach that will provide the structure to make a statistical analysis more straightforward.

In (Anderson and Philpott 2002) an offer stack is modelled as a parameterized curve. However for the purposes of this paper, it is useful to regard an offer stack $\mathbf{s}$ as being defined by a continuous planar *offer curve*: a connected, totally ordered (with respect to the order $(q_1,p_1) \geq (q_2,p_2) \iff q_1 \geq q_2$ and $p_1 \geq p_2$), subset of the plane (i.e. we are considering increasing curves). It will often be convenient, if x and y are points on the offer curve $\mathbf{s}$ to write $x \leq y$ if they are so ordered. Note that an offer stack is an offer curve consisting of a union of horizontal and vertical lines. This definition ensures that the point of dispatch (Q,P) will always lie on the offer curve $\mathbf{s}$.

For a given offer curve $\mathbf{s}$, the distribution of the point of dispatch (Q,P) on $\mathbf{s}$ can be described by its cumulative distribution function $\Psi_\mathbf{s}$. That is, $\Psi_\mathbf{s}(q,p) = P((Q,P) \leq (q,p))$ for all (q,p) on $\mathbf{s}$. The problem of optimizing generator returns then becomes

$$\text{maximize}_\mathbf{s}\ \overline{R}(\mathbf{s}) = \int_s R(q,p)d\Psi_\mathbf{s}. \tag{1}$$

Fortunately, in many circumstances, including the setting of an electricity market, it is possible to simplify the description of the dispatch distribution by specifying a single function Ψ, defined on the whole (q,p) plane, such that $\Psi_{\mathbf{s}}(q,p)=\Psi(q,p)$ for every $\mathbf{s}$ and (q,p) on $\mathbf{s}$. This is clearly desirable as it would simplify the optimization problem (1) to

$$\text{maximize}_{\mathbf{s}}\ \overline{R}(\mathbf{s})=\int_s R(q,p)d\Psi. \tag{2}$$

The function Ψ is easily recognized to be the market distribution function. The extension of $\Psi_{\mathbf{s}}$ to the whole (q,p) plane is made possible by the following proposition.

Proposition. Suppose the market chooses points of dispatch in the following way. First, a random non-increasing right-continuous function $U(q)$ is generated. The point of dispatch (Q,P) chosen for any offer curve $\mathbf{s}$ will then be the point where $\mathbf{s}$ intersects the graph of $p=U(q)$. (The total order on $\mathbf{s}$ allows us to write this more precisely as

$$(Q,P)=\operatorname{argmin}\{(q,p)\text{ on }\mathbf{s}|U(q)\leq p\}.$$

The minimum is achieved since $\{(q,p)|U(q)\leq p\}$ is a closed set. This takes care of cases where the intersection is not a unique point, or does not exist.) Then for any point (q,p), $\Psi_{\mathbf{s}}(q,p)$ takes a common value $\Psi(q,p)$ for all offer curves $\mathbf{s}$ passing through (q,p). The function Ψ is non-decreasing in both q and p.

Remark. U may be thought of as a marginal utility function, giving the market's willingness to pay for each additional unit of quantity from this producer. The actual construction of U will depend on market structure, but U is ultimately a representation of the consumer demand function and competing offers from other producers. Note that this construction occurs when the point of dispatch arises as the solution to a convex optimization problem which is (implicitly or explicitly) solved by the market. This occurs in a pool-type electricity market – see (Anderson and Philpott 2002) – in which case U is referred to as a "residual demand function".

Proof. For any point (q,p) on an offer curve $\mathbf{s}$, $P((Q,P)\leq(q,p))=P(U(q)\leq p)$. This quantity is clearly non-decreasing in both q and p and independent of the offer curve $\mathbf{s}$.

It is worthwhile to note that the distribution of the residual demand function U, as a random function in a function space, contains much more information than is contained in Ψ. However, Ψ is necessary and sufficient for the formulation of the offer curve problem.

3 Estimation of Ψ

Suppose that a sample of past points of dispatch, perhaps arising from several different offer curves, is available for the generator in question. In what follows we describe a method for constructing a grid estimator for the market distribution function Ψ based on the historical dispatch information and analyze the quality of this estimator. To this end, we make the following assumptions

1. The generator submitted their offers bids in the offer stack form required by the market. That is an offer curve which is a step function, starting at $(0,0)$ (or some other standard point) and finishing at some (q_M,p_M) where q_M denotes that maximum production quantity and p_M denotes the maximum price (perhaps $p_M=\infty$).

2. Each observation of dispatch comes in the form (q, p, r) where q indicates the dispatch quantity, p denotes the clearing price and $r \in \{h, v\}$ denotes whether the point of dispatch (q, p) was located on a horizontal $(r = h)$ or a vertical $(r = v)$ segment of the offer stack.

3. The dispatch observations are independent and they come from the same underlying market distribution function Ψ.

A maximum likelihood approach to this problem would choose $\hat{\Psi}$ from among the monotone functions taking values on $[0, 1]$ in such a way as to maximize the probability of having observed the sample. Let (q, p, h) be an observed sample point. Then using the definition of Ψ, let

$$\rho(q, p, h) = \lim_{\epsilon \to 0} (\Psi(q + \epsilon, p) - \Psi(q - \epsilon, p)).$$

Similarly For the observed sample point (q, p, v) let

$$\rho(q, p, v) = \lim_{\epsilon \to 0} \Psi(q, p + \epsilon) - \Psi(q, p - \epsilon)).$$

Therefore in order to maximize the likelihood of the observed sample it is sufficient to maximize

$$\prod_{i=1}^{n} \rho(q_i, p_i, r_i), \text{ or equivalently } \sum_{i=1}^{n} \log(v_i^1 - v_i^0) \tag{3}$$

where $v_i^1 = \lim_{\epsilon \to 0} \hat{\Psi}(q_i + \epsilon, p_i)$, $v_i^0 = \lim_{\epsilon \to 0} \hat{\Psi}(q_i - \epsilon, p_i)$ if $r_i = h$ and $v_i^1 = \lim_{\epsilon \to 0} \hat{\Psi}(q, p_i + \epsilon)$, $v_i^0 = \lim_{\epsilon \to 0} \hat{\Psi}(q, p_i - \epsilon)$, if $r_i = v$. Therefore the problem of selecting a monotone $\hat{\Psi}$ that maximizes the probability of having observed the sample reduces to selecting a finite number of values v_i^1 and v_i^0 that comply with the monotonicity constraints. This is fortunate as we can now restrict ourselves to monotone functions that have piecewise constant value and select from among these, one that maximizes (3) subject to the bound constraints.

It is easy to see that the restriction of $\hat{\Psi}$ to piecewise constant monotone functions will not compromise the objective value of the maximum likelihood problem (3) as long as a sufficient number of cells (pieces) are available to assign values to. In particular, for any subdivision of the plane into cells, it is possible to achieve the maximum likelihood with a function that is constant on each cell, provided $v_1^0, \ldots, v_n^0, v_1^1, \ldots, v_1^n$ represent the values on distinct cells.

In order to formulate the maximum likelihood estimator for the market distribution function Ψ, we construct a grid of cells on which the piecewise constant $\hat{\Psi}$ will be defined.

Definition of the grid estimator. In the (q, p) plane, draw a vertical line through each point (q_i, p_i) with $r_i = h$ and a horizontal line through each point (q_i, p_i) with $r_i = v$. This divides the plane into a number of rectangular cells. We will take $\hat{\Psi}$ to be constant on each cell and denote the value of $\hat{\Psi}$ on cell c by v_c. The maximum likelihood *grid estimator* of $\hat{\Psi}$ is then the solution to

$$\begin{array}{ll} \max & \sum_{i=1}^{n} \log(v_{c_i^1} - v_{c_i^0}) \\ \text{s/t} & v_c \le v_d \quad \text{when } c \le d \\ & 0 \le v_c \le \quad \forall c, \end{array} \tag{4}$$

where for the ith observation, c_i^1 is the cell above the sample point (q_i, p_i) if $r_i = h$ and c_i^0 is the cell below, and similarly, c_i^1 is the cell to the right of the sample point (q_i, p_i) if $r_i = v$ and c_i^0 is the cell to the right. The *enhanced grid estimator* is defined the same way, except that during the subdivision into cells, extra horizontal and vertical lines are added where necessary, to ensure that the cells $c_1^0, \ldots, c_n^0, c_1^1, \ldots, c_n^1$ are all distinct.

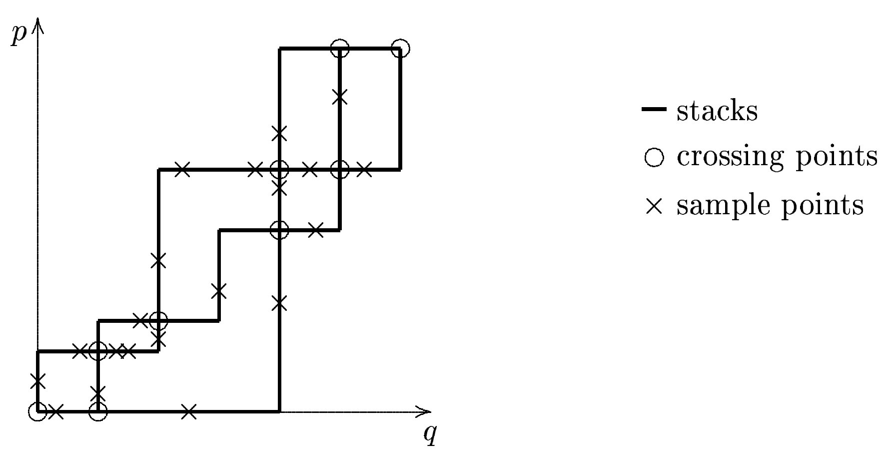
4 Statistical properties of the maximum likelihood grid estimator

In this section and the next we examine the quality of the estimator defined above. We take two approaches to this problem. First in this section we consider the classical statistical approach: bias, consistency and central limit behaviour.

Theorem 1. Statistical properties of the grid estimator. Let Ψ be a market distribution function which is strictly monotone in the plane (i.e. $(q_1, p_1) \leq (q_2, p_2)$ and $\Psi(q_1, p_1) = \Psi(q_2, p_2)$ only if $(q_1, p_1) = (q_2, p_2)$). Let $s_1, \ldots s_m$ be fixed offer stacks. Suppose a sample of $n = \sum_{j=1}^m N_j$ independent points of dispatch is obtained, by presenting each stack s_j to the market N_j times, and let $\hat{\Psi}$ be the grid estimator (or the enhanced grid estimator) derived from this sample. Then for any point $x = (q, p) \in \bigcup_{j=1}^m s_j$:

- $\hat{\Psi}(x)$ is asymptotically unbiased (i.e. $E\left[\hat{\Psi}(x)\right] \to \Psi(x)$);
- $\hat{\Psi}(x)$ is consistent (i.e. $\hat{\Psi}(x) \to \Psi(x)$ in probability);
- $\hat{\Psi}(x)$ obeys a central limit theorem: $\sqrt{n}(\hat{\Psi}(x) - \Psi(x)) \xrightarrow{D} N(0, \sigma_x^2)$ for some σ_x

as $\min_j N_j \to \infty$.



A typical example with $m = 3$ distinct stacks and a total of $n = 20$ dispatch points in the sample.

In order to establish these properties, it will be helpful to consider a simplified version of the estimator.

Suppose we have stacks $s_1, \ldots, s_m$ as in the theorem. The intersection of any two stacks is a union of isolated points and piecewise linear curves. We refer to these isolated points, and the endpoints of the piecewise linear curves, as *crossing points*. Let C denote the (finite) set of all crossing points that arise from pairwise intersections among $s_1, \ldots, s_m$. Let A denote the set of (disjoint) arcs which connect the crossing points, each arc being a subset of one or more of the stacks. For $a \in A$, let $a^0 \in C$ and $a^1 \in C$ be the endpoints of a, with $a^0 \leq a^1$.

Note that if we wish to find a maximum-likelihood estimate of Ψ on C, we can take the function $\phi : C \to [0,1]$ solving

$$\begin{array}{rl} \max & \sum_{a\in A} n_a \log(\phi(a^1) - \phi(a^0)) \\ \text{s.t} & \phi(c) \leq \phi(d) \qquad \text{for all } c \leq d \end{array}$$

where n_a is the number of dispatch points in the sample which fall on a.

Having done this, we can extend ϕ to $S = \bigcup_{j=1}^m s_j$ as follows. If $x = (q,p)$ is a point on $a \in A$, let $n_a(x)$ be the number of dispatch points in the sample which fall on a between a^0 and x. Then let

$$\phi(x) = \frac{n_a(x)\phi(a^0) + (n_a - n_a(x))\phi(a^1)}{n_a}$$

if $n_a > 0$, and

$$\phi(x) = \frac{(\Psi(a^1) - \Psi(x))\phi(a^0) + (\Psi(x) - \Psi(a^0))\phi(a^1)}{\Psi(a^1) - \Psi(a^0)}$$

if $n_a = 0$.

It should be noted that, although theoretically sound, the above construction (which we will refer to as the *simplified estimator*) has some unsatisfactory features as a practical statistic. For one thing, a (small) part of its definition involves the very Ψ it is intended to estimate. More seriously, it may not be monotone: we may have $(q_1,p_1) \leq (q_2,p_2)$ but $\phi(q_1,p_1) > \phi(q_2,p_2)$, if (q_1,p_1) and (q_2,p_2) belong to different stacks. Our grid estimator may be thought of as a modified version of the simplified estimator that corrects these defects. The price to be paid for this is the introduction of a little (asymptotically vanishing) bias.

Lemma. The simplified estimator ϕ discussed above has the following properties:

- Asymptotic unbiasedness: for each $x \in S$, $E[\phi(x)] \to \Psi(x)$;
- Uniform (strong) consistency: $\sup_{x\in S} |\phi(x) - \Psi(x)| \to 0$ with probability 1;
- A central limit theorem: for each $x \in S$, $\sqrt{n}(\phi(x) - \Psi(x)) \xrightarrow{\mathcal{D}} N(0, \sigma_x^2)$ for some σ_x;

as $\min_j N_j \to \infty$.

Proof of Lemma. The standard theory of maximum likelihood estimators (see e.g. (Kendall and Stuart 1979)) gives the result for each $x \in C$, since here we are considering only a parametric estimation problem with finitely many parameters. We have only to extend the results to all $x \in S$.

Let x be a point on the arc $a \in A$. Note that

$$\begin{aligned} \phi(x) - \Psi(x) &= (1-p_x)(\phi(a^0) - \Psi(a^0)) + p_x(\phi(a^1) - \Psi(a^1)) \\ &+ 1_{n_a>0}(\phi(a^1) - \phi(a^0))\left(\frac{n_a(x)}{n_a} - p_x\right), \end{aligned} \tag{5}$$

where $p_x = (\Psi(x) - \Psi(a^0)/(\Psi(a^1) - \Psi(a^0))$.

For the unbiasedness, let the σ-field $\mathcal{G} = \sigma\{n_r : r \in A\}$. Note that $\phi(c) \in \mathcal{G}$ for all $c \in C$, while $E[n_a(x) \mid \mathcal{G}] = p_x n_a$. (The sample may be thought of as being generated by first generating the arc counts n_r, and then (independently) generating the exact locations of the dispatch points on each arc.) Hence from (5) we obtain

$$E[\phi(x) - \Psi(x) \mid \mathcal{G}] = (1 - p_x)(\phi(a^0) - \Psi(a^0)) + p_x(\phi(a^1) - \Psi(a^1)).$$

Since $E[\phi(c)] \to \Psi(c)$ for all $c \in C$, taking a further expectation yields $E[\phi(x)] \to \Psi(x)$.

For the uniform consistency, note from (5) that

$$\sup_{x \in S} |\phi(x) - \Psi(x)| \leq \max_{c \in C} |\phi(c) - \Psi(c)| + \max_{a \in A} \sup_{x \in a} \left| \frac{n_a(x)}{n_a} - p_x \right|.$$

For each $a \in A$, we have $\sup_{x \in a} \left| \frac{n_a(x)}{n_a} - p_x \right| \to 0$ a.s. by the Glivenko-Cantelli theorem (Durrett 1996, p. 59). Also, $\phi(c) \to \Psi(c)$ a.s. for each $c \in C$. Since A and C are finite sets, the result follows.

For the central limit theorem, note that

$$\sqrt{n} \begin{pmatrix} \phi(a^0) - \Psi(a^0) \\ \phi(a^1) - \Psi(a^1) \\ \frac{n_a(x)}{n_a} - p_x \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

for a suitable 3×3 matrix Σ. Hence

$$\begin{aligned} &\sqrt{n}((1 - p_x)(\phi(a^0) - \Psi(a^0)) + p_x(\phi(a^1) - \Psi(a^1)) \\ &\quad + (\Psi(a^1) - \Psi(a^0)) \left(\tfrac{n_a(x)}{n_a} - p_x \right)) \xrightarrow{\mathcal{D}} N(0, \sigma_x^2) \end{aligned}$$

for a suitable number σ_x. The result then follows by (5) and the converging-together lemma (Durrett 1996, p. 91). ∎

Proof of Theorem 1. Let D_1 be the event that each cell in the grid either contains exactly one crossing point of C, or intersects exactly one arc of A, or does not intersect any arc. For either the grid estimator or the enhanced grid estimator, the strict monotonicity condition on Ψ implies that $P(D_1) \to 1$ as $\min_j N_j \to \infty$.

Supposing the event D_1 to have occurred, consider the problem 4, which is solved to determine the values of the grid estimator. Suppose we relax (i.e. ignore) all the monotonicity constraints $v_c \leq v_d$, except for those where cells c and d intersect a common arc $a \in A$. Note that the values assigned to cells which do not intersect any arc are now completely free; these cells will be ignored from now on. Also, for each arc a, those cells which intersect only a form a "chain" subproblem of the form

$$\begin{aligned} \max \quad & \textstyle\sum_{i \in I} \log(v_i - v_{i-1}) \\ \text{s.t.} \quad & \alpha = v_0 \leq v_1 \leq \cdots \leq v_k = \beta \end{aligned}$$

the optimum of which determines values for these cells independently of the rest of the problem. Here $I \subseteq \{1, \ldots, k\}$ represents those cell-cell boundaries where a dispatch point of the sample occurs on the arc a, and α and β are the cell values at the endpoints of the arc, which for the purposes of the subproblem may be assumed to be fixed. It is straightforward to see that the optimum of this subproblem is given by $v_i = \#(I \cap \{1, \ldots, i\})/(\#I)$ (i.e. take equal-sized steps at each dispatch point), and the optimal value is $(\#I)\log((\beta-\alpha)/(\#I))$. That is, we can express the optimal cell values for all cells along an arc in terms of the optimal values for cells at the endpoints. Taking this into account, the main problem simplifies to one in which only the values for cells containing crossing points need to be found. This simplified problem is identical (apart from an additional constant in the objective) to that used to construct the "simplified estimator" above. It follows that the optimum of our relaxed problem agrees with the simplified estimator ϕ on any arc containing at least one dispatch point.

Let D_2 be the event that each arc contains at least one dispatch point. It is clear from the strict monotonicity of Ψ that $P(D_2) \to 1$ as $\min_j N_j \to \infty$.

Let D_3 be the event that the optimum of the relaxed problem also satisfies the constraints that were relaxed (and, therefore, agrees with the optimum of the problem as originally stated). The strict monotonicity of Ψ, together with the uniform consistency of ϕ as shown in the above lemma, gives that $P(D_3) \to 1$ as $\min_j N_j \to \infty$.

It is now clear that $P\left(\hat{\Psi}(x) = \phi(x)\ \forall x \in S\right) \geq P(D_1 \cap D_2 \cap D_3) \to 1$. From this (together with the fact that $\left|\hat{\Psi} - \phi\right| \leq 1$), all the conclusions of the theorem follow:

The asymptotic unbiasedness follows as

$$E\left[\left|\hat{\Psi}(x) - \Psi(x)\right|\right] \leq E\left[|\phi(x) - \Psi(x)|\right] + P\left(\hat{\Psi}(x) \neq \phi(x)\right) \to 0.$$

The consistency follows as

$$\left|\hat{\Psi}(x) - \Psi(x)\right| \leq |\phi(x) - \Psi(x)| + 1_{\hat{\Psi}(x)\neq\phi(x)} \xrightarrow{P} 0.$$

The central limit theorem follows as

$$\sqrt{n}(\hat{\Psi}(x) - \Psi(x)) = \sqrt{n}(\phi(x) - \Psi(x)) + \sqrt{n}(\hat{\Psi}(x) - \phi(x)) \xrightarrow{D} N(0, \sigma_x^2),$$

by the converging-together lemma. ∎

5 Expected excess cost

Although it is important and in itself interesting to measure the quality of the grid estimator $\hat{\Psi}$ in classical statistical terms, we would also like to measure the impact of this estimation in the context of optimization.

Once the grid estimator of the market distribution function is found, we can solve

$$\text{maximize}_{\mathbf{s}} \int_{\mathbf{s}} R(q,p)\,d\hat{\Psi} \tag{6}$$

by restricting the offer curves $\mathbf{s}$ to offer stacks defined by the grid and using dynamic programming.

At each step of the dynamic program, we consider a particular cell c, and we determine which direction (vertical or horizontal) to transition from that cell in order to maximize the return along the offer stack that originates in cell c and

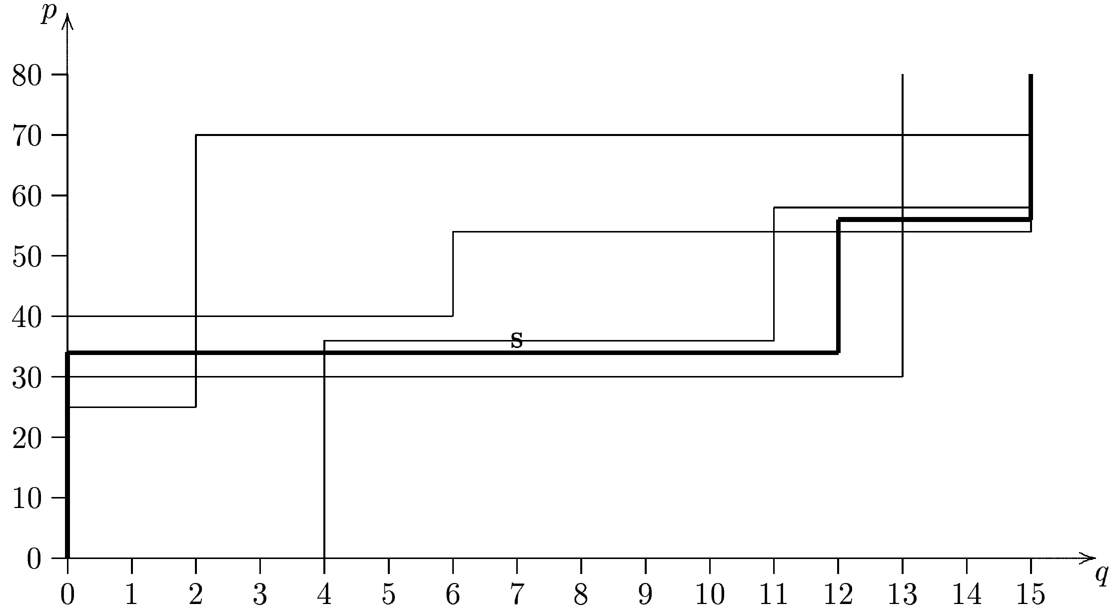


Figure 1: Offer stacks used to generate sample dispatch points.

terminates in the top right call. Once the bottom left cell of the grid is reached, we would have the solution to (6).

Suppose that R is increasing in both components, e.g. $R(q,p) = q * p$. Then the optimal offer stack for the grid estimator $\hat{\Psi}$ transitions from cell c at right corner which we label (q_c, p_c). If the direction of transition is vertical the incremental return is

$$R(q_c, p_c) * (\hat{\Psi}_{c^1} - \hat{\Psi}_c)$$

where c^1 is the cell above cell c. Similarly if the transition is horizontal, the incremental return is

$$R(q_c, p_c) * (\hat{\Psi}_{c^0} - \hat{\Psi}_c)$$

where c^0 is the cell to the right of cell c.

Once $\hat{\mathbf{s}} = \text{argmax}_{\mathbf{s}} \int_{\mathbf{s}} R(q,p) d\hat{\Psi}$ is obtained it is crucial to have an idea of how it performs under the real market distribution function Ψ. The expected excess cost of $\hat{\mathbf{s}}$ is defined by

$$\text{maximize}_{\mathbf{s}} \int_{\mathbf{s}} R(q,p) d\Psi - \int_{\hat{\mathbf{s}}} R(q,p) d\Psi.$$

We have experimented with constructing $\hat{\Psi}_1$ based on observations that came from a single offer stack repeatedly submitted to a market simulator versus constructing $\hat{\Psi}_5$ based on observations that came from 5 different offer stacks. Figure 1 contains the 5 offer stacks with the single offer stack $\mathbf{s}$, used for constructing $\hat{\Psi}_1$, highlighted. In each case the total number of observation where 100. We constructed estimated optimal offer stacks $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_5$ based on the above algorithm. We then simulated the return from these stacks under the "real" market simulator Ψ. The normalized expected excess cost associated with $\hat{\mathbf{s}}_1$ was 13.3% and the normalized expected excess cost of $\hat{\mathbf{s}}_5$ turned out to be 6%.

Although further numerical tests need to be carried out to verify the above, it seems that the grid estimator, even with a limited number of offer stacks that generate sample points, results in a reasonable "approximately optimal" offer stack. Furthermore, in all of experiments, using more offer stacks to generate sample points have resulted in better estimators of Ψ in the sense of generating better "approximately optimal" offer stacks.

6 Acknowledgments

The authors would like to thank Mark Wilson for helpful comments on the presentation of this paper.

References

Anderson, E. J., and A. B. Philpott. 2002. "Optimal offer construction in electricity markets." *Mathematics of Operations Research* 27:82–100.

Durrett, R. 1996. *Probability, theory and examples.* 2nd.

Kendall, M., and A. Stuart. 1979. *The advanced theory of statistics.* 4th. Volume 2.