Non-parametric estimation of market distribution functions in electricity pool markets

Geoffrey Pritchard*, Golbon Zakeri, and Andrew Philpott
University of Auckland

May 6, 2005

Abstract

The market distribution function is a probabilistic device that can be used to model the randomness in dispatch and clearing price that generators in electricity pool markets must take account of when submitting offers. We discuss techniques for estimating the market distribution function, and ways of measuring the quality of these estimators, using both classical statistical approaches and an expected-foregone-revenue approach.

Key words and phrases: electricity market, market distribution function, non-parametric estimation

1 Introduction

A common form of electricity market structure that has emerged in recent years is the electricity pool market. Examples are the Pennsylvania-Jersey-Maryland (PJM) market, the Nordpool, and the electricity markets in New Zealand and Australia. In electricity pool markets, generators submit offers of generation to a central system operator, who dispatches some of the offered generation in such a way as to minimize the total cost of the power generated. 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 non-decreasing 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

*Corresponding author. Dept. of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand.

This work has been supported by the New Zealand Public Good Science Fund.

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 that 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 [1], where a *market distribution function* $\Psi(q,p)$ is defined as the probability of a generator not being fully dispatched if it were to offer a single quantity q at price p. In that paper, it is shown that the expected profit from offering a stack defined by a curve $\mathbf{s}$ is

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

As an aside, we note that the market distribution function is not, as one might expect from the name, solely a property of the market. Rather, it describes the market situation faced by a particular generator. Thus, there is a different market distribution function for each generator.

The market distribution function in [1] is assumed to be continuous. Under further smoothness assumptions on Ψ it is possible to derive optimality criteria for $\mathbf{s}$. In [3], the theory is extended to the case of discontinuous Ψ. For this case, the existence of an optimal offer $\mathbf{s}$ is not guaranteed, and ϵ-optimal solutions are sought instead. The present paper mostly considers this more general case, in which market distribution functions will not be assumed to have any continuity or smoothness properties except where this is explicitly stated.

In this paper we develop a methodology based on maximum likelihood estimation for constructing estimates of $\Psi(q,p)$ from historical dispatch data. The layout of the paper is as follows. In the next section, we provide 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 unbiased, and obeys a central limit theorem. Some numerical investigation of the performance of the estimator is conducted in section 5.

2 Market distribution functions

Market distribution functions have been formally defined in [1]. In this paper, we adopt a slightly different formulation that will provide the structure to make a statistical analysis more straightforward.

In [1] an offer stack is modelled as a parameterized curve. We will 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. That is, an increasing curve. 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. An offer stack is an offer curve consisting of a union of horizontal and vertical lines.

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 expected 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 (a single trading period in) 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 of [1].

The extension of $\Psi_{\mathbf{s}}$ to the whole (q,p) plane is made possible by the following argument. 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

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

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.) 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 [1] – in which case U is referred to as a "residual demand function".

Proposition. Suppose the market chooses points of dispatch in the way described above. 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.

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 worth noting 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. Part of the appeal of the market distribution function is that it contains just enough (i.e. necessary and sufficient) information to formulate the offer curve optimization problem.

An important special case is that of a *price-taker*: a generator too small to influence the market price by any action it might take. For a price-taker, $\Psi(q, p)$ is a function of p only. However, the methods in the present paper are intended to apply to the general (*price-making*) case, where $\Psi(q, p)$ may depend on both q and p.

It is also worth noting that the function Ψ need not, in general, be continuous. If, for example, the distribution of the random function U were partly discrete, the resulting Ψ would have discontinuities at all points along the graph of any function u with $P(U = u) > 0$. In many real electricity markets, it is observed that some generators often choose their offers at "favourite" prices. This will give rise to a Ψ function with discontinuities at those prices. In particular, the rules of the Australian electricity market – with variable quantities, but offer prices for each generator that are fixed throughout the day – make this behaviour almost universal. All that can be said about Ψ in general is that it is a monotone increasing function of each of its variables. This is sufficient for the line integral in (2) to be well-defined, and to coincide with the expected return.

3 Estimation of Ψ

In this section we are concerned with the statistical problem of estimating a market distribution (Ψ) function from dispatches which have occurred under similar conditions in the past.

A traditional statistical approach to the estimation problem might begin with a parametric model for Ψ, and attempt to estimate the parameters using the maximum likelihood (or another) technique. Since observations on the dispatch are being obtained in every trading period, it is natural to use these in a Bayesian approach to update a prior belief about Ψ. One way of doing this using a smooth parametric model is described in [2].

However, smooth parametric models are not easy to adapt to the discontinuous market distribution functions that we wish to consider. One of the main difficulties lies in constructing a natural model that can cogently be defended. It is known that a market distribution function Ψ must be a monotone function in both its variables, but it is unclear how to introduce any model structure beyond this. Only if the approximate form of the Ψ function (along with its discontinuities) is already known or estimated does the parametric approach become reasonable.

Our approach is to estimate the form of Ψ using non-parametric maximum likelihood methods. Given a sample of outcomes in historical trading periods, we will choose $\hat{\Psi}$ from among the monotone functions taking values in $[0, 1]$ in such a way as to maximize the probability of having observed the sample.

Suppose that a sample of past points of dispatch, perhaps arising from several different offer curves, is available for the generator in question. More precisely, we assume that the data have been collected in the following way.

1. The offer stack associated with each point of dispatch has the standard step-function form required by the market, starting at $(0, 0)$ (or some other standard point) and finishing at some (q_M, p_M) where q_M denotes the maximum production quantity and p_M denotes the maximum price (which could be $p_M = \infty$).

2. For the ith sample point we record (q_i, p_i, r_i), where q_i, p_i are the dispatch quantity and clearing price, and $r_i \in \{h, v\}$ indicates whether the point of dispatch (q_i, p_i) occurred on a horizontal ($r_i = h$) or a vertical ($r_i = v$) segment of the offer stack.

3. The dispatch observations are independent and have been generated with the same underlying market distribution function Ψ.

The indicator r_i is not public information in most markets, but may be assumed to be available to a generator considering its own historical offers.

The independence assumption may be questionable in practice, and it is pertinent to consider the potential effect of possible dependence on estimators constructed under such an assumption. While the estimate of Ψ itself might be little affected, the standard error of such an estimate (which we do not consider in the present paper) might well be rather different if the underlying data were dependent. This ought to be borne in mind in future work.

Let (q, p, h) be an observed sample point occurring on a horizontal segment of its associated offer stack. Then the probability of having observed this particular point is

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

This quantity is well-defined for any possible function Ψ, although it could, of course, be zero. Similarly, for the observed sample point (q, p, v) on a vertical segment of its associated offer stack,

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

Therefore, in order to maximize the likelihood of the observed sample, it is sufficient to choose the estimate $\hat{\Psi}$ to maximize

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

where $v_{c_i^1} = \lim_{\epsilon \to 0} \hat{\Psi}(q_i + \epsilon, p_i)$ and $v_{c_i^0} = \lim_{\epsilon \to 0} \hat{\Psi}(q_i - \epsilon, p_i)$ if $r_i = h$, and $v_{c_i^1} = \lim_{\epsilon \to 0} \hat{\Psi}(q, p_i + \epsilon)$ and $v_{c_i^0} = \lim_{\epsilon \to 0} \hat{\Psi}(q, p_i - \epsilon)$ if $r_i = v$. The problem of selecting a monotone $\hat{\Psi}$ that maximizes the probability of having observed the sample thus reduces to selecting a finite number of values $v_{c_i^1}$ and $v_{c_i^0}$ that maximize (3) and satisfy some monotonicity constraints placed on them by the requirement that $\hat{\Psi}$ be monotone.

It is easy to see that the restriction of $\hat{\Psi}$ to piecewise constant monotone functions will not alter the objective value of the maximum likelihood problem (3) as long as sufficiently many cells (pieces) are available to enable $v_{c_1^0}, \ldots, v_{c_n^0}, v_{c_1^1}, \ldots, v_{c_n^1}$ to be all distinct. We will construct an estimator of this kind in which the pieces are given by a rectangular grid of cells.

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 Ψ to be constant on each cell, and denote the value of Ψ on cell c by v_c. In a natural extension of the notation above, the cells adjacent to the sample points can then be denoted $c_1^0, \ldots, c_n^0$ and $c_1^1, \ldots, c_n^1$. (See Figure 1.) The maximum likelihood *grid estimator* of Ψ is found by solving

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

The optimum of (4) will not usually be unique. (Some cell values do not appear in the objective, so are determined only by the monotonicity constraints.) In order to make an essentially arbitrary choice among optimal solutions, we suggest the following rule. For each cell c, let

$$[A_c, B_c] = \{t : \exists \text{ an optimal point of (4) with } v_c = t\},$$

and let the preferred solution be the one with $v_c = (A_c + B_c)/2$ for each c.

One advantage of this rule is that it is easy to implement. Observe that the value of any cell which appears in the objective of (4) is uniquely determined by (4). For any other cell c,

$$A_c = \max \{v_d : d \leq c \text{ and } d \text{ appears in the objective of (4)}\},$$

where $d \leq c$ indicates that d is below and/or to the left of c. Similarly,

$$B_c = \min \{v_d : d \geq c \text{ and } d \text{ appears in the objective of (4)}\}.$$

We may therefore proceed as follows: after an optimal solution to (4) has been found, identify those cells which do not appear in the objective of (4), and set their values according to the above rule.

Note that the solution so found clearly satisfies the monotonicity constraints, and so is itself an optimal solution of (4).

The grid estimator as defined above may not maximize the likelihood among all monotone functions, since there is no guarantee that the cells $c_1^0, \ldots, c_n^0, c_1^1, \ldots, c_n^1$

will be all distinct. This deficiency is overcome by a minor modification: the *enhanced grid estimator*. This is defined the same way as above, except that during the subdivision into cells, extra horizontal and vertical lines are added according to the following convention. Wherever two sample points lie on the boundary of a common cell, an extra line is added through the midpoint of each line segment of the shorter path along the boundary between them. If three or more points lie on the boundary of a common cell, extra lines are similarly introduced between each successive pair. These extra lines create sufficiently small cells that $c_1^0, \ldots, c_n^0, c_1^1, \ldots, c_n^1$ are all distinct.

The construction of the grid estimator is illustrated in Figure 1.

4 Theoretical properties of the grid estimator

In this section we examine the quality of our estimator from a theoretical point of view. We begin with the classical statistical approach: bias, consistency and central limit behaviour.

Theorem 4.1. Let Ψ be a market distribution function. Suppose that the set $U = \{(q,p) : 0 < \Psi(q,p) < 1\}$ is bounded and that on U, Ψ is continuous and strictly monotone (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 $\hat{\Psi}$ has the following behaviour on $S = \bigcup_{j=1}^m s_j$:

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

as $\min_j N_j \to \infty$.

Remark. The requirement that $\min_j N_j \to \infty$ means that these asymptotic properties hold (naturally enough) only at points (q,p) belonging to stacks which are sampled infinitely often.

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*. (See Figure 2.) 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

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

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 $(q_1,p_1) \le (q_2,p_2)$ 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 4.2. The simplified estimator ϕ discussed above has the following properties:

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

as $\min_j N_j \to \infty$.

Proof of Lemma 4.2. The standard theory of maximum likelihood estimators (see e.g. [7]) gives the required results for each fixed $x \in C$, 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 $E[n_a(x) \,|\, \mathcal{G}] = p_x n_a$. (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

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

It follows by taking a further expectation that

$$E[\phi(x) - \Psi(x)] = (1 - p_x)(E\left[\phi(a^0)\right] - \Psi(a^0)) + p_x(E\left[\phi(a^1)\right] - \Psi(a^1)),$$

and so

$$\sup_{x \in S} |E[\phi(x)] - \Psi(x)| \leq \sup_{x \in C} |E[\phi(x)] - \Psi(x)|.$$

The result then follows from the finiteness of C.

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 [6, 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

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

for a suitable 3×3 matrix Σ. Hence

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

for a suitable number σ_x. The result then follows by (5) and the converging-together lemma [6, p. 91]. ∎

Lemma 4.3. Let

$$Z = \{(x, y) : \exists i, j, z \text{ with } x \in s_i, y \in s_j, z \in s_i \cap s_j, x \leq z \leq y\}$$

and

$$K = \{(x, y) : x, y \in S \cap U, x \leq y, (x, y) \notin Z\}$$

Then

$$\inf_{(x,y) \in K} (\Psi(y) - \Psi(x)) > 0.$$

Proof. Suppose the contrary: that a sequence of pairs $(x_n, y_n) \in K$ has $\Psi(y_n) - \Psi(x_n) \to 0$. By passing to subsequences if necessary, we may assume that $x_n \to x$,

$y_n \to y$ where $x, y \in S$ and $x \leq y$. Also, since S consists of a finite union of line segments, we may assume (passing to further subsequences if necessary) that $x_n \in L$ and $y_n \in M$, where L and M are horizontal or vertical closed line segments of finite length. We have $x \in L$ and $y \in M$, and the condition $(x_n, y_n) \in K$ implies that L and M are disjoint. Hence $x \neq y$ and so $\Psi(y) > \Psi(x)$. The continuity of Ψ then provides a contradiction. ∎

Proof of Theorem 4.1. 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

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

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 $v_i = |I \cap \{1, \ldots, i\}| / |I|$ (where $|I|$ denotes the cardinality of the set 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). We have

$$\inf_{(x,y) \in K} (\phi(y) - \phi(x)) \geq \inf_{y \in S} (\phi(y) - \Psi(y)) + \inf_{(x,y) \in K} (\Psi(y) - \Psi(x)) + \inf_{x \in S} (\Psi(x) - \phi(x)),$$

where K is as in Lemma 4.3. Hence by Lemmas 4.2 and 4.3,

$$\underline{\lim}_{n} \inf_{(x,y) \in K} (\phi(y) - \phi(x)) \geq \inf_{(x,y) \in K} (\Psi(y) - \Psi(x)) > 0.$$

Since ϕ is monotone on each individual stack s_i, we see that if (x, y) belongs to the set Z defined in Lemma 4.3, then $\phi(y) \geq \phi(x)$. It thus follows that

$$\inf_{x,y \in S \cap U,\, x \leq y} (\phi(y)) - \phi(x)) \geq 0 \qquad \text{eventually, with probability 1,}$$

and hence 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 uniform asymptotic unbiasedness follows as

$$\sup_{x \in S} \left| E\left[\hat{\Psi}(x)\right] - \Psi(x) \right| \leq \sup_{x \in S} \left| E\left[\hat{\Psi}(x)\right] - E[\phi(x)] \right| + \sup_{x \in S} \left| E[\phi(x)] - \Psi(x) \right|.$$

The second term is covered by Lemma 4.2, while the first can be estimated as follows:

$$\begin{aligned} \sup_{x \in S} \left| E\left[\hat{\Psi}(x)\right] - E[\phi(x)] \right| &\leq \sup_{x \in S} E\left[\left|\hat{\Psi}(x) - \phi(x)\right|\right] \\ &\leq \sup_{x \in S} P\left(\hat{\Psi}(x) \neq \phi(x)\right) \\ &\leq P\left(\hat{\Psi}(x) \neq \phi(x) \text{ for some } x \in S\right). \end{aligned}$$

The uniform consistency follows as

$$\sup_{x \in S} \left| \hat{\Psi}(x) - \Psi(x) \right| \leq \sup_{x \in S} |\phi(x) - \Psi(x)| + 1_{\hat{\Psi}(x) \neq \phi(x) \text{ for some } x \in S} \xrightarrow{P} 0.$$

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. ∎

Theorem 4.1 shows that our estimator has some desirable properties at points (q, p) which belong to stacks that appear frequently in the data. But at points which are far from any such stack, there is no such guarantee. In order to obtain good behaviour uniformly in the plane, it is (reasonably enough) necessary to have a data set which explores the whole plane. This is formalized in the following result.

Theorem 4.4. Let $s_1, s_2, \ldots$ be a sequence of offer stacks such that $\bigcup_{j=1}^{\infty} s_j$ is dense in the bounded set $U = \{(q, p) : 0 < \Psi(q, p) < 1\}$. Suppose Ψ is continuous and strictly monotone on U. Let $\hat{\Psi}_m$ be the estimator of Ψ derived from a sample of $n = \sum_{j=1}^{m} N_j$ points, with N_j points from stack s_j for each $j = 1, \ldots, m$. Then there exists a sequence (ν_m) of integers such that when $\min_{j=1}^{m} N_j \geq \nu_m$, the corresponding $\hat{\Psi}_m$ are uniformly asymptotically unbiased and consistent over the whole plane. That is,

- $\sup_{x \in [0,\infty)^2} \left| E\left[\hat{\Psi}_m(x)\right] - \Psi(x) \right| \to 0;$

- $\sup_{x\in[0,\infty)^2} \left|\hat{\Psi}_m(x) - \Psi(x)\right| \to 0$ in probability

as $m \to \infty$.

Remark. Theorem 4.4 extends the asymptotic properties of Ψ to all points of the (q,p)-plane, including those which do not belong to any stack (*cf.* the remark following Theorem 4.1).

Proof. For each $x = (q,p) \in U$, let x_m^+ and x_m^- denote the closest points to x in $\left(\bigcup_{j=1}^m s_j \cup \{x : \Psi(x) = 1\}\right) \cap ([q,\infty) \times [p,\infty))$, and $\left(\bigcup_{j=1}^m s_j \cup \{x : \Psi(x) = 0\}\right) \cap ([0,q] \times [0,p])$ respectively. For $x \in [0,\infty)^2 \setminus U$, let $x_m^+ = x_m^- = x$. We first establish that our hypotheses are sufficient to imply

$$\sup_{x\in[0,\infty)^2} \left(\Psi(x_m^+) - \Psi(x_m^-)\right) \to 0 \qquad \text{as } m \to \infty.$$

For this, note that if $[0,\infty)^2$ is covered with a grid of spacing δ, then each grid cell $E_{ij} = [i\delta,(i+1)\delta] \times [j\delta,(j+1)\delta]$ with $E_{ij} \cap U \neq \emptyset$ must intersect $\bigcup_{j=1}^\infty s_j$, so there is some finite m such that $\bigcup_{j=1}^m s_j$ intersects each of the finitely many such E_{ij}. For this m, we have $\sup_{x\in[0,\infty)^2} \|x_m^+ - x_m^-\| \leq 3\delta$. Since $\delta > 0$ was arbitrary, it follows that $\sup_{x\in[0,\infty)^2} \|x_m^+ - x_m^-\| \to 0$ as $m \to \infty$. Moreover, our continuity assumption on Ψ actually implies that Ψ is uniformly continuous on $[0,\infty)^2$; hence $\sup_{x\in[0,\infty)^2} (\Psi(x_m^+) - \Psi(x_m^-)) \to 0$ as $m \to \infty$.

Now, for each fixed m, choose (by Theorem 4.1) ν_m such that when $\min_{j=1}^m N_j \geq \nu_m$, we have

$$\sup_{x\in\bigcup_{j=1}^m s_j} \left|E\left[\hat{\Psi}_m(x)\right] - \Psi(x)\right| < 1/m$$

and

$$P\left(\sup_{x\in\bigcup_{j=1}^m s_j} \left|\hat{\Psi}_m(x) - \Psi(x)\right| > 1/m\right) < 1/m.$$

Henceforth, we assume that $\min_{j=1}^m N_j \geq \nu_m$.

We can establish the unbiasedness result by observing that

$$E\left[\hat{\Psi}_m(x)\right] - \Psi(x) \leq E\left[\hat{\Psi}_m(x_m^+)\right] - \Psi(x_m^-) \leq 1/m + \Psi(x_m^+) - \Psi(x_m^-)$$

and

$$\Psi(x) - E\left[\hat{\Psi}_m(x)\right] \leq \Psi(x_m^+) - E\left[\hat{\Psi}_m(x_m^-)\right] \leq \Psi(x_m^+) - \Psi(x_m^-) + 1/m.$$

(Note that when $\Psi(x) = 0$ or 1, $\hat{\Psi}_m(x) = \Psi(x)$ with probability 1.) Hence

$$\lim_m \sup_{x\in[0,\infty)^2} \left|E\left[\hat{\Psi}_m(x)\right] - \Psi(x)\right| \leq \lim_m \sup_{x\in[0,\infty)^2} \left(\Psi(x_m^+) - \Psi(x_m^-)\right) = 0.$$

For the consistency result, let $\epsilon_m = 2\sup_{x\in[0,\infty)^2} (\Psi(x_m^+) - \Psi(x_m^-))$. Then

$$P\left(\sup_{x\in[0,\infty)^2} \left|\hat{\Psi}_m(x) - \Psi(x)\right| > \frac{2}{m} + \epsilon_m\right)$$

$$
\begin{aligned}
&\leq P\left(\sup_{x\in[0,\infty)^2}\left(\left(\hat{\Psi}_m(x_m^+)-\Psi(x_m^-)\right)_+ + \left(\Psi(x_m^+)-\hat{\Psi}_m(x_m^-)\right)_+\right) > \frac{2}{m}+\epsilon_m\right)\\
&\leq P\left(\sup_{x\in[0,\infty)^2}\left(\hat{\Psi}_m(x_m^+)-\Psi(x_m^+)\right)_+ + \sup_{x\in[0,\infty)^2}\left(\Psi(x_m^-)-\hat{\Psi}_m(x_m^-)\right)_+ > \frac{2}{m}\right)\\
&\leq P\left(\sup_{x\in[0,\infty)^2}\left|\hat{\Psi}_m(x_m^+)-\Psi(x_m^+)\right| > \frac{1}{m}\right) + P\left(\sup_{x\in[0,\infty)^2}\left|\Psi(x_m^-)-\hat{\Psi}_m(x_m^-)\right| > \frac{1}{m}\right)\\
&\to 0 \qquad \text{as } m\to\infty.
\end{aligned}
$$

∎

4.1 Expected foregone revenue

Although it is important and in itself interesting to measure the quality of the grid estimator $\hat{\Psi}$ in classical statistical terms, it is perhaps more important to measure its performance in the context of optimization. The theory of stochastic optimization provides a standard way to do this, which we outline below. (For a more comprehensive treatment, see [11, 10].)

Suppose the market has market distribution function Ψ. For any offer curve $\mathbf{s}$, let

$$F(\mathbf{s}) = \int_{\mathbf{s}} R(q,p)\, d\Psi.$$

Let $\mathbf{s}^*$ be a maximizer of F, i.e. an optimal offer curve. One of the main uses of any estimate $\hat{\Psi}$ of Ψ is to give us an estimate $\hat{\mathbf{s}}$ of $\mathbf{s}^*$, found by solving

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

The quality of $\hat{\mathbf{s}}$, and hence of $\hat{\Psi}$, can be gauged by the quantity $F(\mathbf{s}^*) - F(\hat{\mathbf{s}})$, the loss in objective value that will be experienced when using these estimates in place of the true Ψ and $\mathbf{s}^*$. This loss in the objective value is termed *foregone revenue*. In the present context, $\hat{\Psi}$ is a random object, since it is constructed from a random sample as in the previous section. Similarly, $\hat{\mathbf{s}}$ is random, and so $F(\mathbf{s}^*) - F(\hat{\mathbf{s}})$ is a non-negative valued random variable. Its expectation $\epsilon = F(\mathbf{s}^*) - E\left[F(\hat{\mathbf{s}})\right]$ may be used as a figure of merit for the estimation process. We will refer to ϵ as the *expected foregone revenue*; this is essentially the same thing as the "expected excess cost" (for minimization problems) in [9]. The following theorem shows that if $\hat{\Psi}$ is the grid estimator constructed based on the hypotheses of theorem 4.4 then the expected excess cost diminishes to zero as the estimator improves.

Theorem 4.5. Let $s_1, s_2, \ldots$ be a sequence of offer stacks such that $\bigcup_{j=1}^{\infty} s_j$ is dense in the bounded set $U = \{(q,p) : 0 < \Psi(q,p) < 1\}$. Suppose Ψ is continuous and strictly monotone on U. Let $\hat{\Psi}$ be the estimator of Ψ derived from a sample of $n = \sum_{j=1}^{m} N_j$ points, with N_j points from stack s_j for each $j = 1, \ldots, m$. Let R be a bounded function of p and q which is increasing in p as well as q. Also let $\mathbf{s}^*$ and $\hat{\mathbf{s}}$ be maximizers of F and $\hat{F}$, as defined above, respectively. Then $F(\mathbf{s}^*) - E\left[F(\hat{\mathbf{s}})\right] \to 0$.

Proof Note that

$$F(\mathbf{s}^*) \;\leq\; \hat{F}(\mathbf{s}^*) + ||F - \hat{F}||_\infty$$

$$\begin{aligned} &\leq \hat{F}(\hat{\mathbf{s}}) + ||F - \hat{F}||_\infty \\ &\leq F(\hat{\mathbf{s}}) + 2||F - \hat{F}||_\infty. \end{aligned} \tag{7}$$

Furthermore,

$$\begin{aligned} ||F - \hat{F}||_\infty &= \sup_{\mathbf{s}} |\int_{\mathbf{s}} R\, d\Psi - \int_{\mathbf{s}} R\, d\hat{\Psi}| \\ &= \sup_{\mathbf{s}} |\int_{\mathbf{s}} R\, d(\Psi - \hat{\Psi})| \end{aligned}$$

Now Ψ and its estimate $\hat{\Psi}$ both take on the value 0 at the origin, which is the start of the curve $\mathbf{s}$, and they evaluate to 1 at the end of $\mathbf{s}$. Using this and integrating by parts we obtain

$$\begin{aligned} \sup_{\mathbf{s}} |\int_{\mathbf{s}} Rd(\Psi - \hat{\Psi})| &= \sup_{\mathbf{s}} |\int_{\mathbf{s}} (\Psi - \hat{\Psi}) dR| \\ &\leq ||\Psi - \hat{\Psi}||_\infty \sup_{\mathbf{s}} \int_{\mathbf{s}} dR \\ &\leq 2||\Psi - \hat{\Psi}||_\infty ||R||_\infty \end{aligned} \tag{8}$$

From (7) and (8) together we derive that

$$F(\mathbf{s}^*) - F(\hat{\mathbf{s}}) \leq 4||R||_\infty ||\Psi - \hat{\Psi}||.$$

Theorem 4.4 states that $\hat{\Psi}$ constructed as described here will uniformly converge to Ψ in probability, therefore we obtain

$$F(\mathbf{s}^*) - F(\hat{\mathbf{s}}) \to 0 \text{ in probability as } m \to \infty.$$

Furthermore, the random variables $F(\hat{\mathbf{s}})$ are bounded as R is bounded. Therefore

$$F(\mathbf{s}^*) - E\left[F(\hat{\mathbf{s}})\right] \to 0 \text{ as } m \to \infty.$$

∎

5 Numerical example: the three-node network

In this section, we use our method to estimate a market distribution function from synthetic (simulated) data. We use an example system which is sufficiently small for all quantities to be computed analytically.

Consider the three-node network shown in Figure 3. The network represents a (very small) power system operated according to a nodal pool power market. In this system, "we" are the only generator large enough to influence prices, and we have the ability to generate power at no marginal cost. A competitive fringe comprising many smaller generators provides additional supply "G1" and "G2"; which is offered to the market via fixed aggregated offer curves. For simplicity, assume that these take the forms $q = \beta_1 p$ and $q = \beta_2 p$ at nodes 1 and 2 respectively, with $\beta_1 = \beta_2 = 1$. Demand D is located at node 3 and is random, with a probability distribution function F; we will use a uniform distribution on $[180, 300]$ for this. There are no line losses, and no line capacity constraints other than the 100 megawatt limit on the line between nodes 2 and 3. The three lines have equal admittances. (The significance of this last point is that 1/3 of any power injected at node 1, and 2/3 of any power injected at node 2, must flow via the limited-capacity line to reach the load.)

The system is simple enough to admit a simple closed-form expression for $\Psi(q,p)$. Indeed, Ψ turns out to be piecewise linear. There are only two regimes: either the transmission constraint between nodes 2 and 3 is binding, or it is not. If it is not, the price p of electricity will be the same at all nodes; this is possible when $(q+\beta_1 p)/3 + 2\beta_2 p/3 < 100$, i.e. $q + 3p < 300$ (where q is our own generation). The total generation from all sources is then $q + (\beta_1 + \beta_2)p$, and so

$$\Psi(q,p) = F(q+2p) = \frac{1}{120}(q+2p-180)_+.$$

If $q + 3p \geq 300$, then the line between nodes 2 and 3 must be at capacity, and there will be a different price at each node. Let p be the price at our node, and p_2 be the price faced by the generators G2. The line constraint requires $(q+\beta_1 p)/3+2\beta_2 p_2/3 = 100$, and so $p_2 = 150 - (q+p)/2$. To maintain $p_2 \geq 0$ requires $q + p \leq 300$. (Note that no solution is possible for $q+p > 300$, since then the generation at node 1 alone would be greater than the transmission network could handle.) The total generation from all sources is $q + \beta_1 p + \beta_2 p_2 = 150 + (q+p)/2$. Hence

$$\Psi(q,p) = F(150 + (q+p)/2) = \frac{1}{240}(q+p-60).$$

To summarize:

$$\Psi(q,p) = \begin{cases} 0 & \text{, if } q+2p \leq 180 \\ (q+2p-180)/120 & \text{, if } q+2p \geq 180 \text{ and } q+3p \leq 300 \text{ (region I)} \\ (q+p-60)/240 & \text{, if } q+3p \geq 300 \text{ and } q+p \leq 300 \text{ (region II)} \\ 1 & \text{, if } q+p \geq 300\ . \end{cases} \tag{9}$$

The regions corresponding to the different linear pieces of Ψ are indicated by the dashed lines in Figure 4.

Suppose we wish to maximize our gross revenue $R(q,p) = qp$. Our optimal offer curve s^* can be found using the technique of [1]. According to that technique, the optimal curve must at each point be either horizontal, vertical, or a solution of

$$\frac{\partial R}{\partial q}\frac{\partial \Psi}{\partial p} - \frac{\partial R}{\partial p}\frac{\partial \Psi}{\partial q} = 0.$$

For this problem, the latter equation reduces to $q = 2p$ in region I, and $q = p$ in region II. This gives us two non-intersecting line segments; to produce an offer curve which is monotone increasing in both q and p, they must be joined by a vertical line segment at some fixed quantity q_0. The value of q_0 should be chosen so as to maximize

$$\int_{s^*} R\, d\Psi = \int_{90}^{q_0} \frac{q^2}{2}\frac{1}{60}\, dq + \int_{q_0/2}^{100-q_0/3} q_0 p \frac{1}{60}\, dp + \int_{100-q_0/3}^{q_0} q_0 p \frac{1}{240}\, dp + \int_{q_0}^{150} q^2 \frac{1}{120}\, dq,$$

the expected gross revenue. Elementary calculus suffices to show that this is achieved for $q_0 = 100$ giving an optimal expected revenue of $10127\frac{7}{9}$. The resulting optimal curve s^* is shown in Figure 4.

We now describe the results of a number of numericalexperiments that investigate the behaviour of our Ψ-estimator when it is computed from
(a) dispatch points simulated from a single "good" stack;
(b) dispatch points simulated from each of six stacks; and
(c) dispatch points simulated from a single "bad" stack.
The stacks used are shown in Figure 5. Note that the "good" stack used for (a) is quite close to optimal, but the "bad" stack used for (c) is rather far from the optimal stack. Both the "good" and the "bad" stacks are among the six used for (b).

In each experiment the following procedure was followed. For the set of stacks corresponding to the experiment a set of dispatch points was calculated by sampling from Ψ defined by (9). In each experiment sample sizes of 60 and 240 dispatch points were chosen. For each sample the non-parametric maximum likelihood procedure was followed to produce an estimate $\hat{\Psi}$ of Ψ. The computation time required to produce $\hat{\Psi}$ increased from a few minutes to over an hour, roughly a 20-fold increase, as the sample size quadrupled.

To assess the quality of the estimator we repeated each experiment 100 times to obtain 100 estimates of Ψ. These estimates are plotted on the graphs shown in Figures 6, 7, 8, 9, 10 and 11. These figures depict the true value of Ψ as shown by the piecewise linear curve, compared with the 100 sampled estimates. (The horizontal axis on these graphs represents the parameter t as $(q(t), p(t))$ moves along the optimal stack from (90, 45) where $\Psi = 0$ to (150, 150) where $\Psi = 1$.)

Figures derived from a sample size of 60 depict very similar results (in terms of bias of the estimators) to that depicted in the sample size of 240. The only thing different is that the spread decreases as the sample size is increased.

The differences between cases (a), (b), and (c) are most obvious in the lower left corners of Figures 6 to 11. In case (c) there is a clear bias, in case (a) the bias is much smaller, and no bias is noticeable in case (b). This suggests that bias reduction may be an advantage of more diverse data.

In each of the experiments we computed the estimate $\hat{\mathbf{s}}$ of the optimal offer stack $\mathbf{s}^*$ by solving (6) using $\hat{\Psi}$ rather than Ψ. Since $\hat{\Psi}$ is piecewise constant on the cells of a rectangular grid, the line integral reduces to a finite sum, with one term arising each time the curve $\mathbf{s}$ crosses a boundary between cells. Furthermore, if R is increasing in both components (as is usually the case, e.g. $R(q,p) = qp$ here), then the optimal curve will have each such boundary crossing as near as possible to the upper or right-hand end of the edge. (Note that the optimal value of (6) may not actually be attained, but must be understood as the supremum of all possible values.) These considerations reduce (6) to a simple, finite problem.

Figure 12 shows estimates of the optimal offer stack $\hat{\mathbf{s}}$ (for clarity we show only the first 25 of these). Note that for experiments (a) and (c), where the dispatches come from a single offer stack, the estimated optimal offer stacks seem to have almost converged even with a sample size of 60. However the limit to which they converge is not necessarily the optimal offer stack. This is particularly the case for (c), where the limit is quite far from the optimal offer stack. In (b), convergence is slower, but

Table 1: Average expected returns of estimated offer stacks.

	sample size 60	*sample size 240*
(a). *"good" stack*	9291.63	9184.07
(b).*six stacks*	9743.01	9874.59
(c). *"bad" stack*	7542.69	7718.66

the limit is closer to the optimal offer stack.

The performance of each $\hat{\mathbf{s}}$ was then evaluated under the true market distribution function to give an expected revenue $F(\hat{s})$. Table 1 provides the sample average of these values over the 100 experiments. The difference between these values and the optimal expected revenue is an estimate of the expected forgone revenue. For (c), the performance of the estimated stacks is poor, and Figure 12 suggests that increasing the sample size is not likely to improve the returns significantly. This is consistent with the theoretical results - the increased sample size improves the resulting estimator $\hat{\Psi}$, but only on the offer stacks that the dispatch points came from. The estimator here gets closer to the true Ψ on the single "bad" stack, but remains far from the true Ψ everywhere else, including in the vicinity of $\mathbf{s}^*$.

The returns from (a) are better, but Figure 12 again indicates that increasing the sample size will not result in further improvement. The six stacks in (b) produce the best returns. This illustrates the value of data diversity - sampling from many parts of the plane produces better results than the same number of observations confined to a single curve. Here Figure 12 suggests that increasing the sample size might produce better results still.

6 Future directions

The results above demonstrate that in order to improve on a "current" offer stack, it is necessary to know the value of Ψ not only on that stack, but also elsewhere in the plane. A relevant question would be how much revenue would be lost (or gained) by submitting a variety of offers in order to garner more information on the market distribution function. This is a "multi-armed bandit" problem, of the kind described in [4]. The loss or gain must be defined relative to the strategy of repeatedly submitting the same "current" stack, and so depends very much on how good that stack was to begin with. If the original stack was close to optimal, then it is likely that a loss will be incurred by varying it. Furthermore, the eventual gains resulting from a better knowledge of Ψ are likely to be small. If, on the other hand, the original stack was poor, then it is likely that experimenting with other stacks will result in an immediate gain, with further gains in prospect once Ψ is better understood.

In a real electricity market, a generator must be thought of as facing not a single Ψ function, but many Ψ functions corresponding to different times of day, hydrological conditions, etc. Thus the generator may have available data relating to Ψ functions other than – but perhaps similar to – the one being estimated. A valuable extension of the techniques presented in this paper would be to make use of the additional data in some way.

Market distribution functions, as described in this paper, are intended to apply to a single-period market clearing. They are not applicable to market dynamics in a repeated gaming situation. An extension to such a situation would have to consider how Ψ might vary in future periods in response to offers made now.

7 Acknowledgments

The authors would like to thank the anonymous referees and the associate editor for their valuable comments on the content and presentation of this paper.

References

[1] E.J. Anderson and A.B. Philpott, Optimal offer construction in electricity markets, *Math. of OR* 27 (2002), 82–100.

[2] E.J. Anderson and A.B. Philpott, Estimation of electricity market distribution functions, *Annals of Operations Research* 121 (2003), 21–32.

[3] E.J. Anderson and H. Xu, ϵ-Optimal Bidding in an Electricity Market with Discontinuous Market Distribution Function, Working Paper, Australian Graduate School of Management, June 2004.

[4] D. A. Berry and B. Fristedt *Bandit problems: sequential allocation of experiments* (1985), Chapman/Hall.

[5] Chao, H-P. and Huntington, H.G. *Designing Competitive Electricity Markets* (1998), Kluwer.

[6] R. Durrett, *Probability, theory and examples* 2nd ed.(1996), Duxbury Press.

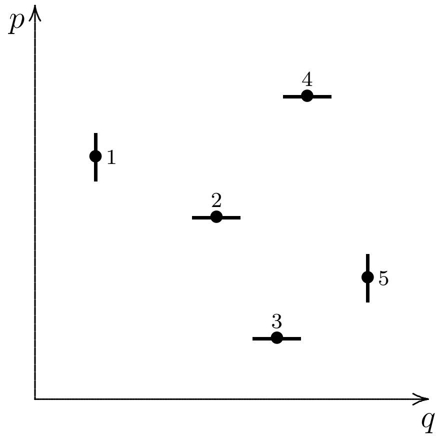
[7] M. Kendall and A. Stuart, *The advanced theory of statistics* 4th ed. vol. 2 (1979), Macmillan.

[8] A. Philpott and G. Pritchard, Financial transmission rights and market power, forthcoming in *Decision Support Systems.*

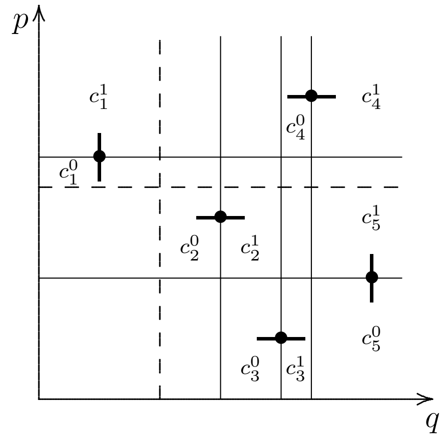
[9] G. Pritchard and G. Zakeri, Stochastic optimization: excess cost and importance sampling, in: Proceedings of 4th Hungarian colloquium on limit theorems in probability and statistics (1999), Birkhäuser.

[10] A. Shapiro and J.F. Bonnans, *Perturbation analysis of optimization problems* (2000), Springer-Verlag.

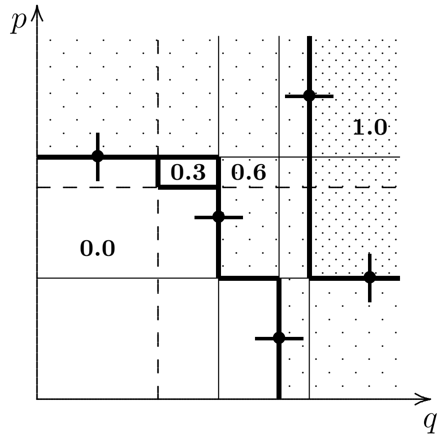
[11] A. Shapiro, *Monte Carlo sampling methods,* in A. Rusczynski and A. Shapiro (editors) Stochastic Programming, volume 10 of Handbooks in Operations Research and Management Science (2003), North Holland.



1. Original data



2. Subdivision into cells



3. Cell values giving $\hat{\Psi}$

Figure 1: Example showing the construction of the (enhanced) grid estimator from a sample of five points of dispatch. First, the plane is subdivided into rectangular cells (the extra lines of the enhanced estimator are shown dashed). Then cell values are determined (the contours of $\hat{\Psi}$ are shown as heavy lines). In this example, only one cell's value is not completely determined by (4); it has been given a value of 0.3 in accordance with the convention described in the text, although any value in $[0, 0.6]$ would give the maximum likelihood.

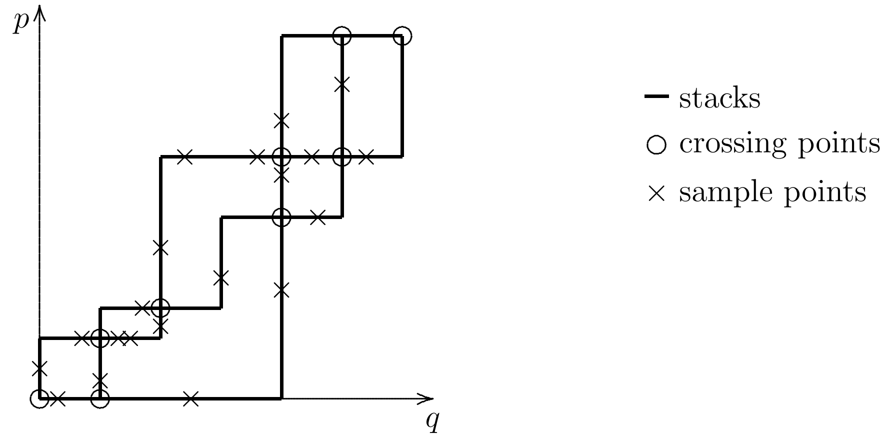


Figure 2: A typical example with $m = 3$ distinct stacks and a total of $n = 20$ dispatch points in the sample.

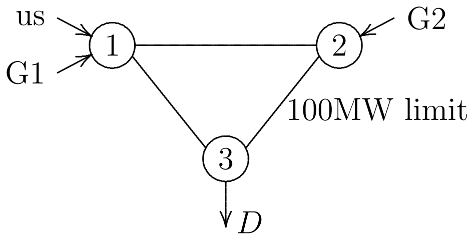


Figure 3: A three-node network.

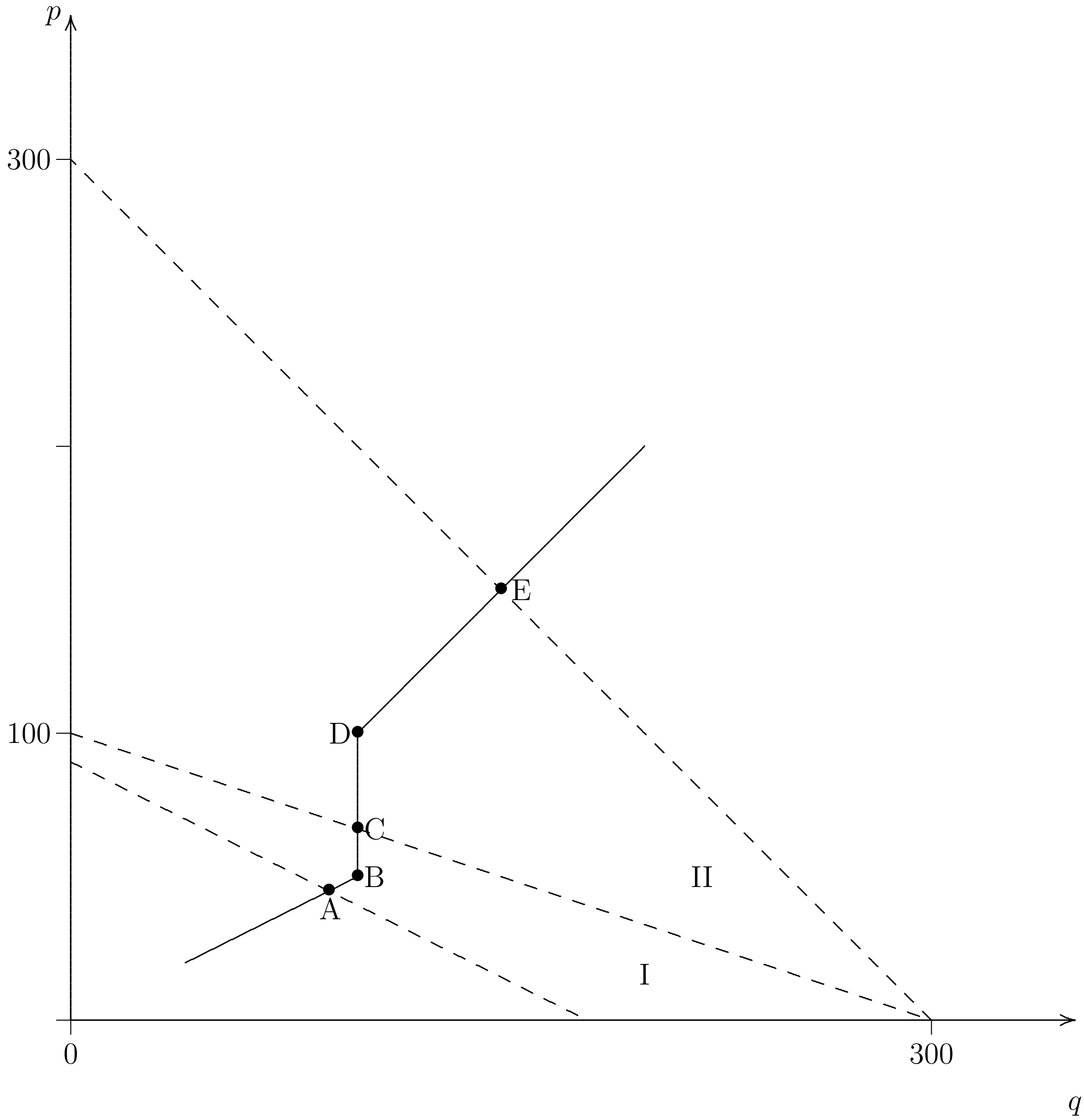
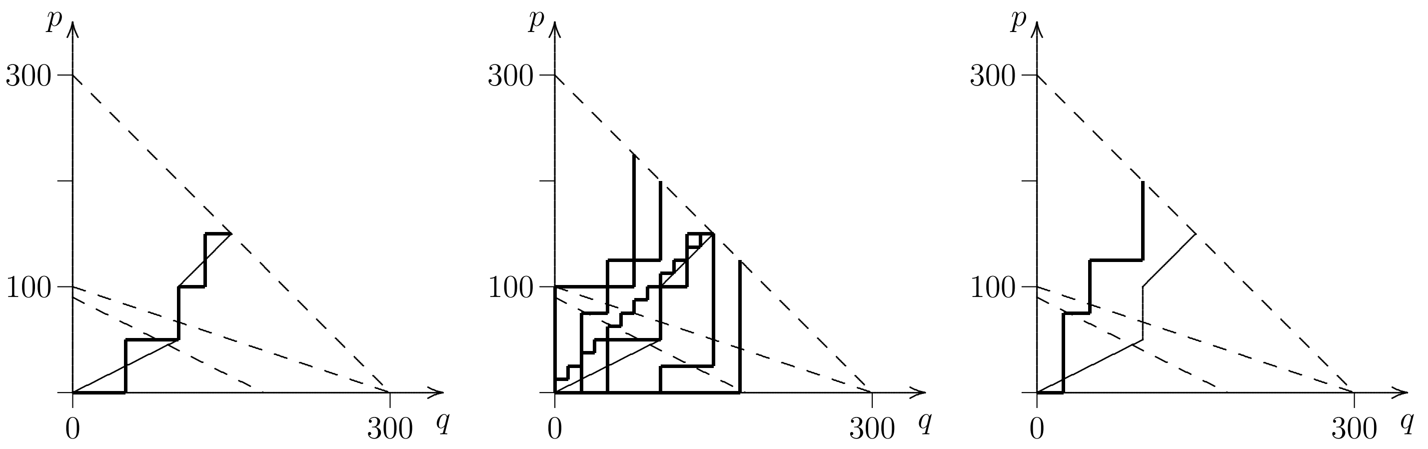


Figure 4: The (q, p) plane for the three-node network problem, with optimal offer curve shown.



(a) One "good" stack (b) Six stacks (c) One "bad" stack

Figure 5: Stacks used for obtaining simulated data in the three-node network.

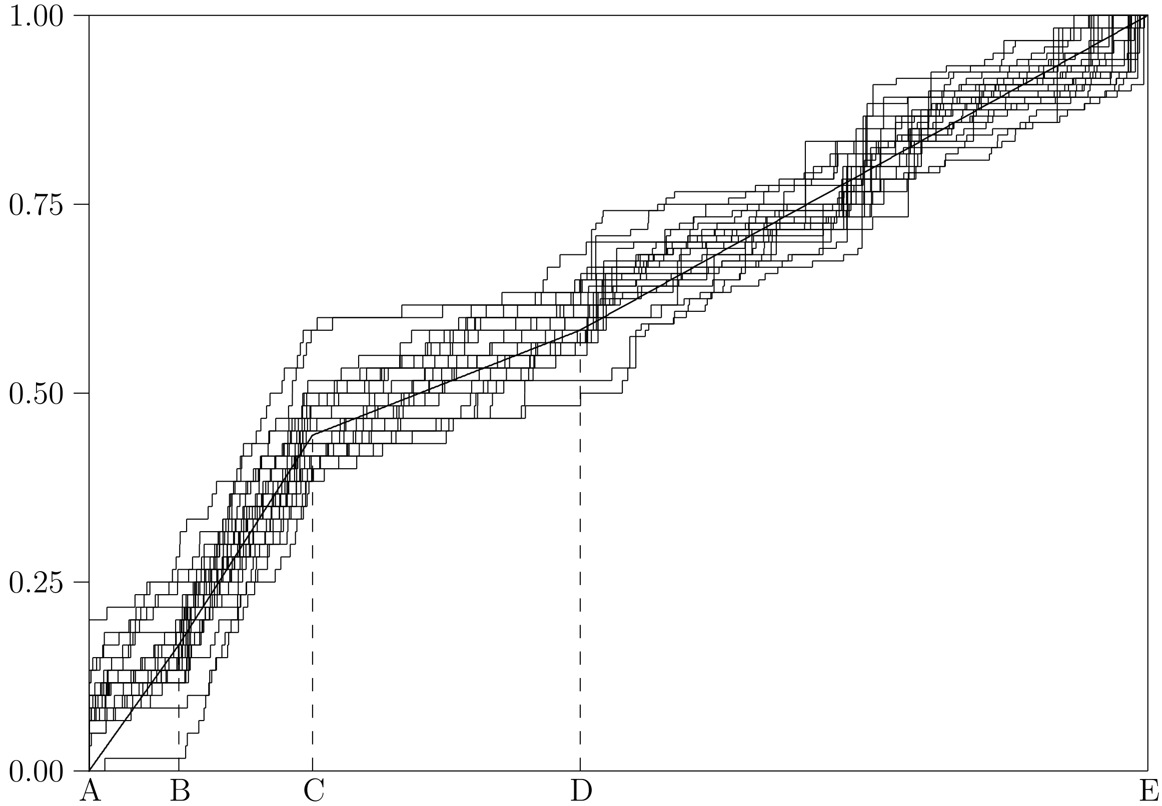


Figure 6: Values of the estimated Ψ computed from 60 dispatch points on the single "good" stack.

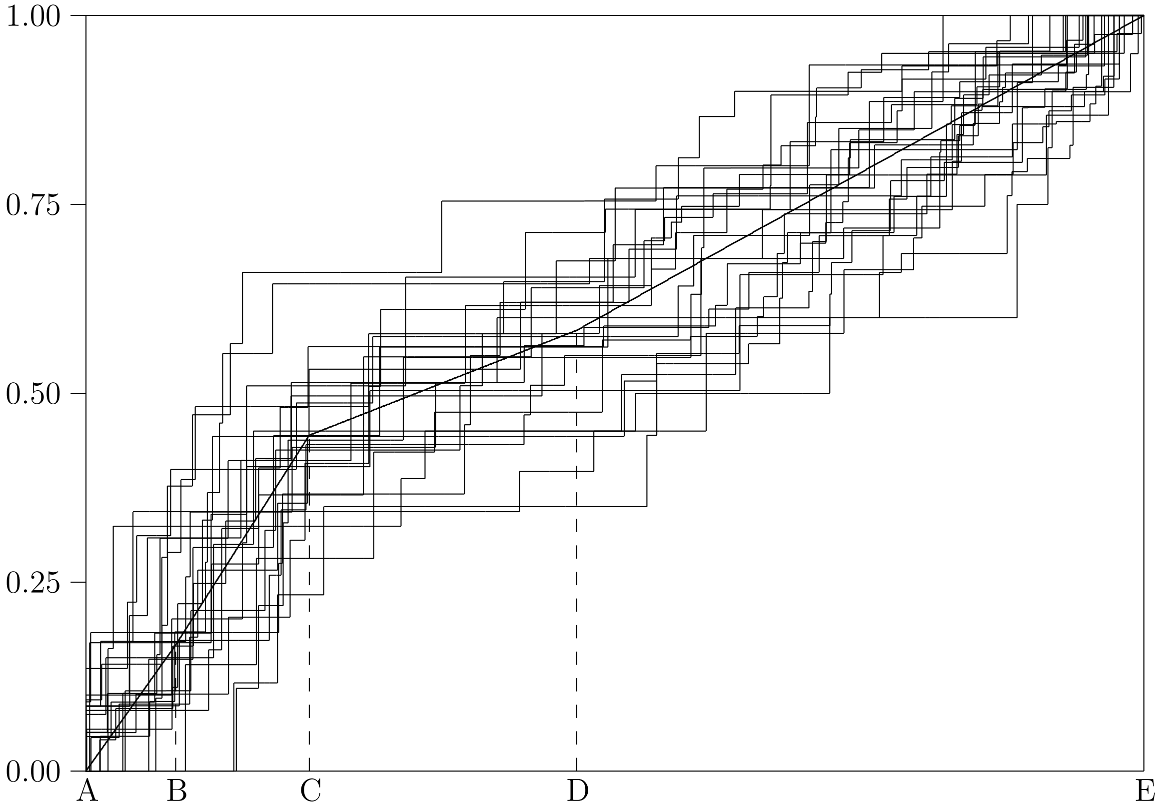


Figure 7: Values of the estimated Ψ computed from 60 dispatch points on all six stacks.

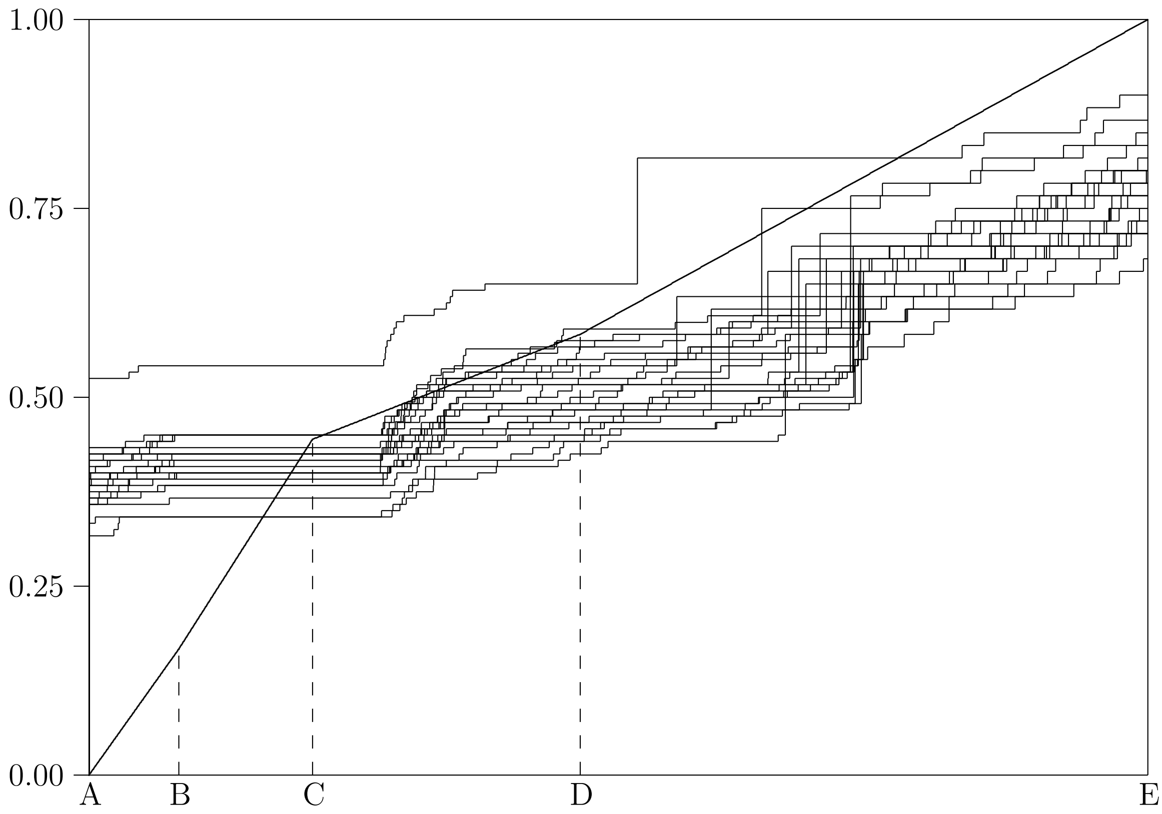


Figure 8: Values of the estimated Ψ computed from 60 dispatch points on the single "bad" stack.

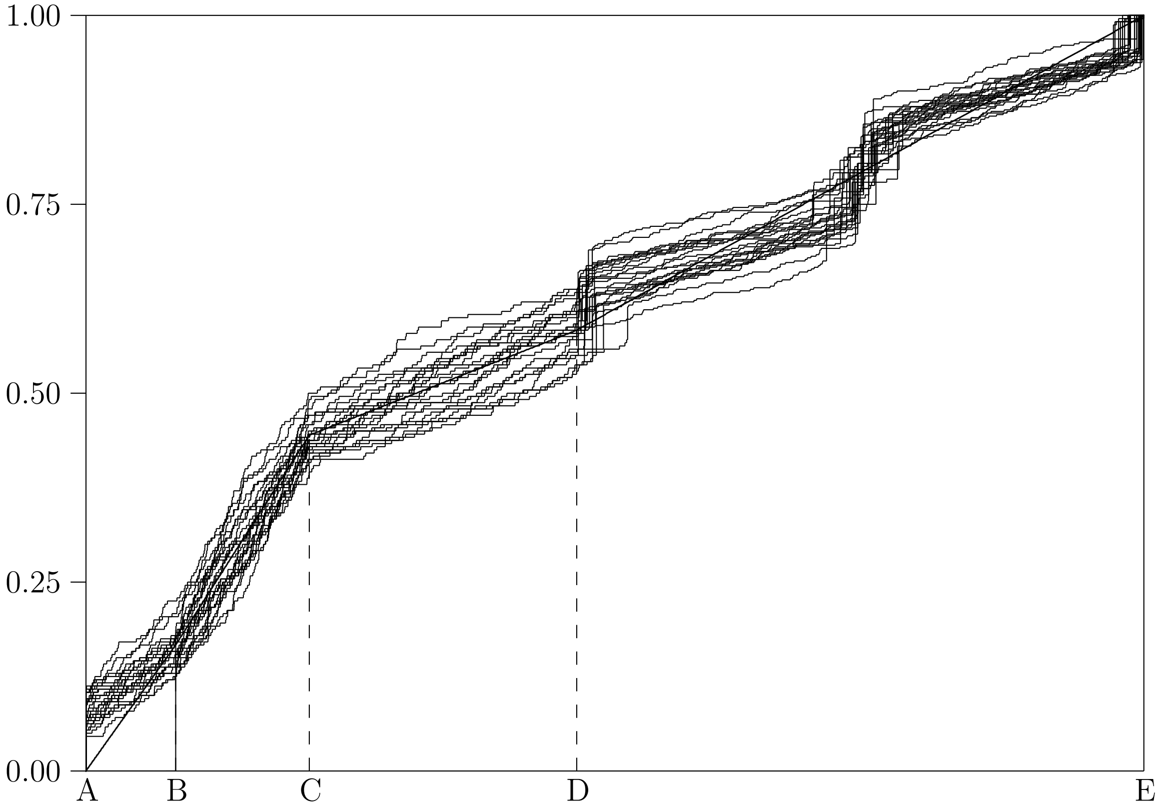


Figure 9: Values of the estimated Ψ computed from 240 dispatch points on the single "good" stack.

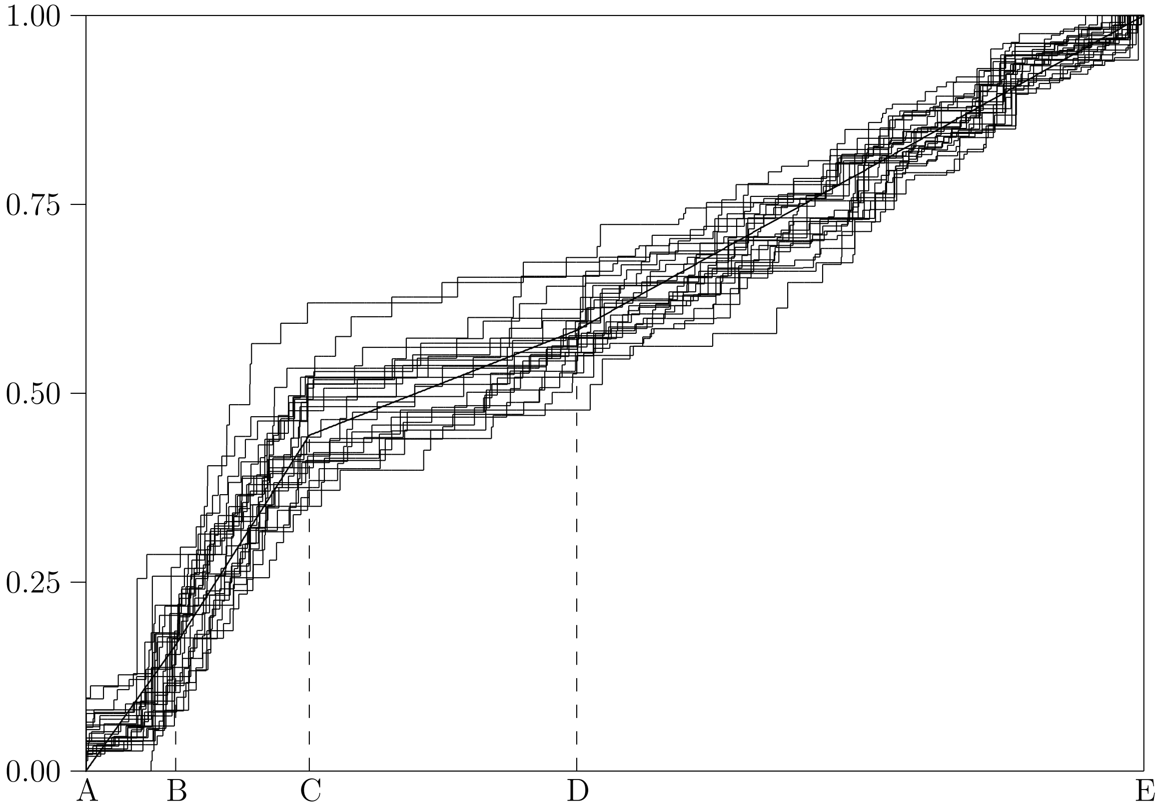


Figure 10: Values of the estimated Ψ computed from 240 dispatch points on all six stacks.

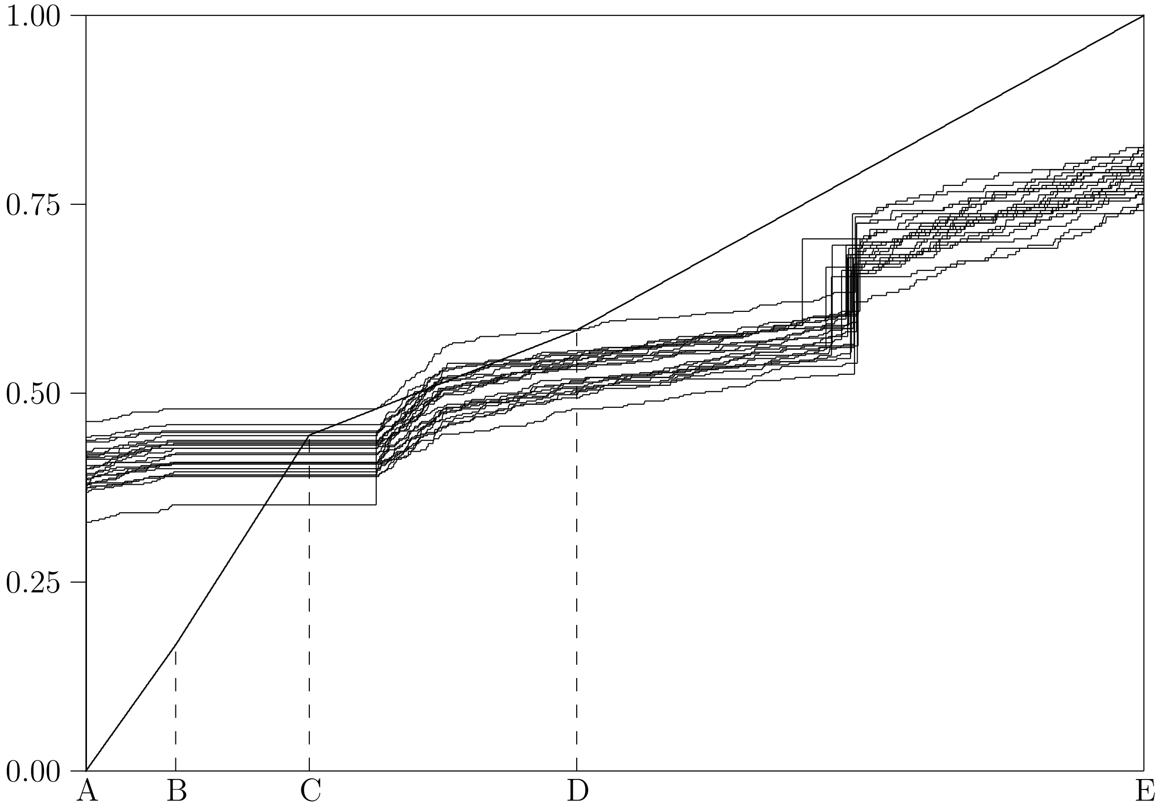
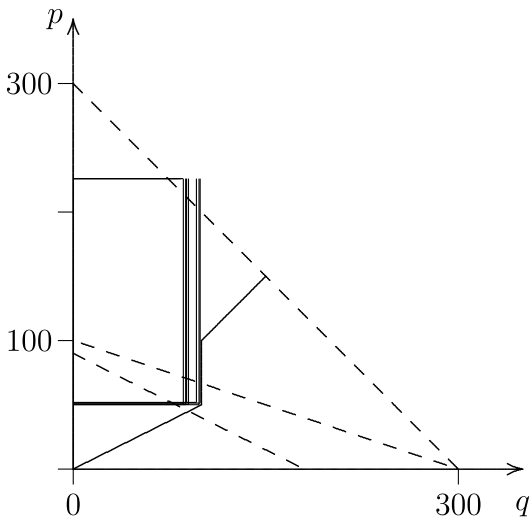
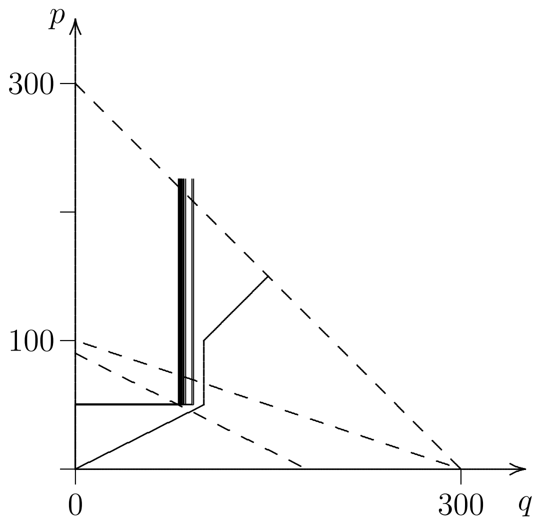


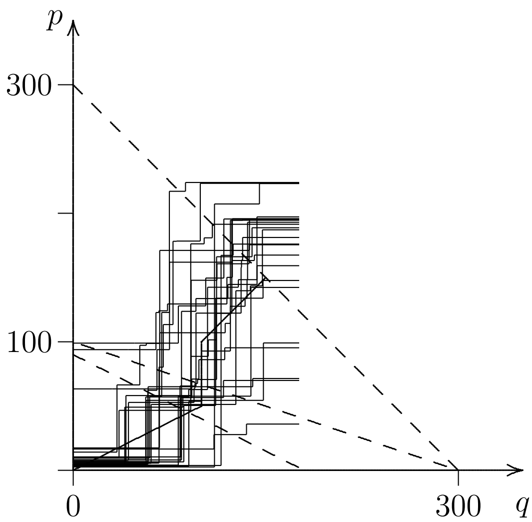
Figure 11: Values of the estimated Ψ computed from 240 dispatch points on the single "bad" stack.



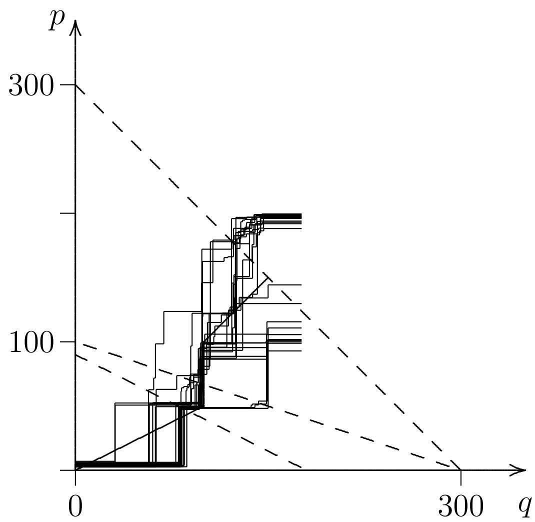
sample size 60, from one "good" stack



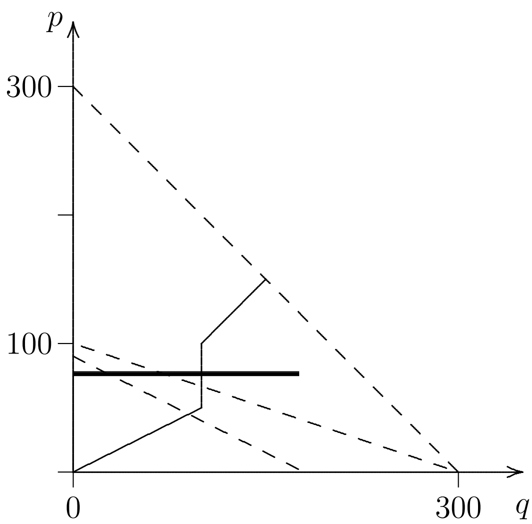
sample size 240, from one "good" stack



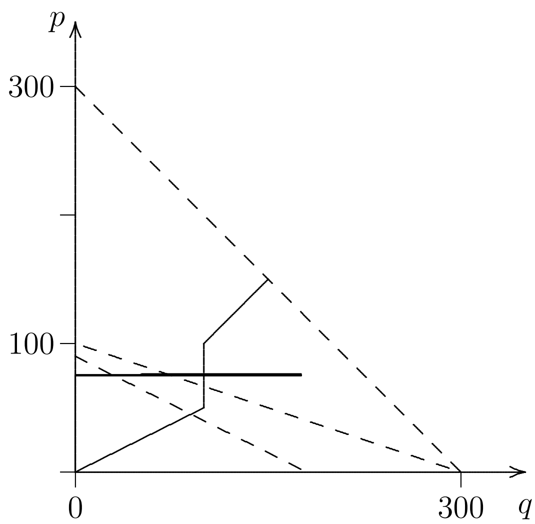
sample size 60, from six stacks



sample size 240, from six stacks



sample size 60, from one "bad" stack



sample size 240, from one "bad" stack

Figure 12: Estimated optimal offer stacks