COMPUTATION OF NONPARAMETRIC CONCAVITY-RESTRICTED ESTIMATORS

Rosa L. Matzkin*
Department of Economics
Northwestern University
Evanston, IL 60208

October 1999

Abstract

This paper introduces a computation method that can be used to estimate various nonparametric concavity-restricted functions. In particular, the method can be used to calculate estimators of utility, production, profit, and cost functions that do not impose any parametric structure on these functions.

The new computation method requires less computer resources, in terms of computer memory and/or CPU time, than required when using standard algorithms. The method is used in a small experiment with simulated data to evaluate the performance of a shape-restricted estimator.

---

* The support of NSF is gratefully acknowledged. I have benefited from my interaction with Vassilis Hajivassiliou, Kurt Anstreicher, Emil Moffa, and the participants in the several seminars and conferences where this paper was presented.

1 Introduction

This paper presents a computation method to calculate several nonparametric, concavity-restricted estimators. The method requires less computer resources than some standard algorithms. Using the new technique, one can, for example, estimate utility, production, profit, and cost functions, without imposing any parametric structure on these functions.

The nonparametric estimators to which the computation method presented in this paper can be applied use properties, such as concavity, monotonicity, and homogeneity of degree one, about the shape of the nonparametric functions in the models. These properties are used to define restrictions on the set of possible values that the nonparametric functions may attain. The restrictions are imposed by means of a large system of inequality constraints on the values and subgradients of the nonparametric functions. Each inequality in the system constraints the values and subgradients of a nonparametric function at two points of its domain. These inequalities are similar in spirit to the revealed-preference type inequalities that were first introduced by Samuelson (1938) and Houthakker (1950) and that have more recently been studied by Afriat (1967a,1967b, 1972, 1973, 1981)), Chiappori and Rochet (1987), Diewert (1972), Diewert and Parkan (1985), Matzkin and Richter (1991), and Varian (1982, 1983).

Examples of nonparametric estimators that use shape restrictions on the functions being estimated and whose computation can be performed using the methods described in this paper are Matzkin (1990, 1991a, 1991b, 1992, 1993, 1994). Other nonparametric estimation and testing methods that also use shape restrictions and whose computation can be performed using modifications of the method presented in this paper are Brown and Matzkin (1990, 1995), Epstein and Yatchew (1985), and Varian (1985, 1988).

In the methods developed in Matzkin (1990, 1991a, 1991b, 1992, 1993, 1994), estimators of nonparametric concave functions are obtained by a two-step procedure. In the first step, a criterion function is maximized over the values and subgradients that a concave function attains at a finite number of points. In the second step, an estimator for the nonparametric concave function is obtained by interpolating between the optimal values and subgradients obtained in the firststep maximization. The maximization problem in the first step of the procedure is a large-scale constrained optimization problem. The objective of this paper is to present a method to find a solution to large-scale problems of the type that arise in such estimation problems.

The statistical methods of Brown and Matzkin (1990, 1995), Epstein and Yatchew (1985), and Varian (1985, 1988) also use two-step computation procedures where the first step involves a maximization problem over the values and subgradients of concave functions. So, with appropriate modifications, the methods described in this paper can be applied also to those methods.

The computation method presented in this paper can be used to solve optimization problems whose criterion functions are either continuous or step functions. The values of the step functions are required to depend only on the ordering of the values of their arguments. The constraints of each particular optimization problem are determined by the shape restrictions that the estimation method imposes on the nonparametric functions being estimated. Typically, when the number of observations is N and the domain of the estimated nonparametric function is a subset of R^K, the optimization problem involves at least NK variables and N^2 constraints.

The computation method is based on a random search procedure that exploits the particular properties of the objective function and constraints of the optimization problems that arise when the estimated nonparametric functions is required to be concave. In particular, the method exploits the convexity of the constraint set, the sparsity of the matrix defining the constraints, and the fact that the constraint set can be characterized using functions that are the minimum of a finite number of linear functions.

The paper includes results of a small experiment that has been performed using the new computation method. The experiment shows that a nonparametric shape-restricted estimator for the binary threshold-crossing model may perform better than parametric and semiparametric counterparts, when the data size is as small as 100 observations.

In the special case where the shape-restricted function that one wishes to estimate is a conditional expectation function, one can use, besides the method presented in this paper, the methods presented in Goldman and Ruud (1992) and Ruud (1995) (see also Hildreth (1954) and Dykstra (1983)).

The outline of the paper is as follows. In the next section, we describe the typical optimization problem that this paper is concerned with. In Section 3, we present the computation method and establish some of its convergence properties. A small experiment performed using the new computation method is described in Section 4. The main conclusions of the paper are summarized in Section 5. The proofs of the lemmas and theorems are presented in the Appendix.

2 The optimization problem

The optimization problems concerning this paper are of the following form:

$$
\begin{aligned}
(1)\quad & \underset{\{h^i\}\,\{T^i\}}{Maximize}\ L(h^1, ..., h^N) \\
& \quad \textit{subject to} \\
& \quad (R.1)\ h^i - h^j - T^j \cdot (x^i - x^j) \leq 0 \qquad ij = 1, ..., N \\
& \quad (R,2)\ CY \ \leq \ c
\end{aligned}
$$

where $Y = (h^1, ..., h^N; T^1, ..., T^N)$, $h^i \in R$, $T^i \in R^K$, $x^i \in R^K$, C is an $M \times (N + NK)$ matrix, and c is an $M \times 1$ vector. The values of $h^1, ..., h^N$ are interpreted as the values of a function h at observed vectors $x^1, ..., x^N$ in the domain of h. The values of $T^1, ..., T^N$ are interpreted as the subgradients of h at $x^1, ..., x^N$. A typical case where this problem arises is when calculating an estimator, $\widehat{h}$, for a function h^* that is concave and satisfies also other properties. Such an estimator is obtained by interpolating between the values of a solution, Y^*, to the above problem. The interpolated function has to satisfy the properties that h^* is assumed to satisfy. The first set of inequality constraints, $(R.1)$, imposes restrictions on Y guaranteeing that the values of $h^1, ..., h^N$ and $T^1, ..., T^N$ correspond to the values and subgradients of a concave function. This assures the existence of an interpolation that is concave. The second set of inequality constraints, $(R.2)$, imposes restrictions guaranteeing that the resulting interpolation satisfies the other additional properties that h^* is assumed to possess. For example, if an estimator for h^* is required to be monotone increasing, in addition to being concave, then the set of inequalities $(R.2)$ will take the form

$$T^i \geq 0 \qquad i = 1, ..., N$$

If the estimator for h^* is required to be homogenous of degree one, the constraints in $(R.2)$ will include the following inequalities:

$$T^i \cdot x^i = h^i \qquad i = 1, ..., N$$

To guarantee that the estimator for h^* attains a given value, α, at some point x^*, one can define $x^{N+1} = x^*$, redefine Y to be $(h^1, ..., h^N, h^{N+1}; T^1, ..., T^N, T^{N+1})$, and include the following constraints in $(R.2)$:

$$h^{N+1} = \alpha,$$

$$h^i - h^{N+1} - T^{N+1} \cdot (x^i - x^{N+1}) \leq 0 \qquad i = 1, ..., N$$

$$h^{N+1} - h^j - T^j \cdot (x^{N+1} - x^j) \leq 0 \qquad j = 1, ..., N$$

(See the papers mentioned in the introduction to see how to impose other shape restrictions on estimators for h^*.) Note that in all the cases considered above, each inequality imposes restrictions on at most two coordinates of $(h^1, ..., h^N, h^{N+1})$ and two coordinates of $(T^1, ..., T^N, T^{N+1})$.

The objective function, L, in (1) can be either continuous or a step function whose value changes only when the order of the values of $(h^1, ..., h^N, h^{N+1})$ change.

2.1 An example where the typical optimization problem arises

To provide an example of an estimation method where the optimization problem described above arises, consider the problem of nonparametrically estimating the cost function a typical firm faces when it undertakes a particular project. Suppose that the only available data relate to whether or not the firm decides to undertake the project and the price vector faced by the firm for the inputs required to perform the project. The firm is assumed to undertake the project if the revenue it yields exceeds its cost. The econometrician does not observe the potential revenue of the project but knows that it is distributed independently of the vector of input prices. Matzkin (1992) introduces a method of strongly consistently estimating the cost function of the typical firm in this model. The method does not require any parametric assumptions about either the cost function or the distribution of revenue. It is only assumed, for normalization purposes, that the cost is known at one particular vector of input prices.

Let $(x^1, ..., x^N)$ denote the observed vectors of input prices faced by N sampled firms. And let $(h_1^*, ..., h_N^*)$ and $(T_1^*, ..., T_N^*)$ denote, respectively, the unobservable values and subgradients at $(x^1, ..., x^N)$ of a cost function h^*. Then, the estimator for the cost function is obtained by a two-step procedure. In the first step, a step function is maximized over the values $(h^1, ..., h^N)$ and subgradients $(T^1, ..., T^N)$ of the cost function of the typical firm at the observed vectors $(x^1, ..., x^N)$. To guarantee that the vectors $(h^1, ..., h^N)$ and $(T^1, ..., T^N)$ correspond to the vectors of values and subgradients of a function, h, that is a cost function, these vectors are constrained to satisfy concavity, monotonicity, and homogeneity of degree one constraints. In the second step, one obtains a function that interpolates between the optimal values of $(h^1, ..., h^N)$ and $(T^1, ..., T^N)$ calculated in the first step. This interpolation is a concave, monotone, and homogeneous of degree one function.

Let $x^0 = 0$ and $h^0 = 0$. To normalize the value of the estimated function, it is assumed that its value is known at one point. Let x^{N+1} be any given (either observed or unobserved) vector of input prices. Assume that the value of the cost function at x^{N+1} is a known value, denoted by h^{N+1}. Let T^{N+1} denote the subgradient of the cost function at x^{N+1}. Then, the constraints on $(h^1, ..., h^N)$ and $(T^0, ..., T^{N+1})$ derived from the homogeneity of degree one, concavity, and monotonicity restrictions on h are

(2) $h^i = T^i \cdot x^i \qquad i = 0, 1, ..., N+1,$

(3) $h^i \leq T^j \cdot x^i \qquad i, j = 0, ..., N+1,$ and

(4) $T^i \geq 0 \qquad i = 0, ..., N+1.$

The optimization problem in the first step of the estimation of h^* is then the following:

$$(5) \quad \underset{\{h^i\}\,\{T^i\}}{Maximize} \; L(h^1, ..., h^N)$$

$$subject\ to$$

$$h^i = T^i \cdot x^i \qquad i = 0, 1, ..., N+1,$$

$$h^i \leq T^j \cdot x^i \qquad i, j = 0, ..., N+1, \text{ and}$$

$$T^i \geq 0 \qquad i = 0, ..., N+1.$$

The objective function in (5) is the optimal value of the following optimization problem

$$(6) \quad \underset{\{F^i\}}{Maximize} \; \sum_{i=1}^{N} \{y^i \ln(F^i) + (1y^i)\ln(1F^i)\}$$

$$subject\ to$$

$$F^i \leq F^j \quad if\ h^i < h^j \qquad i, j = 1, ..., N,$$

$$F^i = F^j \quad if\ h^i = h^j \qquad i, j = 1, ..., N, \text{ and}$$

$$0 \leq F^i \leq 1 \qquad i = 1, ..., N.$$

where each y^i is an observable dichotomous variable that equals 0 if the ith sampled firm undertakes the project and equals 1 otherwise. The maximization in (6) is performed for given values of $h^1, ..., h^N$ over the real variables $F^1, ..., F^N$. An algorithm to calculate the value of the optimal value function, $L(h^1, ..., h^N)$, of (6) has been developed by Asher et al. (1955). The optimal value function $L(\cdot)$ is a step function, whose value changes only if the ordering between the values of $h^1, ..., h^N$ changes.

To obtain in the second step an estimator, $\widehat{h}$, for the cost function, one can use the following interpolation:

$$\widehat{h}(x) = \min\{\widehat{T^i} \cdot x^i |\ i = 0, 1, ..., N+1\}$$

where$\widehat{T^0}$,...,$\widehat{T^{N+1}}$ are the optimal values of T^0,...,T^{N+1}, which were obtained in the first step. This is a function that is homogeneous of degree one, concave, and monotone increasing. Figure 1 presents a typical graph of the resulting function when $x \in R^2$.

In this example, the first step optimization problem of the cost function falls into the category of the problems described by (1). In this case, the value of the objective function of the optimization problem depends on the values $h^1, ..., h^N$ of the unobserved cost function at the finite number of points $x^1, ..., x^N$. The constraints of the optimization problem are determined by concavity, homogeneity of degree one, and monotonicity restrictions.

Using the equality constraints in (5) corresponding to the homogeneity of degree one restrictions, one can eliminate $h^1, ..., h^N$. So, the problem in (5) can be rewritten as

$$\begin{array}{lll} (7) & \underset{\{T^i\}}{Maximize}\ L(T^1, ..., T^N) & \\ & \quad subject\ to & \\ & T^i \cdot x^i \leq T^j \cdot x^i & i, j = 0, ..., N+1, \text{ and} \\ & T^i \geq 0 & i = 0, ..., N+1. \end{array}$$

This is a problem with $(N+2)K$ variables and $(N+2)^2 + (N+2)K$ constraints. Suppose, for example, that the number of observations, N, is 100, and the number of input prices is 2. In this case, the problem in (7) requires maximization of a step function over a vector of 204 coordinates subject to 10,608 constraints.

2.2 Some alternative algorithms

If the objective function were differentiable, one could use, for example, a Quasi-Newton method to find a solution. These methods assign a Lagrange multiplier to each of the constraints and find the values of the variables and the Lagrange multipliers that satisfy the Kuhn-Tucker conditions of the problem. In an optimization problem such as the one described by (7), with 100 observations, this method requires manipulating a matrix of order 10,812 x 10,812, which requires a considerable amount of computer memory.

When the objective function is discontinuous, random search methods are typically used to find a solution. To find a solution x^* to the maximization of a function $L(x)$ by a standard random search method, one can proceed as follows: First, starting from an initial vector x, randomly draw another vector x'. Second, find a point x" in the line connecting x and x' such that $L(x$") is larger than the value of L at any other point along the line. If $L(x") > L(x)$, take x" as a new starting point, return to the first step, and repeat the procedure. If $L(x") > L(x)$, take x as the starting point, return to the first step, and repeat the procedure. The algorithm stops when in a consecutive sequence of passes through step 1 the initial point is the same. The last starting point is then determined to be the solution. The memory requirements of this method are relatively small. When the number of coordinates of x is large, however, this method may require a considerable amount of CPU time, because randomly generating a number is relatively expensive in terms of CPU time.

It is easy to modify the standard random search algorithm to find a solution to a constrained optimization problem. In the second step of the random search algorithm described above, when performing the line search, one needs to consider only the points along the line that satisfy the constraints. But when the number of constraints is large, this modification may require a large amount of CPU time. There are two main reasons for this. First, when the number of constraints is large, the constraint set is small. Therefore, the probability that any given line will intersect the constraint set is small, and the majority of line searches may be performed over directions that are outside the constraint set. Second, when the number of constraints is large, checking whether each point satisfies all the constraints will be very expensive in terms of CPU time.

We could consider using other methods such as simulated annealing and genetic algorithms, which are more sophisticated random search type methods, or the Nelder and Mead (1965) method, which is a nonrandom, derivative-free method. But, usage of these methods would require evaluating the constraint functions a very large number of times. Hence, they are expensive in terms of CPU time.

3 A modified random search method

In this section, we describe a random search method that can be used to find a solution to optimization problems of the sort described in (1). The method can be used when the objective function of the large-scale constrained optimization problem is either a continuous function or a step function.

The new algorithm involves maximization of the objective function along randomly determined directions, as does the standard random search method mentioned in the previous section. But, unlike that standard random search method, the algorithm searches along directions that are inside the constraint set, and it does not always determine each direction of search by randomly drawing values for all the variables. The improvements are effected by exploiting particular properties of maximization problems described by (1).

3.1 Useful properties of the typical optimization problem

The typical optimization problem described by (1) satisfies the following properties:

(i) The constraint set is convex.

(ii) The matrices defining the constraint set are sparse.

(iii) The constraint set consists of points whose coordinates are the values and subgradients ofparametric concave functions.

(iv) The constraint set can be characterized by a set of vectors whose coordinates are the values and subgradients of functions that are the minimum of linear functions.

The constraint sets of optimization problems described by (1) are convex because they are determined by a finite number of linear inequality constraints. The matrices defining these constraint sets are sparse because most entries in each of their rows are zero. The only nonzero values are those that correspond to at most two coordinates of $(h^1, ..., h^N)$ and at most two coordinates of $(T^1, ..., T^N)$. Clearly, points determined by the values and subgradients of parametric functions satisfying all the shape restrictions that are used to define the constraint set are points in the constraint sets. Finally, as the next lemma establishes, the set of all points $(h^1, ..., h^N; T^1, ..., T^N)$ that satisfy the constraints in $(R.1)$ is the set of points determined by the values and subgradients of functions that are the minimum of linear functions. Hence, when these linear functions satisfy the restrictions defined by $(R.2)$, the points determined by the minimum of them are exactly the set of points in the constraint set.

LEMMA 1: The set of all vectors $(h^1, ..., h^N; T^1, ..., T^N)$ that satisfy the set of constraints in $(R.1)$ is the set of all vectors $(h^1, ..., h^N; T^1, ..., T^N)$ for which there exists a function, h, that is the minimum of N linear functions and it is such that the values and subgradients of h at $x^1, ..., x^N$ are respectively $h^1, ..., h^N$ and $T^1, ..., T^N$.

PROOF OF LEMMA 1: (See also the proof of Lemma 1 in Matzkin (1991).) Suppose that $(h^1, ..., h^N; T^1, ..., T^N)$ satisfies $(R.1)$. For each i, define the function v_i by

$v_i(x) = h^i + T^i \cdot (x - x^i)$.

Define the function h by

$h(x) = \min_i\{v_i(x)\}$.

Then, since $(h^1, ..., h^N; T^1, ..., T^N)$ satisfies $(R.1)$, it follows that for each i, $h(x^i) = v_i(x^i) = h^i$ and T^i is a subgradient of h at x^i. Hence, h is the minimum of N linear functions and it is such that the values and subgradients of h at $x^1, ..., x^N$ are respectively $h^1, ..., h^N$ and $T^1, ..., T^N$.

Conversely, if h is a function that is the minimum of linear functions, then h is concave. It follows that its values and subgradients at any N points satisfy $(R.1)$.

3.2 A method that exploits that exploits the particular properties of the typical optimization problem

We next describe how the above properties can be used to search for the optimal values of the optimization problems that possess those properties, in a more efficient way than using standard algorithms.

First, the convexity of the constraint set guarantees that whenever two points are in the constraint set, the segment connecting them is also in the constraint set. Hence, when searching along such a segment, there is no need of checking whether the constraints are satisfied.

Second, the sparsity of the constraint set can be used to determine in a relatively fast way the boundary points of any segment that is parallel to any one of the axes. Since the matrices defining the constraint set are sparse, any coordinate of the vector $Y = (h^1, ..., h^N, T^1, ..., T^N)$ enters into relatively a small number of constraints. Hence, starting from a point that is inside the constraint set and leaving all but the value of one coordinate variable, one can determine what are the maximum and minimum feasible values for that particular coordinate by considering only a few number of constraints. For example, in (7), any T^i enters into only $N+3$ of the $(N+2)2+(N+2)K$ constraints. Hence, to determine the maximum and minimum feasible values of any coordinate of T^i one only needs to look at $N+3$ constraints.

In the two paragraphs above, we have described how one can take advantage of the convexity and sparsity of the constraint set to determine segments that are inside the constraint set. These segments are determined using one or two points that are known to belong to the constraint set. Properties (iii) and (iv) can be used to generate such points.

To generate a point in the constraint set using parametric functions, one needs to find one or several families of parametric functions in which it is easy to impose the desired shape restrictions by imposing restrictions on the values of their parameters. Any function in such families can be selected randomly by drawing the values of its parameters within this restricted set. Hence, since the selected function satisfies the desired shape restrictions, the vector whose coordinates are the values and subgradients of such a function at the points $x^1, ..., x^N$ is a point in the constraint set. Note the contrast between this easy and cheap method of finding points in the constraint set and the standard random method, which in a problem such as (7) proceeds by first randomly drawing values for each of the 204 coordinates and then checking whether the obtained values satisfy each of the 10,608 constraints.

The drawback of using parametric functions to find points in the constraint set is that this method does not guarantee that all points in the constraint set have the possibility of being drawn. To guarantee this, one needs to use the slower method of generating points using the minimum of linear functions. Using this method is also necessary to guarantee the convergence of the proposed random search method.

The next lemma, which can be easily proved, establishes conditions guaranteeing that the probability density of points Y generated using the minimum of randomly drawn linear functions is everywhere positive on the interior of the constraint set.

LEMMA 2: Let $x^1, ..., x^N$ be given points in the domain of a concave function h^*. Let $Z = (a^1, ..., a^N, b^1, ..., b^N) \in R^N \times R^{NK}$ be a random vector possessing an everywhere positive density. Let the random vector Y=$(h^1, ..., h^N$,$T^1, ..., T^N)$ be defined from Z as follows:

If $a^i + b^i \cdot x^i \leq a^j + b^j \cdot x^i$ for all $j = 1, ..., N$,

set $(h^i, T^i) = (a^i + b^i \cdot x^i, b^i)$.

Otherwise, set $(h^i, T^i) = (a^j + b^j \cdot x^i, b^j)$, where j is the first index for which

$a^j + b^j \cdot x^i \leq a^k + b^k \cdot x^i$ for all $k = 1, ..., N$.

Then,

(2.1) the coordinates of Y correspond to the values and subgradients at $x^1, ..., x^N$ of a function h defined by

$h(x) = \min\{a^j + b^j \cdot x^i | j = 1, ..., N\}$, and

(2.2) the probability density of Y is everywhere positive on the constraint set defined by $(R.1)$ in (1).

Using the facts described above we can design a method that can be applied to find a solution to optimization problems such as (1). The method proposed in this paper searches for the maximum of the objective function over randomly determined segments that are contained in the constraint set. Three types of segments are used,

1. parallel segments
2. parametric segments
3. min_lin segments.

And three different types of searches along the segments are considered,

1. grid/line search
2. deep search
3. extreme point search

Parallel segments are segments parallel to one of the axes. They are obtained by calculating the boundary points of the segment that is the intersection between the constraint set and a line parallel to one of the axes. To calculate such a segment, one needs to start from a point that is inside the constraint set. This point will not necessarily be one of the extreme points of the segment. To determine parametric and min_lin segments, it is also necessary to start from a point that is inside the constraint set. In these cases, the starting point will be one of the extreme points of the segment. Parametric segments are determined by using parametric families of functions to draw the other extreme point of the segment. Min_lin segments are determined by using the minimum of linear functions to determine this other extreme point.

When the search for a point that attains the maximum value of the objective function over a segment is found by using either a grid or a line search or both, we call the search a grid/line search. The deep search method selects few points along the segment and starting from each of these points searches along the parallel segments that are determined starting from these points and moving along each of the axes. In the extreme point search, one calculates the value of the objective functions only at one or both of the extreme points of the segment. This kind of search is used to find the maximum of monotone objective functions over parallel segments. Since along parallel segments only one coordinate of Y varies in value, the maximum will be obtained in the upper (lower) extreme if the objective function is monotone increasing (decreasing) in that coordinate. Hence, there is no need of checking the value of a monotone objective function over all points along such segments.

The relative number of times that one should draw and search over each of the possible sets of segments and the type of search used along each of them should depend on the particular problem and how important the accuracy of the solution is.

Drawing parallel segments is extremely fast, but if the search is performed exclusively along these segments, some important directions in which the objective function increases may be missed. Drawing parametric segments is slower than drawing parallel segments, but there is a much higher chance of finding directions of increase of the objective function. The fewer the number of parameters defining a parametric family, the faster it is to draw such segments, but the higher the probability of missing important directions of increase of the objective function. One could start out drawing parametric segments using parametric families defined by a small number of parameters, and gradually draw more segments determined by higher-dimensional, more flexible parametric families. The more accurate one wants the solution to be, the more flexible the parametric families that should be added. Note that even when the only segments used are parallel and parametric segments, the resulting solution may not lie in the space of solutions that can be obtained by maximization over a family of parametric functions. The reason for this is that the extreme points of parallel axes do not necessarily belong to point attained by a parametric family. Hence, convex combinations of one of these points with a parametric point will not necessarily be a parametric point. Nevertheless, one could think of this procedure as justifying maximizing the objective function over points generated by all possible parametric functions that satisfy a particular set of shape restrictions.

Finally, as the convergence results in Subsection 3.4 indicate, to guarantee that the probability of finding a solution tends to one as the number of draws tends to infinity, it is required that min_lin segment points be also used.

3.3 Description of the algorithm

A procedure that incorporates the features mentioned above into an algorithm is next described.

In this algorithm, the objective function is denoted by $L(\cdot)$ and its argument by $Y \in R^L$. The algorithm requires an initial point, called Y^0, that lies inside the constraint. Once this point is selected and the value that the objective function attains at this point is determined, the algorithm tries to find a new point, inside the constraint set, at which the objective function attains a larger value. To find such a new point, the objective function of the optimization problem is maximized over randomly drawn segments generated using the initial point Y^0. When a higher-valued point is found, this new point takes the place of the initial point Y^0 and the search for a new point is undertaken. The algorithm stops when no higher-valued point is found after a predetermined number of consecutive searches.

Let Y^0 denote the initial point inside the constraint set. For all $k \in \{1, ..., L\}$, let $1_k = (0, ..., 0, 1, 0, ..., 0) \in R^L$ denote the vector whose coordinates are all equal to 0 except for the k^{th} coordinate, whose value equals 1. Let NUMBER_REPETITIONS, NUMBER_AXES, NUMBER_PARA_POINTS, and NUMBER_MIN_LIN_POINTS be four nonnegative numbers. Then, the steps of the algorithm are as follows:

(0.0) Set REPETITIONS = 0

(0.1) Set REPETITIONS = REPETITIONS + 1

(1.0) Set AXES = 0

(1.1) Set AXES = AXES + 1.

(1.2) Randomly draw a number $k \in \{1, ..., L\}$

(1.3) Find numbers d_l and d^u in R such that $[Y^0 + d_l\ 1_k,\ Y^0 + d^u\ 1_k]$ is the intersection of the constraint set with the line $\{Y \mid Y = Y^0\ + d\ 1_k,\ d \in R\}$.

(1.4) If the objective function is strictly increasing in the k^{th} coordinate of Y, set $Y^1 = Y^0 + d^u\ 1_k$. If the objective function is strictly decreasing in the k^{th} coordinate of Y, set $Y^1 = Y^0 + d_l\ 1_k$. If the objective function is constant in the k^{th} coordinate of Y, set $Y^1 = Y^0$. Otherwise, use either a grid/line search to find a point in the segment $[Y^0 + d_l\ 1_k,\ Y^0 + d^u\ 1_k]$ at which $L(\cdot)$ attains its maximum value over the segment, and call this point Y^1.

(1.5) If $L(Y^1) > L(Y^0)$, set $Y^0 = Y^1$ and go to step (1.0). Otherwise, perform a deep search over $[Y^0 + d_l\ 1_k,\ Y^0 + d^u\ 1_k]$ and if a point Y^2 is found with $L(Y^2) > L(Y^0)$ set $Y^0 = Y^2$ and go to step (1.0).

(1.6) If AXES < NUMBER_AXES go to (1.1).

(2.0) Set PARA_POINTS = 0

(2.1) Set PARA_POINTS = PARA_POINTS + 1.

(2.2) Randomly draw a parametric function whose values and subgradients satisfy the constraints, and calculate the point Y^2 corresponding to the values and subgradients of that function.

(2.3) Use either a grid or a line search to find a point in the segment $[Y^0,\ Y^2]$ at which $L(\cdot)$ attains its maximum value over the segment. Call this point Y^1.

(2.4) If $L(Y^1) > L(Y^0)$, set $Y^0 = Y^1$ and go to step (1.0). Otherwise, perform a deep search over $[Y^0 + d_l\ 1_k,\ Y^0 + d^u\ 1_k]$ and if a point Y^3 is found with $L(Y^3) > L(Y^0)$ set $Y^0 = Y^3$ and go to step (1.0).

(2.5) If PARA_POINTS < NUMBER_PARA_POINTS go to (2.1).

(3.0) Set MIN_LIN_POINTS = 0

(3.1) Set MIN_LIN_POINTS = MIN_LIN_POINTS + 1.

(3.2) Randomly generate a point Y^2 using the procedure described in Lemma 2 to generate Y.

(3.3) Use either a grid or a line search to find a point in the segment $[Y^0, Y^2]$ at which $L(\cdot)$ attains its maximum value over the segment. Call this point Y^1.

(3.4) If $L(Y^1) > L(Y^0)$, set $Y^0 = Y^1$ and go to step (1.0). Otherwise, perform a deep search over $[Y^0 + d_l\ 1_k, Y^0 + d^u\ 1_k]$ and if a point Y^2 is found with $L(Y^2) > L(Y^0)$ set $Y^0 = Y^2$ and go to step (1.0).

(3.5) If MIN_LIN_POINTS < NUMBER_MIN_LIN_POINTS go to (2.1).

(4.0) If REPETITIONS < NUMBER_REPETITIONS go to (0.1)

(4.1) Set the solution to the maximization problem to be Y^0.

3.4 Convergence properties of the algorithm

In this section, we present some convergence results of the algorithm described in the previous subsection. Roughly stated, Theorem 1 states that when the objective function of the optimization problem is continuous, the probability that the value of the objective function at the solution vector is ε away from the optimal value converges to one as the number of min-lin segments drawn converges to infinity. This result does not require that the number of searches over parallel and parametric segments is positive. But, searching over those segments will improve the rate of convergence to the solution.

THEOREM 1: Suppose that the objective function, L, in (1) is continuous. Let Y^* be a solution to the optimization problem (1). Let D denote the constraint set of the optimization problem. Let $\varepsilon > 0$. Let $A = \{Y \in D|\ |L(Y) - L(Y^*| < \varepsilon\}$. If, in the above algorithm, NUMBER_REPETITIONS > 0, then the probability that the solution obtained from the above algorithm belongs to A converges to 1 as NUMBER_MIN_LIN_POINTS converges to infinity.

The result of Theorem 1 is, of course, not surprising. Since the distribution of the points generated by the minimum of linear functions is everywhere positive over the constraint set, the probability that at least one of the drawn points will belong to the set A should converge to one as the number of draws tends to infinity. Note that the existence of Y^* is easily established if the constraints in $(R.2)$ bound the values of $(h.T)$, since then the objective function is continuous and the constraint set if compact.

Theorem 2 presents a convergence property of the above algorithm for optimization problems whose objective function is a step function. The theorem requires that no $K+1$ of the x^i vectors lie on a common hyperplane. Under these conditions, the result of the theorem is that the probability that the solution obtained from the algorithm attains the maximum value of the objective function tends to one as the number of searches over min_lin segments tends to infinite.

THEOREM 2: Suppose that the objective function, L, in (1) is a step function whose value changes only when the order of $h^1, ..., h^N$ changes. Let $Y^*{=}(h^*, T^*)$ be a solution to the optimization problem in (1) where the constraints in $(R.2)$ impose lower and upper bounds on (h, T). Assume that for all $i = 1, ..., N$, $T^{*i} > 0$. Let $A = \{Y \in D| \ |L(Y) - L(Y^*| < \varepsilon\}$. Suppose that no $K + 1$ elements of the set $\{x^1, ..., x^N\}$ lie in a common hyperplane. Then, if in the above algorithm NUMBER_REPETITIONS > 0, the probability that the solution obtained from the above algorithm belongs to A converges to 1 as NUMBER_MIN_LIN_POINTS converges to infinity.

Note that a solution to the optimization problem referred to in the above theorem always exists, because the number of different values that the objective functions attains is finite.

4 Use of the algorithm in a small experiment

In this section we present results of a small experiment that has been performed using a version of the algorithm described in the previous section. In the experiment, simulated data have been used to evaluate the performance of a nonparametric estimator. Results of this experiment show that nonparametric estimation methods may perform better than their parametric counterparts, with as few as 100 observations. We proceed to describe the experiment.

We used a binary threshold-crossing model to generate simulated data. In this model, an unobservable latent variable y^* is determined by the value of a systematic function h^*, at a vector of observable explanatory variables x, and the value of an unobservable random term δ , according to

$$y^* = h^*(x) - \delta.$$

The random variable δ is assumed to be independent of x. The c.d.f. of δ is denoted by F^*. Instead of observing y^*, we observe the value of a variable y, which is defined by: $y = 1$ if $y^* \geq 0$; $y = 0$ otherwise.

The model described in Section 2 about the decision of a firm to undertake a project, is an example of a binary threshold-crossing model. In that case, x is the vector of input prices, h^* is the cost function the typical firm faces, δ is the revenue, and y is a binary indicator that equals 0 when the firm decides to undertake the project and 1 otherwise.

In the experiment, $x = (x_1, x_2)$ was specified to be a 2-dimensional vector and the function h^* was specified, first, to be a nonlinear function and, second, to be a linear function. The nonlinear specification of h^* was

$$h^*(x_1,x_2) = 1.0897\ x_1 + 0.2179\ x_2 \text{ if } x_2/x_1 \geq 1.5$$

$$h^*(x_1,x_2) = 0.1667\ x_1 + 0.8333\ x_2 \text{ otherwise.}$$

The ratio of the coefficient of x_1 to the coefficient of x_2 in this nonlinear function is 5 when $x_2/x_1 \geq 1.5$ and it is 1/5 when $x_2/x_1 < 1.5$. Some isovalue sets of this function are graphed in

Figure 2. The linear specification used for h^* was

$$h^*(x_1,x_2) = 0.5\ x_1 + 0.5\ x_2\ .$$

The distribution function F^* of δ was specified to be the logistic distribution:

$$F^*(\delta) = [\ 1 + \exp(\text{-}(\delta - \mu)/\sigma)\ ]^{-1}\ .$$

The values of μ and σ were selected such that the mean of δ would equal the mean of h^* and the variance of δ would equal twice the variance of h^*.

One hundred points $\{x^i = (x_1^i, x_2^i)\}_{i=1}^{100}$ were selected and kept fixed across repetitions of the experiment. These points formed a "perturbed" grid over the square [20,120]x[20,120]. The perturbation was obtained by first generating a grid and then perturbing the first coordinate of the points in the grid. The perturbation was generated using a uniform distribution with zero mean and very small variance. The value of h^* at $x^{n+1} = (1,1)$ was taken to be 1. In each repetition of the experiment, data were generated by first drawing values for $\{\delta^i\}_{i=1}^{100}$ by using the distribution F^*, and then using h^*, $\{x^i\}_{i=1}^{100}$, and the randomly drawn $\{\delta^i\}_{i=1}^{100}$, to calculate $\{y^i\}_{i=1}^{100}$. In each repetition, the simulated data $\{(y^i, x^i)\}_{i=1}^{100}$ were used to estimate h^* by three different methods. The methods used to estimate h^* were a fully parametric method (FP), the distribution-free method (DF) introduced in Cosslett (1983), and the fully nonparametric method (FNP) introduced in Matzkin (1992). The fully parametric and the distribution-free methods impose a parametric structure on the function h^*. We chose a linear-in-parameters specification for h^* in this methods. Hence, in the fully parametric and distribution-free methods, the function h^* was specified to be

$$h^*(x_1,x_2) = \beta_1 x_1 + \beta_2 x_2.$$

When the true function h^* is a nonlinear function, this parametric specification allows us to evaluate the relative error produced when the function h^* is, as is typically the case, parametrically misspecified. When, on the other hand, the true function h^* is a linear function, this specification allows us to evaluate the decrease in MSE that occurs from relaxing parametric specifications.

The fully parametric and distribution-free methods estimate the values of the parameters

β_1 and β_2. The fully nonparametric method (FNP) only requires that the estimator for h^* be homogeneous of degree one, concave, monotone increasing, and attain the value 1 at $x = (1,1)$. Neither the distribution-free method nor the fully nonparametric method impose a parametric structure on F^*. The fully parametric method specifies a parametric structure for F^*, which was chosen to be the true one with the true values of μ and σ. (Note that by assigning to μ and σ their correct values, the FP method was given an advantage it does not usually have.) To guarantee that the parametric and nonparametric specifications for h^* were normalized in the same way, it was specified that

$$\beta_1 + \beta_2 = 1.$$

Applying the FP and DF methods, the estimated values for β_1 are the estimates for the first coordinate of the subgradient of h^* at any point x. When h^* is nonparametric, to estimate the first coordinate of the subgradient of h^* , at any point, we first used the estimated values for $(h^1, ..., h^N)$ and $(T^0, ..., T^{N+1})$ to obtain an estimate for h^*, and then calculated at each x the subgradient of the estimate for h^*. This was easily done using as an estimate for h^* the function defined by

$$h(x_1,x_2)=\min\{\widehat{T_1^i}x_1 + \widehat{T_2^i}x_2 | i = 0, 1, ..., N + 1\},$$

where $\widehat{T^0}$,...,$\widehat{T^{N+1}}$ are the estimated values for T^0,...,T^{N+1}.

To study the performance of the three estimation methods, we calculated statistics on the probability density of the estimates for the first and second coordinates of the subgradient of h^* at five different points. The points, X1=(20,120), X2=(45,95), X3=(70,70), X4=(95,45), and X5=(120,20), are graphed in Figure 2. Table 1 presents the statistics for the estimators obtained using the data generated by the nonlinear function, while Table 2 presents the statistics for the estimators obtained using the data generated by the linear function. The graphs of the probability densities of these estimators are presented, respectively, in the pages entitled "Nonlinear Function" and "Linear Function". In these tables and plots, DiXj denotes the estimator for the i-th coordinate of the subgradient of $\widehat{h}$ at the point Xj (i=1,2;j=1,...,5), FP denotes the fully parametric estimator, whose estimated density is graphed in a "_ . _" line, DF denotes the distribution-free estimator, whose estimated density is graphed in a "· · ·" line, and FNP denotes the fully-nonparametric estimator, whose estimated density is graphed in a "_ _ _" line. The results were obtained using 100 observations and 100 replications.

It took approximately one hour in a DELL Dimension XPS R450 computer to find a solution in each repetition, using, in the above algorithm, NUMBER_AXES = 500, NUMBER_PARA_POINTS = 10,000, NUMBER_MIN_LIN_POINTS = 70,000, and NUMBER_REPETITIONS = 1. This time should increase considerably if the number of observations or the number of explanatory variables increases, or if fewer restrictions are imposed on the nonparametric function. Our experience suggests that the speed of convergence can be considerably increased by first finding the maximum over the parametric segments, and only then using the min-lin segments to fine-tune the solution. In the reported experiment, the linear functions for the min-lin segments were generated using the point that was found to have the highest value before the segment was generated as the mean of the generated lines. I.e., the coefficients of the i-th linear function were drawn from a distribution whose mean was the vector of subgradients at observation x^i of the best found point.

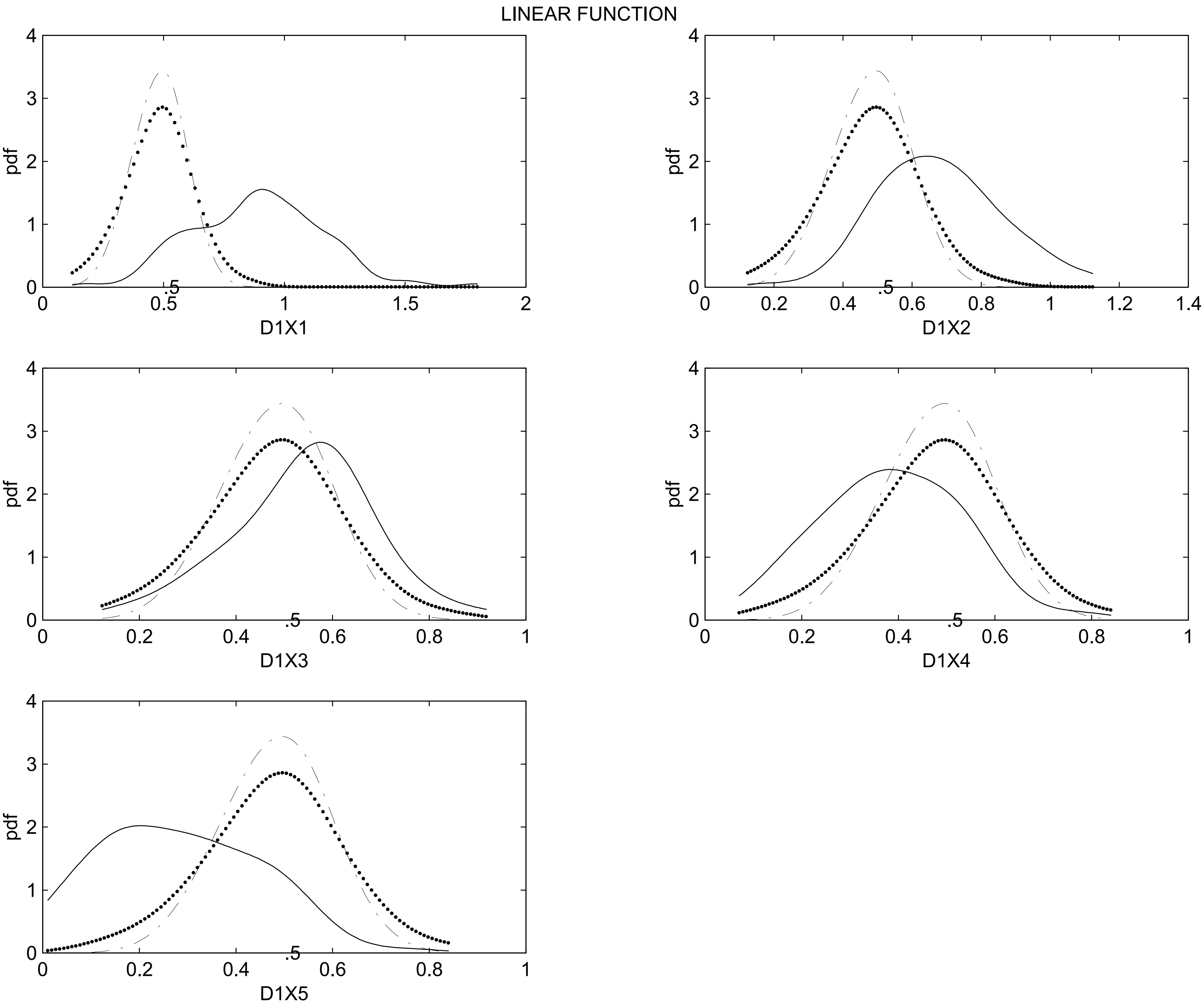
These results show that, when the data is generated by the nonlinear function, the FNP estimator may perform better, in terms of root-mean-squared-error, than the FP and DF estimators. When the data is generated by the linear function, the bias of the FNP and DF estimators is small, while the FNP estimator has some bias in addition to a larger standard deviation. Except at the endpoints, the ratio of the RMSE between the FNP and DF estimators is, in the linear case, between 1.1 and 2. The ratio between the RMSE of the DF and the FNP is 1.5. Hence, this provides some evidence that relaxing the parametric assumption on the systematic function may generate an increase in the RMSE that is similar to the increase generated by relaxing the parametric assumption on the distribution of the random term.

TABLE 1

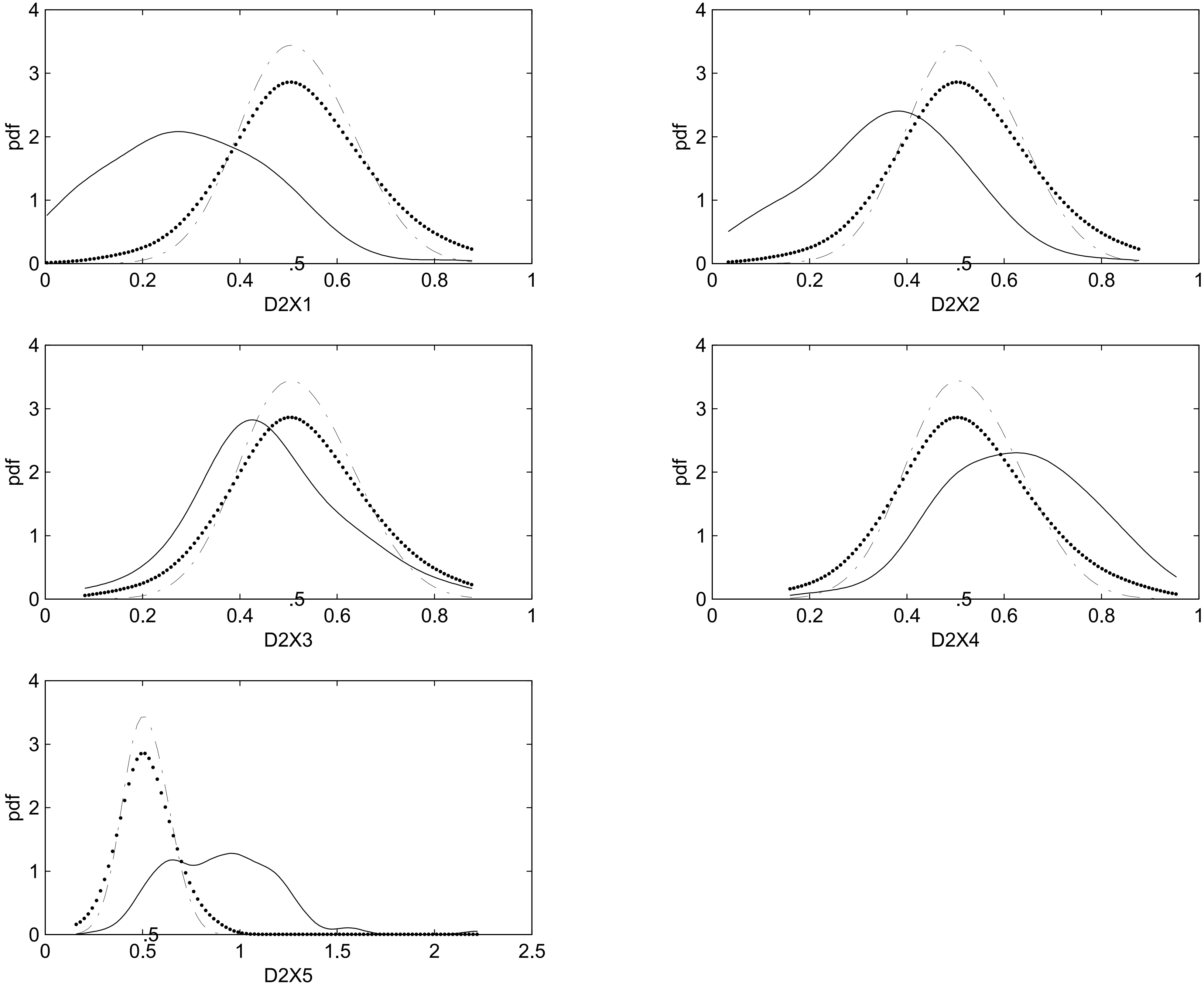
	FP			DF			FNP		
	$BIAS$	STD	$RMSE$	$BIAS$	STD	$RMSE$	$BIAS$	STD	$RMSE$
$D1X1$	−.721	.090	.727	.723	.145	.737	−.015	.398	.398
$D1X2$	−.721	.090	.727	.723	.145	.737	−.445	.206	.490
$D1X3$	.202	.090	.221	.200	.145	.247	.252	.160	.299
$D1X4$	.202	.090	.221	.200	.145	.247	.078	.137	.158
$D1X5$	.202	.090	.221	.200	.145	.247	−.009	.131	.131
$D2X1$	.414	.090	.221	.200	.145	.247	.064	.193	.203
$D2X2$	.414	.090	.221	.200	.145	.247	.201	.178	.269
$D2X3$	−.202	.090	.221	.200	.145	.247	−.252	.160	.299
$D2X4$	−.202	.090	.221	.200	.145	.247	−.061	.137	.150
$D2X5$	−.202	.090	.221	.200	.145	.247	.166	.203	.263

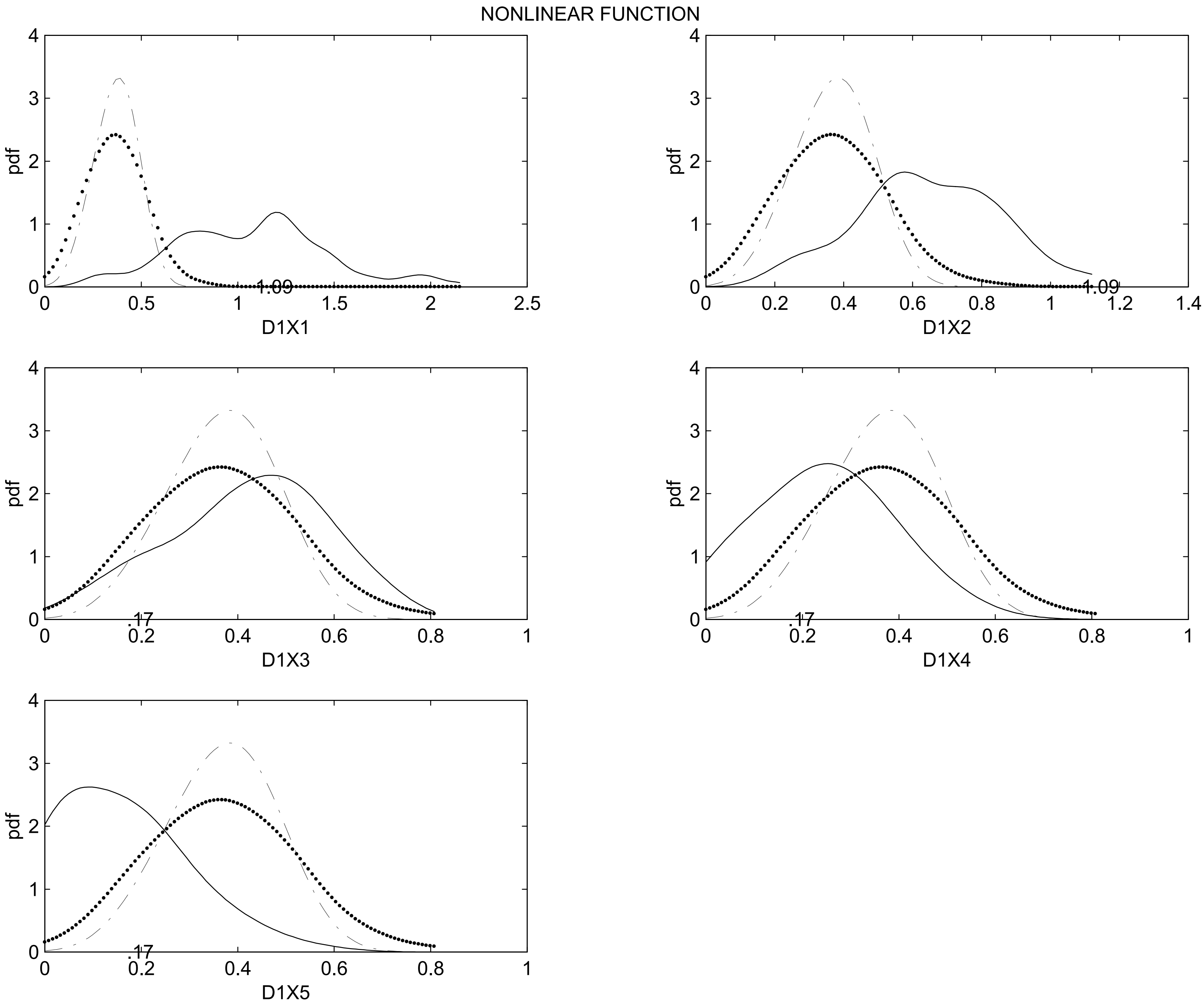
TABLE 2

	FP			DF			FNP		
	$BIAS$	STD	$RMSE$	$BIAS$	STD	$RMSE$	$BIAS$	STD	$RMSE$
$D1X1$	−.019	.087	.089	−.021	.132	.134	.391	.273	.478
$D1X2$	−.019	.087	.089	−.021	.132	.134	.179	.179	.253
$D1X3$	−.019	.087	.089	−.021	.132	.134	.039	.146	.151
$D1X4$	−.019	.087	.089	−.021	.132	.134	−.115	.142	.182
$D1X5$	−.019	.087	.089	−.021	.132	.134	−.216	.163	.271
$D2X1$	.019	.087	.089	.021	.132	.134	−.207	.165	.264
$D2X2$	.019	.087	.089	.021	.132	.134	−.140	.157	.210
$D2X3$	.019	.087	.089	.021	.132	.134	−.039	.146	.151
$D2X4$	.019	.087	.089	.021	.132	.134	.127	.145	.192
$D2X5$	.019	.087	.089	.021	.132	.134	.399	.294	.496

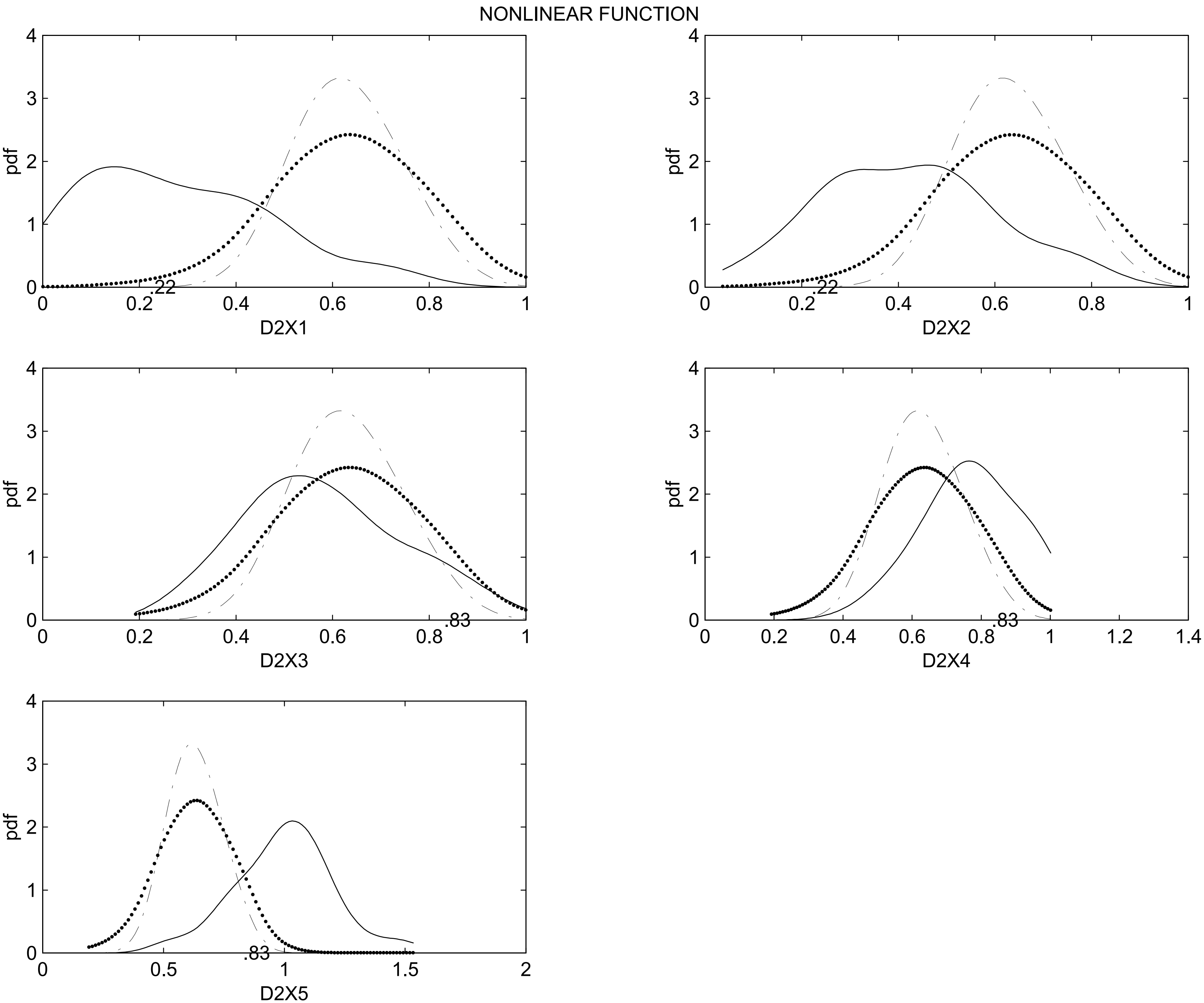

LINEAR FUNCTION
pdf
D1X1
D1X2
D1X3
D1X4
D1X5
.5
0
0.5
1
1.5
2
0.2
0.4
0.6
0.8
1.2
1.4
3
4

LINEAR FUNCTION





NONLINEAR FUNCTION
pdf
D1X1
1.09
D1X2
1.09
D1X3
.17
D1X4
.17
D1X5
.17


NONLINEAR FUNCTION
pdf
D2X1
.22
D2X2
.22
D2X3
.83
D2X4
.83
D2X5
.83

5 Conclusion

We have described a method for solving constrained optimization problems of the sort that arise in the estimation of nonparametric concavity-restricted functions. The method exploits particular properties of these problems, such as the convexity of the constraint set, the sparsity of the matrix defining the constraint set, and the fact that points in the constraints can be obtained from parametric functions and from functions that are the minimum of linear functions.

We have used the method in a small experiment. The objective of the experiment was to evaluate the performance of a nonparametric estimator for threshold-crossing models, which imposes concavity, homogeneity of degree one, and monotonicity on one of the functions being estimated.

Results from the small experiment show that the new nonparametric estimator may perform better than parametric estimators, with as few as 100 observations. It is to be expected that when fewer restrictions are imposed on h^* , and whe n the dimensionality of x is larger than 2, the relative superiority of the fully nonparametric method will decrease. Nevertheless, these results present at least some evidence that nonparametric estimation methods that use concavity restrictions can be useful as either substitutes or complements for parametric methods.

6 Appendix

PROOF OF LEMMA 1: Suppose that (h, T) satisfies $(R.1)$. Define $m_i : R^K \rightarrow R$ by $m_i(x) = h^i + T^i \cdot (x - x^i) \quad (i = 1, ..., N)$. Each m_i is a linear function. Define $m : R^K \rightarrow R$ by $m(x) = \min\{m_i(x)|\ i = 1, ..., N\}$. Then, m is the minimum of N linear functions. By $(R.1)$ it follows that $m(x^i) = m_i(x^i) = h^i$ $(i = 1, ..., N)$ and it is then easily seen that T^i is a subgradient of m at x_i $(i = 1, ..., N)$.

Conversely, let m be a function that is the minimum of N linear functions and it is such that $m(x_i) = h^i$ and T^i is a subgradient of m at x^i $(i = 1, ..., N)$. Then, m is concave. Hence, $h^1, ..., h^N$ and $T^1, ..., T^N$ satisfy $(R.1)$.

PROOF OF THEOREM 1: Let D denote the constraint set. Since L is a continuous function, there exists $\eta > 0$ such that $\|Y - Y^*\| < \eta$ implies $|L(Y) - L(Y^*)| < \varepsilon$. Then, by the properties of D, there exists a neighborhood B in $C \cap \{Y|\ \|Y - Y^*\| < \eta\}$. Since by Lemma 2 the density used to generate the Y^2 points in steps (3.0)-(3.5) is everywhere positive on D, it follows that for some $\mu > 0$, $\Pr(Y^2 \in B) = \mu$. So, the probability that at least one of M Y^2 points so obtained belongs to B is $1 - (1 - \mu)^M$. Hence, the probability of drawing a point in A converges to 1 as NUMBER_MIN_LIN_POINTS converges to infinity.

PROOF OF THEOREM 2: Again, let D denote the constraint set. We will show that the assumptions of the theorem imply that there exists a neighborhood B in $D \cap A$. The result will then follow by the argument used in the proof of Theorem 1.

For any two observations k, s, let $K_{k,s}(Y)$ denote the value of the concavity constraint that relates the coordinates of Y that correspond to observations k and s. Note that the vector h^*=(h^{*1},...,h^{*N}) is a point in the boundary of A iff for some i, j $h^{*i} = h^{*j}$. And the vector Y^*=(h^*,T^*)= (h^{*1},...,h^{*N},$T^{*1}, ..., T^{*N}$) is a point in the boundary of D iff for some k, s

$K_{k,s}(h^*,T^*) = 0$.

Suppose first that Y^*=(h^*,T^*) belongs to the interior of D. Then, it is easy to see that the intersection of any neighborhood around Y^* with the set A will necessarily contain an open set in D.

Suppose next that h^* belongs to the interior of A. Define $m : R^K \to R$ by $m(x) = \min\{\ h^{*i} + T^{*i} \cdot (x - x^i)|i = 1, ..., N\}$. Then, m is a concave function such that no two x^is are in a common isovalue set. The uppercontour sets of m corresponding to each h^{*i} value may not be strictly convex sets. But it is easy to see that strictly convex sets, $C^1, ..., C^N$, can be found such that, for each i, x^i belongs to the boundary of C^i and for each i, j such that $h^{*i} < h^{*j}$ one has that C^j is included in the interior of C^i. Since D is compact, the results in Kannai (1974), MasColell (1974), and Matzkin (1988, 1991c) imply that there exists a concave, strictly quasiconcave function, q, such that the boundaries of the C^i sets are level sets of this function q. Let s be the composition of a strictly increasing and strictly concave function with q. Then, the boundaries of the C^i sets are also level sets of the function s. It then follows that $h^{*i} < h^{*j}$ if and only if $s(x^i) < s(x^j)$, and since s is strictly concave, there exist vectors $T_s^1, ..., T_s^N$ such that $s(x^1), ..., s(x^N)$ and $T_s^1, ..., T_s^N$ satisfy the constraints in $(R, 1)$ with strict inequalities. Hence, $(s(x^1), ..., s(x^N))$ belongs to A and $(s(x^1), ..., s(x^N);T_s^1, ..., T_s^N)$ solves the optimization problem and it belongs to the interior of D. As in the previous paragraph, this implies that the intersection of A with D will necessarily contain an open set in D.

Finally, suppose that it is neither the case that h^* belongs to the interior of A nor the case that Y^* belongs to the interior of D. Define the function m as in the above paragraph. It then follows that there exist at least two x^i's in a common isovalue set of m. Since L is a step function, A contains a point $(h^1, ..., h^N)$ such that for no i, j such that $i \neq j$, h^i =h^j. Since T^* is strictly positive and no subset containing $K + 1$ elements of $\{x^1, ..., x^N\}$ is such that its elements lie on a common hyperplane, it follows that convex sets, $\widetilde{C}_1, ..., \widetilde{C}_N$, can be found such that, for each i, x^i belongs to the boundary of $\widetilde{C}_i$ and for each i,j such that $h^i < h^j$ one has that $\widetilde{C_j}$ is included in the interior of $\widetilde{C}_i$. (See, for example, the techniques used in Matzkin (1988, 1991c).) Since D is compact, the results in Kannai (1974) and MasColell (1974) imply that there exists a concave function, q, such that the boundaries of the $\widetilde{C}_i$ sets are level sets of this function q. It then follows that $h^i < h^j$ if and only if $q(x^i) < q(x^j)$. Since q is concave, there exist vectors $T_q^1, ..., T_q^N$ such that $s(x_1), ..., x(x^N)$ and $T_q^1, ..., T_q^N$ satisfy the constraints in $(R.1)$. Hence, $(q(x_1), ..., q(x^N))$ belongs to the interior of A and $(q(x_1), ..., q(x^N); T_q^1, ..., T_q^N)$ solves the optimization problem. By the arguments in the above paragraph, this implies that the intersection of A with D contains an open set in D. This completes the proof of Theorem 2.

7 References

AFRIAT, S. (1967a), "The Construction of a Utility Function from Demand Data," International Economic Review, Vol. 8.

_______. (1967b), "The Construction of Separable Utility Functions from Expenditure Data," mimeo, Purdue University.

_______. (1972), "The Theory of International Comparison of Real Income and Prices," in J. D. Daly (ed.), International Comparisons of Prices and Output. New York: National Bureau of Economic Research.

_______. (1973), "On a System of Inequalities in Demand Analysis," International Economic Review, Vol. 14.

_______. (1981), "On the Constructability of Consistent Price Indices between Several Periods Simultaneously," in A. Deaton (ed.) Essays in Applied Demand Analysis. Cambridge University Press.

ASHER, M., H. D. BRUNK, G. M. EWING, W. T. REID, and E. SILVERMAN (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," Annals of Mathematical Statistics , 26, 641-647.

BROWN, D.J. and R.L. MATZKIN (1990), "Recoverability and Estimation of the Demand and Utility Functions of Traders when Demands are Unobservable," mimeo, Cowles Foundation, Yale University.

____________________________. (1995), "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an application to Consumer Demand," mimeo, Northwestern University.

CHIAPPORI, P. and J. ROCHET (1987), "Revealed Preference and Differential Demand," Econometrica, 55, 687-691.

COSSLETT, S.R. (1983), "Distribution-free Maximum Likelihood Estimator of the Binary

Choice Model," Econometrica, Vol. 51, No. 3, pp. 765-782.

DIEWERT, E.W. (1973), "Afriat and Revealed Preference Theory," Review of Economic Studies, 40.

DIEWERT, E.W. and C. PARKAN (1978), "Test for Consistency of Consumer Data and Nonparametric Index Numbers," University of British Columbia, Working Paper 7827.

DYKSTRA, R.L. (1983), "An Algorithm for Restricted Least Squares," Journal of the American Statistical Association, 78 (364), 837-842.

EPSTEIN, L.G. and D.J. YATCHEW (1985), "Non-Parametric Hypothesis Testing Procedures and Applications to Demand Analysis," Journal of Econometrics, 30.

GOLDMAN, S.R. and P.A. RUUD (1993), "Nonparametric Multivariate Regression Subject to Constraints," Technical Report 93/213, Department of Economics, University of California, Berkeley.

HILDRETH, C. (1954), "Point Estimates of Ordinates of Concave Functions," Journal of the American Statistical Association, 49, 598-619.

HOUTHAKKER, H. (1950), "Revealed Preference and the Utility Function," Economica, 17, 159-174.

KANNAI, Y. (1974), "Approximation of Convex Preferences," Journal of Mathematical Economics, 1.

MAS-COLELL, A. (1974), "Continuous and Smooth Consumers: Approximatin Theorems," Journal of Economic Theory, 8, 305-336.

MATZKIN, R. L. (1988), "Nonparametric Tests of Maximizing Behavior Subject to Nonlinear Sets," Cowles Foundation Discussion Paper No. 894.

_______. (1990), "Least-concavity and the Distribution-free Estimation of Nonparametric Concave Functions," Cowles Foundation Discussion Paper No. 958, Yale University.

_______. (1991a), "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," Econometrica, 59, 1315-1327.

________. (1991b), "A Nonparametric Maximum Rank Correlation Estimator", in Nonparametric and Semiparametric Methods in Econometrics and Statistics, by W. Barnett, J. Powell, and G. Tauchen (eds.), Cambridge: Cambridge University Press.

________. (1991c), "Axioms of Revealed Preference for Nonlinear Choice Sets," Econometrica, 59.

________. (1992), "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models," Econometrica, 60, 239-270.

________. (1993), "Fully Nonparametric Estimation of Polychotomous Choice Models," Journal of Econometrics, 139-168.

________. (1994), "Restrictions of Economi Theory in Nonparametric Methods," in Handbook of Econometrics, Vol. 4, edited by D.L. McFadden and R.F. Engel.

MATZKIN, R.L. and M.K. RICHTER (1991), "Testing Strictly Concave Rationality, " Journal of Economic Theory, 53, 2, 287-303.

NELDER J.A. and R. MEAD (1965), "A Simplex Method for Function Minimization," Computer Journal, Vol. 7, 308-313.

RICHTER, M.K. (1966), "Revealed Preference Theory," Econometrica, 34, 635-645.

RUUD, P.A. (1995), "Restricted Least Squares Subject to Monotonicity and Concavity Constraints," presented at the 7th World Congress of the Econometric Society.

SAMUELSON, P. (1938), "A Note on the Pure Theory of Consumer Behavior," Economica, 5, 61-71.

VARIAN, H. (1982), "The Nonparametric Approach to Demand Analysis," Econometrica, 50, No. 4.

________. (1983), "Nonparametric Tests of Consumer Behavior", Review of Economic Studies, 50.

________. (1985), "NonParametric Analysis of Optimizing Behavior with Measurement Error," Journal of Econometrics, 30.

_______. (1988), "Goodness-of-Fit in Demand Analysis," CREST Working Paper No, 8911, University of Michigan.