Restricted Least Squares Subject to Monotonicity and Concavity Constraints

Paul A. Ruud
University of California, Berkeley

July, 1995 (Revised January, 1996)

1 Introduction

Economists have devoted a large research effort to the estimation of cost functions and profit functions. Since the popularization of the dual approach to such functions, econometricians have focussed particularly on methods for fitting functions that satisfy the restrictions implied by optimizing behavior, minimizing costs or maximizing profits. For the most part, researchers have sought simple parametric functional forms that are sufficiently flexible to approximate well all possible functions in the families of cost and profit functions. Given such parametric functions, much of the estimation has followed the method of least squares.

In this paper, I will describe a computational approach to fitting cost and profit functions by the method of least squares subject only to the restrictions imposed by the theory of optimizing behavior. Instead of tightly parameterized functional forms, I will use as many parameters as required to cover all the permissible functions. The computational approach has a long history, but this paper grows specifically out of previous, joint research with my colleague, Steven Goldman. I will illustrate the computation with two examples drawn from the empirical literature on the estimation of cost functions.

There are several reasons to pursue the least squares fit of cost and conditional factor demands. First of all, researchers have worked for many years on appropriate functional forms for these functions.[1] Economists have proposed various parametric functional forms designed to exhibit their theoretical properties and to be amenable to conventional parametric statistical estimation methods.[2] This requirement explains the popularity of such parametric cost functions as the translog. In addition, researchers desire flexible functional forms so that their inferences are not misled by statistical artifacts of parametric misspecification. The approach presented here also yields a specification that accommodates factor and output levels of zero naturally.

Another motivation for our line of research is the desire to explore various ways to let the data determine the amount of smoothing in regression estimation. Nonparametric research has focused heavily on such techniques as kernel smoothing. In this paper, we obtain smoothness through structural restrictions of monotonicity and concavity on the regression function. One can also view the fits we compute here as the least smooth function that satisfies these structural constraints. Anything else is a further restriction. Thus, one can learn what additional structure must be imposed to sharpen one's empirical inference.

This paper focuses on computational problems. We do not consider the statistical properties of the least squares estimator. Hanson & Pledger (1976) have demonstrated the consistency of the estimator for the univariate case. Approximate (asymptotic) distributions are non-normal and quite complicated for the simplest cases. See Yazhen (1992) for an analysis. Our purpose is to introduce the computational problem and its history, describe a solution method, and apply this method to estimating a cost function and its associated factor demand functions. See Matzkin (1992) for an alternative approach.

[1] The econometric literature on parametric functional forms is extensive and so we point to some leading examples. Diewert (1973) initiated much of the analysis of flexible functional forms. Lau (1978) proposed a transformation of parameters for imposing monotonicity and convexity in quadratic functions.

[2] Diewert & Wales (1987) call the construction of restricted parametric functions "one of the most vexing problems applied economists have encountered" and comment that the comprises which researchers strike have often proved unsatisfactory. They propose a family of semiflexible functions with which one can impose convexity and choose the degree of flexibility.

2 The Computational Approach

Computing the least squares fit of a cost or profit function is a classic quadratic programming problem. Our computational approach is to compute the solution to a dual problem. The solution to the dual quadratic programming problem is computed by a combination of standard Gauss-Seidel and programming algorithms. However, the dimensions of "typical" problems in econometrics are too large for software engineered for a wide range of applications. We exploit the special structure of our particular problems to overcome this dimensionality.

2.1 Statement of the Problem

2.1.1 Univariate Concave Regression

To introduce our discussion, consider a particular example of restricted least squares, fitting a univariate, concave regression function, as proposed by Hildreth (1954). Given observations $\{(x_n, y_n)\,;\, n = 1, \ldots, N\}$, ordered so that $x_1 < x_2 < \cdots < x_N$, find

$$\min_{z} \|\mathbf{y} - \mathbf{z}\|^2 \tag{1}$$

subject to

$$\frac{z_{n+1} - z_{n+1}}{x_{n+2} - x_{n+1}} \le \frac{z_{n+1} - z_n}{x_{n+1} - x_n},\ n = 1, \ldots, N-2, \tag{2}$$

where $\mathbf{y} \equiv [y_n;\ n = 1, \ldots, N]$ and $\mathbf{z} \equiv [z_n;\ n = 1, \ldots, N]$. For exposition, we have specified that the x_n have distinct values.[3] A direct approach would be to optimize the elements of $\mathbf{z}$. But this is awkward because most z_n must satisfy several constraints simultaneously. Hildreth used an alternative approach.

Instead, Hildreth computed the solution to the dual problem. Let $\mathbf{Rz} \le 0$ denote the inequality constraints (2) gathered into matrix form. The dual program is

$$\min_{\boldsymbol{\lambda} \ge 0} \|\mathbf{y} - \mathbf{R}'\boldsymbol{\lambda}\|^2 \tag{3}$$

[3]If there were any n such that $x_n = x_{n+1}$, then one must insert equality constraints $z_n = z_{n+1}$. Hildreth allowed for these multiplicities. We avoid such additional detail.

which has an analytically simpler form for the constraints: Each element of $\boldsymbol{\lambda}$ must satisfy a single positivity constraint. These scalars are the Lagrange multipliers associated with each constraint. Given the optimal $\hat{\boldsymbol{\lambda}}$, the optimum of the primal problem is calculated simply as

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{R}'\hat{\boldsymbol{\lambda}}. \tag{4}$$

Hildreth applied Gauss-Seidel to (3), optimizing iteratively over each element of $\boldsymbol{\lambda}$ in an (arbitrary) fixed sequence. Each optimization has a simple solution:

$$\max\{0, b_r/c_r\} = \arg\min_{\lambda_r \geq 0} \|\mathbf{y} - \mathbf{R}'\boldsymbol{\lambda}\|^2, \tag{5}$$

where b_r and c_r are taken from the expansion of the objective function in each element λ_r:

$$\|\mathbf{y} - \mathbf{R}'\boldsymbol{\lambda}\|^2 \equiv a_r - b_r\lambda_r + \frac{1}{2}c_r\lambda_r^2.$$

In this way, he converted a relatively difficult problem into a series of very simple ones.

The basic duality is illustrated in Figure 1. The vectors R_1 and R_2 are two rows in the constraint matrix $\mathbf{R}$. The constrained set is $K = \{\mathbf{z} \mid \mathbf{Rz} \leq 0\}$ and its dual is $K^* = \{\mathbf{w} \mid \mathbf{w}'\mathbf{z} \leq 0 \; \forall \mathbf{z} \in K\}$. The optima $\hat{\mathbf{y}}$ and $\mathbf{R}'\hat{\boldsymbol{\lambda}}$ are denoted by the points A and B respectively. Note that these are orthogonal projections of $\mathbf{y}$ onto subspaces.

By updating each element of $\boldsymbol{\lambda}$, λ_r, with (5) sequentially as $r = 1, \ldots, M$ over M constraints, one completes a cycle of Hildreth's algorithm. A convenient starting point is to set $\boldsymbol{\lambda} = 0$. At every step within a cycle, say the i^{th} from $\hat{\boldsymbol{\lambda}}_i$ to $\hat{\boldsymbol{\lambda}}_{i+1}$, one reduces the length of the implicit fit, $\hat{\mathbf{y}}_i \equiv \mathbf{y} - \mathbf{R}'\hat{\boldsymbol{\lambda}}_i$.[4] Hildreth also proved that this Gauss-Seidel sequence converges to the optimum of (3). His proof is a classic demonstration of a contraction mapping, resting on the uniqueness of the optimal $\boldsymbol{\lambda}$, the boundedness of $\boldsymbol{\lambda}$, and the continuity of the objective function. See Goldman & Ruud (1993*b*) for a generalization of his theorem and proof that applies to the problems we consider below.

[4]One cycles through all the constraints repeatedly, in the sequence $\{1, \ldots, M, 1, \ldots, M, \ldots\}$. Thus, the i^{th} cycle will handle constraint $r = (i - 1 \,\text{modulo}\, M) + 1$.

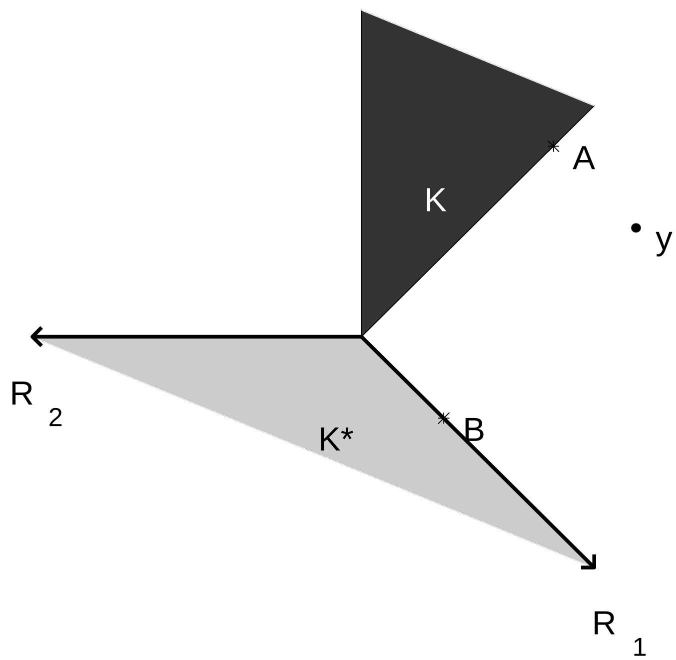

Figure 1: Quadratic Program Duality

2.1.2 Cost Functions

This univariate case is simpler than the general problem that we consider here: fitting a cost function in for production with multiple inputs and outputs. The basic structure is the same, but the constraints are more complicated. Let q denote a vector of J output levels, y denote a vector of M input factor levels, and p denote a vector of M corresponding factor prices. The cost function is

$$c(p, q) = \min_{\substack{(\tilde{q}, x) \in \mathbf{Q}, \\ \tilde{q} \geq q}} p' y$$

where $\mathbf{Q}$ is the production set.[5] Such functions are characterized by four properties:

1. $c(\cdot)$ is monotonically increasing in input prices and output levels;
2. $c(\cdot)$ is homogeneous of degree one in input prices;
3. $c(\cdot)$ is concave; and
4. $c(\cdot)$ is continuous.[6]

Associated with the cost function is a vector of M conditional factor demand correspondences that we denote by $z(p, q) \equiv [z_m(p, q);\ m = 1, \ldots, M]$. These are characterized by

1. $z_m(\cdot)$ is monotonically decreasing in p_m,
2. $z(\cdot)$ is homogeneous of degree zero in p,
3. if $z(\cdot)$ is single-valued at (p_1, q_1), then $z(p_1, q_1) = \nabla_p c(p_1, q_1)$, and
4. if $z(\cdot)$ is differentiable at (p_1, q_1), then $\nabla_p z(p_1, q_1)'$ is a symmetric, negative semidefinite matrix.[7]

---

[5] The production set is assumed to be nonempty and compact.

[6] In mathematical terms, $z \in \mathbf{C}$ if and only if z is continuous and

(a) $p_1 \geq p_2 \Rightarrow\ z(p_1, q) \geq z(p_2, q)$

(b) $p \gg 0,\ \alpha \geq 0 \Rightarrow z(\alpha \cdot p, q) = \alpha \cdot z(p, q)$

(c) $p_1, p_2 \gg 0,\ 0 \leq \alpha \leq 1 \Rightarrow\ z\left[\alpha \cdot p_1 + (1 - \alpha) \cdot p_2\right] \leq \alpha \cdot z(p_1, q) + (1 - \alpha) \cdot z(p_2, q)$

[7] These well-known properties of cost functions and conditional factor demand correspondences are explained in Varian (textbook) and Mas-Colell *et al* (textbook). We use ∇_p to denote the vector of partial derivatives with respect to the elements of p.

Given a data set of N observations $\{(p_n, q_n, y_n)\,; n = 1, \ldots, N\}$ on the price vector, output vector, and input vector, we will seek

$$\hat{\mathbf{y}} = \arg\min_{z \in \mathbf{C}} \sum_{n=1}^{N} \|y_n - z(p_n, q_n)\|^2 \tag{6}$$

where $\mathbf{C}$ is the set of single-valued conditional factor demand functions that are consistent with the properties of cost minimization listed above.[8] This problem is a generalization of the parametric estimators econometricians typically employ. It can be motivated by assuming normally distributed error terms associated with the demand or by specifying that the conditional factor demand function is the conditional expectation of the observable factor levels. We will refer to $\hat{y}$ as the "nonparametric" fit, though we will actually work with a high-dimensional parameterization below.

First of all, consider the constraints that cost minimization imposes on this minimization problem. We will not impose differentiability of z. Clearly, only the points $\{z_n \equiv z(p_n, q_n); n = 1, \ldots, N\}$ are relevant to the least squares problem. Therefore, we can restrict our parameterization, without loss of generality, to

$$\mathbf{C}_N = \left\{[z_{nm}] \in \mathbb{R}^N \times \mathbb{R}^M \mid z_n = z(p_n, q_n),\ z \in \mathbf{C}\right\}.$$

Similarly, the continuity of the cost function is not a binding restriction in this problem.

Imposing linear homogeneity as a restriction on z is straight-forward: One simply chooses one input as numéraire and normalizes all prices by the price of the numéraire. Monotonicity and convexity restrictions result in a nontrivial computational problem. The monotonicity constraints on cost functions require simply that every z_{nm} be nonnegative:

$$z_{nm} \geq 0. \tag{7}$$

As Varian (1984) points out, the concavity constraints imply that[9]

$$p_n' z_n \leq p_n' z_i \quad \forall n, i : q_n \leq q_i. \tag{8}$$

[8] It will be convenient to think of the norm $\|\cdot\|$ as the simple Euclidean norm, but one can generalize to generalized least squares problems straightforwardly.

[9] Varian (1984) calls these constraints the 'Weak Axiom of Cost Minimization' (WACM). He restricts his analysis to a scalar output, but this is not necessary for our analysis.

In words, the cost function $p_n' z_n$ is less than the cost at prices p_n of the conditional factor demands for producing at least the same output q_i at any other prices p_i.

Goldman & Ruud (1993*b*) observe that these constraints (7)–(8) together describe a closed convex polyhedral cone. It is an immediate consequence of the convexity of $\mathbf{C}_N$ (and the strict concavity of the objective function) that $\hat{y}$ is unique. This uniqueness is, of course, a highly desirable property for computation because it rules out the need to seek out multiple potential local optima.

It is also interesting to consider the estimation of the cost function without the factor input level data. Occasionally, input data are not available. In addition, comparisons of cost function estimates with and without input data may serve as a natural specification test. Given a data set of N observations $\{(p_n, q_n, c_n); n = 1, \ldots, N\}$ on the price vector, output vector, and total costs, the cost function estimation program solves

$$\min_{\mathbf{z} \in \mathbf{C}_N} \|\mathbf{c} - \mathbf{z}\|^2 ,$$

where $\mathbf{c} \equiv [c_n]$ is the vector of observed costs and $\mathbf{C}_N$ is a subset of $\mathbb{R}^N$ where $\mathbf{z} \in \mathbf{C}_N$ if

$$z_n \leq z_i \quad \forall n, i : p_n \leq p_i,\ q_n \leq q_i, \tag{9}$$

$$\exists \gamma_n \in \mathbb{R}_+^M : z_n = p_n' \gamma_n,\ z_i \leq p_i' \gamma_n \quad \forall n, i : q_i \leq q_n. \tag{10}$$

We normalize prices and costs so that homogeneity is imposed. The restrictions in (9) describe monotonicity and the restrictions in (10) describe concavity. The latter is a convenient alternative in higher dimensions to explicit expressions like (2), which require extensive computation and work space to apply. Also, (10) is the natural analogue to (8): The γ_n correspond to possible factor demands and the constraint states that the cost of the factor demands at (p_n, q_n) valued at prices p_i must not undercut minimized costs at p_i and any output less than q_n.

Again, the parameter space $\mathbf{C}_N$ is a closed convex cone so that there is a unique fitted $\hat{c}$. Note, however, that the implicit conditional factor demands will be a correspondence, not a function. In general, the values of γ_n that satisfy (10) will not be unique. Instead, the conditional factor demands will be a closed set.

Although the parameterization of the cost or conditional factor demands can be restricted to a finite number of points, the points where prices and outputs are observed in the sample, there are implicit restrictions on the cost function and demands at other prices and outputs as well. These restrictions yield estimators for prices and quantities that are correspondences, in much the same way as for conditional factor demands above.[10] Thus, for conditional factor demands, $\hat{y}(p,q)$ is the set that satisfies constraints exactly analogous to (7)–(8):

$$\begin{aligned} z(p,q) &\geq 0, \\ p'z(p,q) &\leq \min_{i:q\leq q_i} p'\hat{y}_i \\ p_i'\hat{y}_i &\leq p_i'z(p,q) \quad \forall i : q_i \leq q. \end{aligned}$$

For the cost function, $\hat{c}(p,q)$ is the set that satisfies

$$\begin{aligned} \hat{c}(p,q) \leq \hat{c}_i &\quad \forall i : p \leq p_i,\ q \leq q_i, \\ \hat{c}_i \leq \hat{c}(p,q) &\quad \forall i : p_i \leq p,\ q_i \leq q, \\ \exists \gamma \in \mathbb{R}^M : \hat{c}(p,q) = p'\gamma,\ \hat{c}_i \leq p_i'\gamma &\quad \forall i : q_i \leq q, \\ \exists \gamma_i \in \mathbb{R}^M : \hat{c}_i = p_i'\gamma_i,\ \hat{c}(p,q) \leq p'\gamma_i &\quad \forall i : q \leq q_i. \end{aligned}$$

These constraints are analogous to (9)–(10). For any (p,q), either of these sets can be computed easily with linear programming software.[11]

The computational problem that we face is that number and dimension of constraints in a data set is often so enormous that general programs for solving quadratic programs cannot accommodate them. Varian (1985) has computed the solution to a problem similar to (6) for a relatively small data set (18 observations, 3 factors, 1 output) using the MINOS package by Murtagh & Saunders (1967). In one of the examples below, there are over 200,000 restrictions on more than 2,500 parameters for a data set with 630 observations.

Goldman & Ruud (1993*a*) fit factor demands for a somewhat larger problem using Hildreth's method. In their work, each Gauss-Seidel iteration cor-

[10] Generally, we restrict attention within the convex hull of observed values of prices and quantities.

[11] If we treat the fitted restricted least squares values $\{\hat{y}_i;\ i = 1, \ldots, N\}$ as observed data, the boundaries of these cost functions correspond to the under- and overcost functions derived by Varian (1984, p. 593). The boundaries of the conditional factor demands correspond to Varian's $VI(q)$ and $VO(q)$ (p. 591).

responds to updating the fitted factor demands with

$$\begin{aligned} \begin{pmatrix} z_i \\ z_n \end{pmatrix} &= \left\{ \begin{array}{ll} \begin{pmatrix} w_i \\ w_n \end{pmatrix} & \text{if} \left\{ \begin{array}{c} p_n' w_n \leq p_n' w_i, \\ q_n \leq q_i \end{array} \right\} \\ \begin{pmatrix} w_i \\ w_n \end{pmatrix} - \alpha_{ni} \cdot \begin{pmatrix} p_n \\ -p_n \end{pmatrix} & \text{if} \left\{ \begin{array}{c} p_n' w_n > p_n' w_i, \\ q_n \leq q_i \end{array} \right\} \end{array} \right\}, \quad (11) \\ \alpha_{ni} &= \frac{p_n' (w_i - w_n)}{2 p_n' p_n} \end{aligned}$$

where the other components of z are set equal to the corresponding components of w. The projections for the inequalities (7) are even simpler:

$$z_{nm} = \max \{0, w_{nm}\} .$$

These calculations can be done very rapidly, rendering large problems into many, workable subproblems. Goldman & Ruud (1993*a*) found that the convergence of such sequences was extremely slow in the problem they considered. This slowness is a well-studied characteristic of a general class of calculations called *alternating projections*.[12]

3 Generalizing Hildreth's Procedure

The components of Hildreth's computational method have been generalized by Goldman & Ruud (1993*a*).[13] Their basic, and simple, insight is that one need not restrict the iterations of Gauss-Seidel to one element of $\boldsymbol{\lambda}$ at a time. More generally, one can optimize over subsets of the elements of $\boldsymbol{\lambda}$ simultaneously or alter the order in which the elements are taken. These two possibilities make the range of possible algorithms much bigger and their exploitation can substantially improve the speed of convergence. Goldman & Ruud (1993*b*) prove that the contraction property of such generalizations rests solely on the requirement that every constraint appears in at least one subproblem of an iteration.

For calculations like (11), one can understand the slow speed of convergence as a symptom of near multicollinearity among the restrictions. Each

[12] Many authors cite von Neumann (1950) as the originator of the method of alternating projections.

[13] See Goldman & Ruud (1993*b*) for a description of related research.

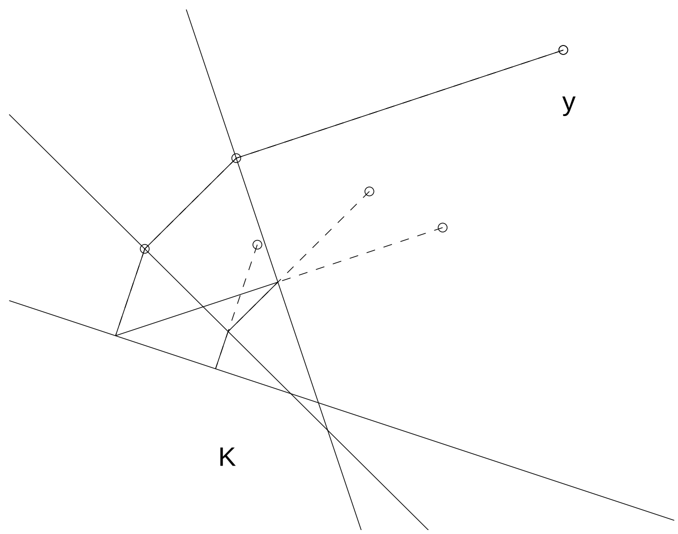

Figure 2: Alternating Projections

iteration of Gauss-Seidel corresponds to a projection onto a half-space.[14] See Figure 2 for an illustration. The solid line depicts the path of two iterations, through all constraints, of the Hildreth algorithm. In the second iteration, the circles depict the location of the fitted vector when $\lambda_r = 0$ and the dashed line represents a projection onto a half-space as λ_r is set to its optimal, positive value. Alternating projections between several highly collinear subspaces is illustrated in Figure 3. Because the intermediate projections are so close to one another, the algorithm makes small incremental steps toward the ultimate solution.

Goldman & Ruud (1993*a*) combine the elements of $\boldsymbol{\lambda}$ in two instances. First of all, for the estimation of cost functions without factor demand data, they constructed a sequence of least squares regressions that optimize with respect to the intersection of the concavity constraints (10) for fixed n. Secondly, when the iterations of Hildreth's procedure cycled over the same active constraints, they check whether the active constraints comprise a basis for the final solution. If so, they calculate the optimum (by direct linear calculation) and end the iterations. In this paper, we extend the principle of optimizing over several elements in $\boldsymbol{\lambda}$ in two additional ways. Firstly, we replace the individual factor demand projections in (11) with a single projec-

[14]This interpretation is emphasized by Dykstra (1983).

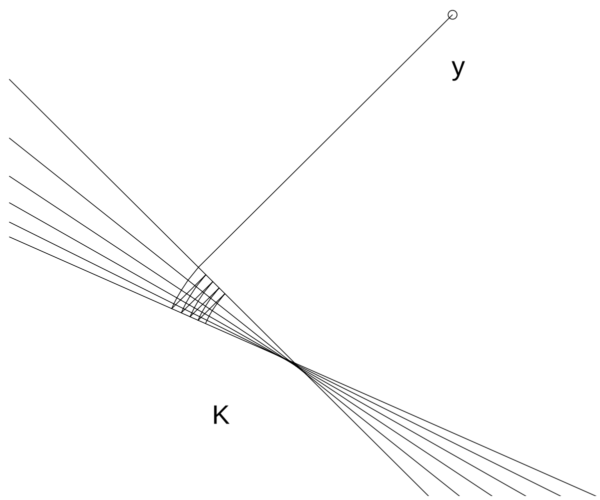

Figure 3: Nearly Collinear Subspaces

tion onto the intersection of a set of concavity restrictions. This is analogous to the projection for costs used by Goldman & Ruud (1993a). Secondly, we reduce the active constraint set by removing as many constraints as possible through periodic optimization over all positive elements of $\boldsymbol{\lambda}$. This procedure is a generalization of the attempt to solve for a final solution.

3.1 Projection onto Concavity Constraints

Our experience shows that the greater speed of such simple projections as (11) can be overcome by the greater improvements achieved by more complex projections onto higher dimensional cones. In our case, we make projections onto the sequence of cones

$$K_r \equiv \{p_r' z_r \leq p_r' z_i \quad \forall i : q_r \leq q_i\}\,,\ r = 1, \ldots, S,$$

the intersections of subsets of the half spaces Goldman & Ruud (1993a) consider. The projection onto this cone is nonlinear, but it can be accomplished efficiently using a conventional quadratic programming algorithm (see Gill, Murray, Saunders & Writhgt (1982)).[15] This is possible because the special

[15]Here is a brief description. Given a feasible starting point for an iteration, the quadratic programming algorithm computes the solution to the restricted least squares

structure of these cones allows us to compute the search direction without numerical inversion of a large Hessian term. The constraints in K_r have a convenient tensor form by virtue of the common price vector p_r. The constraints can always be written in the form

$$(R_r \otimes p_r')\,\mathbf{z} = 0$$

by stacking the fitted factor demand vectors into a single vector. For notational convenience, suppose all the elements of $\mathbf{z}$ are actively constrained elements and suppose z_r is the last element of $\mathbf{z}$. Then

$$R_r = \begin{pmatrix} -I & \boldsymbol{\iota} \end{pmatrix},$$

where I is an identity matrix and ι is a vector of ones. The search direction for an iteration of the quadratic programming algorithm is similar to (11):

$$\mathbf{z} = \mathbf{w} - \frac{1}{p_r' p_r}\left[p_r p_r' \left(w_i - \bar{w}\right)\right]$$

where $\bar{w}$ is the simple average of the elements of w and w_i is the i^{th} element of $\mathbf{w}$. Clearly, this direction can be computed easily for large numbers of active constraints. One simply computes an average of conditional demand vectors. One must iterate through a sequence of such search directions, but the cost of these iterations is overcome by the improvement in search efficiency.

The monotonicity constraints can also be conveniently combined. The projection joint simply replaces every negative entry with zero. The simplicity of this intersection arises from the mutual orthogonality of the constraints. This orthogonality also implies what is obvious here: that there is no gain in the efficiency of the algorithm derived from forming this "monotonicity" cone.

3.2 Optimal Constraint Elimination

There is a general pattern in the paths of the Gauss-Seidel iterations in the dual: The number of active constraints usually declines, especially over

(RLS) problem constructed from imposing the active constraints as equalities. If it satisfies all constraints, then this solution also solves the projection problem. If not, one computes the point on the line segment between the initial value and the RLS solution closest to the initial value where a constraint changes status. The fit at this point becomes the starting value for the next iteration.

the initial iterations. The approach to the optimal solution is monotone in the length of the fitted vector. As one approaches the optimal solution, constraints that are satisfied at the solution are eliminated. Therefore, a general strategy to accelerate Hildreth's procedure seeks ways to eliminate constraints from the current basis.

We have found a rapid and convenient method for such elimination. When the set of active constraints remains unchanged for two successive iterations, we attempt to jump to a final solution by computing the constrained optimum that imposes all active constraints as equalities. Denoting the active constraints by $\mathbf{Rz} = 0$, this point is

$$\begin{aligned} \mathbf{y}^* &= \left(I - \mathbf{R}^- \mathbf{R}\right) \mathbf{y} \\ &= \left(I - \mathbf{R}' \left(\mathbf{R}^-\right)'\right) \mathbf{y} \\ &= \left(I - \mathbf{R}' \left(\mathbf{R}'\right)^-\right) \mathbf{y} \end{aligned}$$

where $\mathbf{R}^-$ is a generalized inverse of $\mathbf{R}$. The corresponding value for the Lagrange multiplier vector $\boldsymbol{\lambda}$ is, therefore, $\boldsymbol{\lambda}^* = \mathbf{R}'^- y$. If the active constraints do not comprise the set of constraints binding at the optimum, but instead include extra constraints, then elements of $\boldsymbol{\lambda}^*$ will be negative. The optimal, constrained, point on the line segment between the current value of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^*$ will set one element of $\boldsymbol{\lambda}$ to zero, effectively eliminating one active constraint and improving the objective function. We repeat this process with the remaining constraints, until $\boldsymbol{\lambda}^*$ contains strictly positive values.

This procedure does not produce a projection onto the intersection of the active constraints. That would require the entry of inactive constraints into the active constraint set. We have not yet explored whether the computational effort would be worthwhile. Our current procedure is rapid and speeds convergence significantly.

3.3 Additional Considerations

Finally, we remark that numerical round-off errors can play a significant role in preventing successful iteration. To summarize our experience, we find it critical to parameterize the computational problem in terms of the Lagrange multiplier vector $\boldsymbol{\lambda}$. Dykstra (1983), for example, uses a theoretically equivalent parameterization in the primal parameter space that often failed us as we developed our algorithm and software. Dykstra writes the Gauss-Seidel sequence as

$$\begin{aligned} \hat{\mathbf{y}}_i &= P\left(\hat{\mathbf{y}}_{i-1} - \hat{\mathbf{u}}_{i-S} \mid K_r\right), \\ \hat{\mathbf{u}}_i &= \hat{\mathbf{y}}_i - \left(\hat{\mathbf{y}}_{i-1} - \hat{\mathbf{u}}_{i-S}\right), \end{aligned}$$

where S is the number of constraint sets K_r, $r = (i - 1 \,\mathrm{modulo}\, S) + 1$, and $P(\mathbf{z} \mid K)$ denotes the orthogonal projection of $\mathbf{z}$ onto K.[16] This formulation has theoretical appeal for generalizations of Hildreth's approach to other optimization problems. However, in practice numerical round-off errors in the $\hat{\mathbf{u}}$ may accumulate so that the fitted value $\hat{\mathbf{y}}_i$ does not satisfy the constraints of the dual problem. Therefore, we retain the parameterization of the problem in terms of the dual, thereby ensuring that these constraints are respected at every step of the calculations.

We have written our programs in Matlab. Although Gauss is generally very similar, Matlab has an ability to handle sparse matrices that is particularly convenient for our algorithm. The constraint matrix $\mathbf{R}$ contains many zeros, because the constraints are pair-wise in the observations. Sparse matrix routines save a great deal of work space and computational time.

4 Examples

We give two examples as applications of the computation of restricted least squares. The first is based on the classic paper by Christensen & Greene (1976), one of the earliest applications of the translog cost function to cost function estimation. Goldman & Ruud (1993*a*) also used this example, but made some computational errors which are corrected here. The second example examines U.S. trucking costs, one of the areas in which cost function estimation has been applied extensively.

4.1 Electric Power Generation

Christensen & Greene (1976) estimated a translog cost function for electricity generation as a function of three factor prices: prices for capital, labor and fuel. Their primary interest was economies of scale. Using seemingly unrelated regressions, they tested the cost function for homogeneity and found

[16] This equivalence requires that one begin with $\hat{\mathbf{y}}_0 = \mathbf{y}$, which corresponds to setting $\hat{\mathbf{u}}_i = 0$ for $i = 1 - S, \ldots, 0$.

convincing evidence against these restrictions. We use the 99 observations from the original Christensen/Greene data set that Berndt (1991) provides. The imposition of homogeneity makes no appreciable difference in the calculations of scale economies. If one checks for concavity at the data points, 12 observations are at points where the translog cost function is not concave. All monotonicity restrictions are satisfied. All in all, this is an application where the translog specification appears to be successful.

We fit the cost function with least squares applied to conditional factor demands. To account for differences in scale, the factor levels were scaled by their empirical standard deviations. We made no attempt to exploit cross-equation covariance in estimation. The quadratic programming problem has 4,852 constraints. Our algorithm converged in 11 iterations, which take less than five minutes on a 486-33 PC microcomputer. No monotonicity constraints are binding. Only 54 of the concavity constraints were binding. Out of the 99 observations, 22 have perfect fits. The sample standard deviations for the residuals in the translog share equations are 0.042, 0.051, and 0.064 for labor, capital, and fuel respectively. For the nonparametric fit, these standard deviations are 0.008, 0.010, and 0.108. As one might expect, the fit of the nonparametric model is much tighter. If we make a ball park adjustment for degrees of freedom, by treating the number of observations minus the active constraints (54) as the number of parameters, the nonparametric fit remains substantially tighter.

In Figures 4 and 5, we graph the fitted average cost functions for the two specifications, computed at the sample average of the prices. The nonparametric fit has two lines, which are the upper and lower bounds of the fitted correspondence. The bulk of the output levels actually observed in the data set are below 20 billion KWH (see Figure 6) so that the second figure, graphing the logarithm of output, gives an expanded view of that region of the average cost function. Without statistical distribution theory, we cannot make a formal statement about the similarity of these functions but we think that there is substantial similarity. The estimated translog function does not appear to be misleading. This occurs in a data set where the usual diagnostics suggest confidence in the translog parameterization is justifiable.

4.2 Transportation by Truck

Our second illustration is somewhat more ambitious in scale. We re-examine the estimation of cost functions for U. S. trucking firms. This industry has

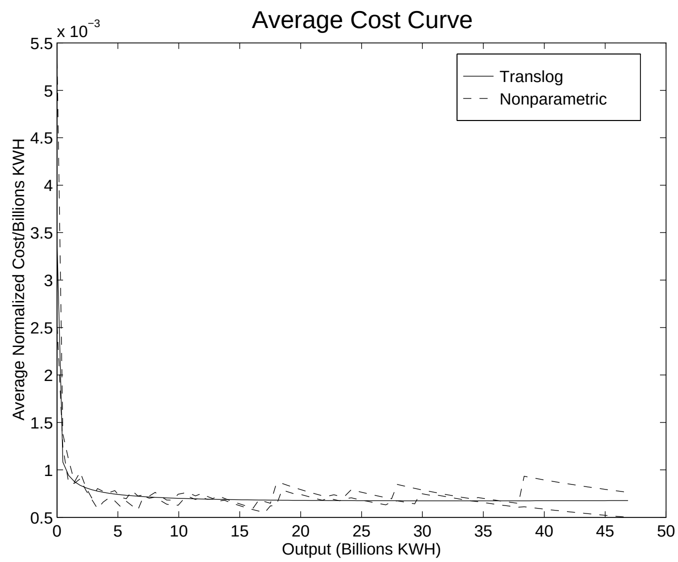

Figure 4: Average Costs for Electric Power Generation

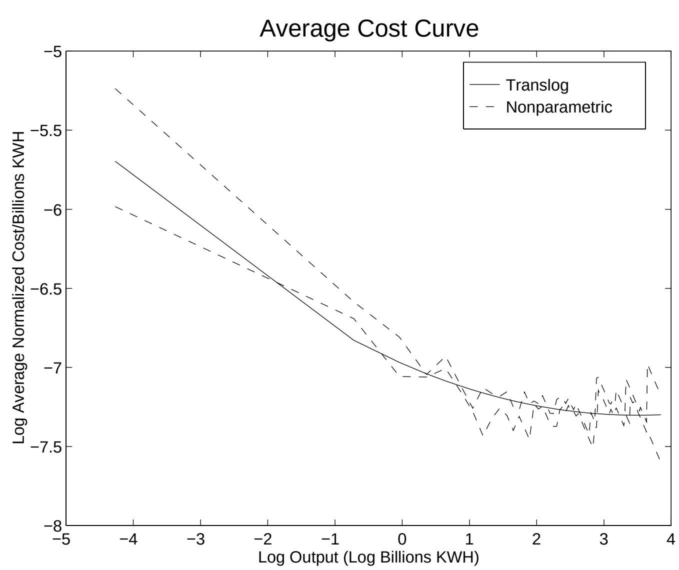

Figure 5: Average Costs for Electric Power Generation

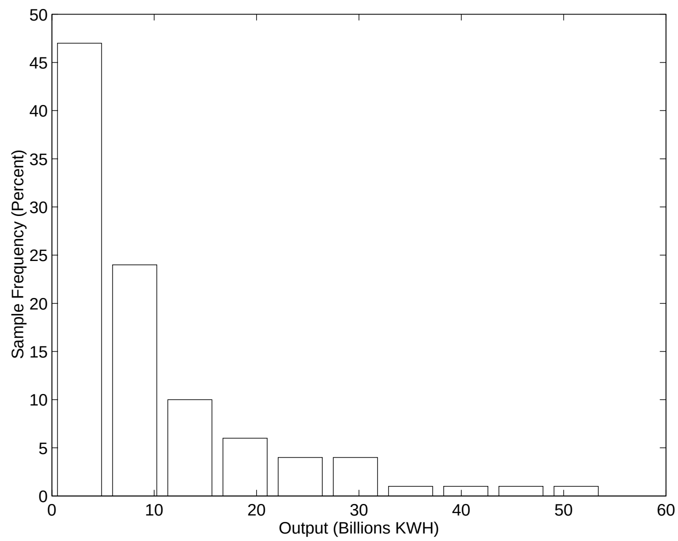

Figure 6: Size Distribution of Electric Utilities

received repeated attention because it has been heavily regulated in the past, so that data describing its costs, outputs, and factor inputs are relatively easy to obtain. Also, recent deregulation gives economists an opportunity to assess its effects on costs. Our reference point is work by Ying (1990), who references many of the earlier studies.

Our data set is a cross-section of trucking firms observed in 1976. This is the first year of a panel that we have been constructing for joint research with Goldman and Keeler on the effects of deregulation on this industry. There are 630 observations and we have followed Ying's methods to construct the data set and to estimate a translog cost function. The data contains Class I and II common carriers of general freight specializing in relatively small "less-than-truckload" shipments. The data are taken from *Trinc's Blue Book of the Trucking Industry*. The factor inputs are fuel, purchased transportation, labor, and capital. Output is measured by revenue ton-miles, although the translog specification includes average length of haul, average shipment size, and average load as additional explanatory variables.

Translog estimates of the cost function suggest that this parametric specification is inappropriate. Although no monotonicity restrictions in output are violated in the sample, 67% of the observations are at points where the estimated cost function is not concave and 9% are fail monotonicity in factor prices. The likelihood ratio test for cross-equation constraints on parameters rejects these restrictions at all conventional levels of significance.

One of the reasons for the failures in monotonicity is that one factor, purchased transportation, is often not used: 226 observations (36%) in the sample do not use purchased transportation. As a result, the translog is fitting a substantial portion of the sample in the vicinity of a factor share equal to zero. It is inevitable that some of these observations will predict negative shares, and thereby violate monotonicity in a factor price. The nonparametric fit has no such difficulty. Factor levels of zero can be a natural outcome of cost minimization and correspond to directions in which the cost function is flat.

Another cause of the failures to meet cost function restrictions may be the simplification that output is the scalar ton-miles. Researchers have routinely added such variables as "average length of haul" to the translog specification to account for differences in shipping environment. These attribute variables also appear to measure such other characteristics of output as the number of trips. Therefore, we consider treating them as additional components of a multi-dimensional output vector.

We fit the nonparametric conditional factor demand functions using the original output of ton-miles and specifying output to be the vector tons, miles, and trips. The average cost functions, translog and nonparametric, for the scalar output measure ton-miles are shown in Figure 7 and a histogram of firm sizes in Figure 8. In this case there appears to be some disagreement between the two specifications. For most of the lower output levels, there is evidence that the translog is under-estimating average costs and understating the potential economies of scale.

The programming problem had 198,135 concavity restrictions and 2,520 monotonicity constraints. The computer program took 44 iterations and approximately 30 hours on a Sun Sparc 10 to converge. We do not give precise timing because the workstation was not dedicated to this one task. If it were, the duration would be this order of magnitude.

It is interesting to find that the nonparametric fit does not appear to over-fit in this case. There are only 13 observations that obtain perfect fits. There are 1138 active constraints in the nonparametric fit, 98 of which are monotonicity constraints. The translog fit has smaller error sums of squares for the cost and share equations, while the nonparametric fit has smaller error sums of squares for the factor demand levels. The statistical evidence against the translog cost specification is consistent with a constrained nonparametric

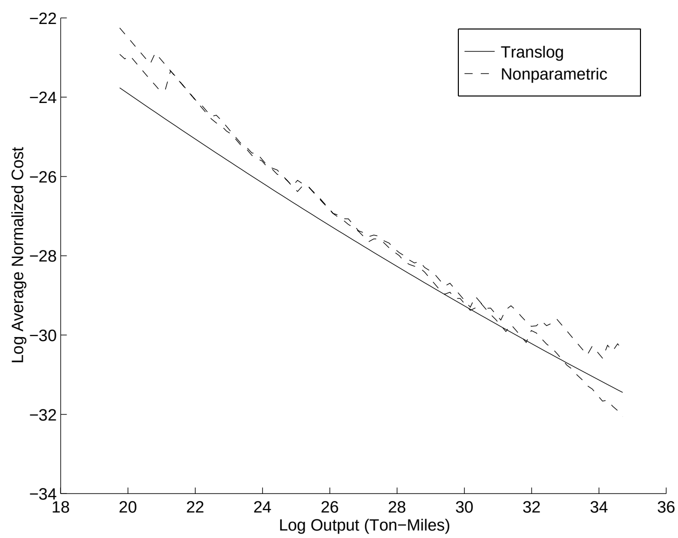

Figure 7: Average Costs for 1976 Trucking (Output Ton-Miles)

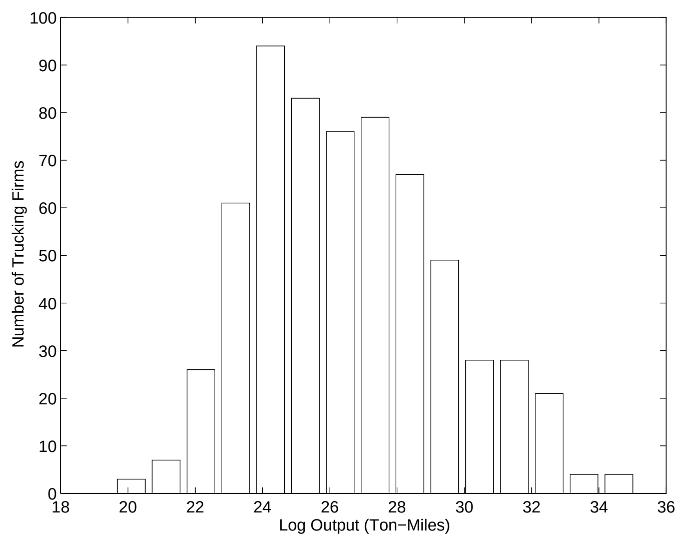

Figure 8: Size Distribution of 1976 Trucking

fit that is looser as we observe here.

The presence of binding monotonicity constraints is worth special note because Varian (1985) does not consider these constraints in his work. This reflects a difference in approach. Varian considers *testing* for cost minimization in the presence of observation error, where the observed factor demands are always positive. We are fitting a cost function with concavity constraints that can push the least squares fit into violations of monotonicity. Thus, the monotonicity constraints may constrain the fit.

We also fit the nonparametric conditional factor demand functions, specifying output to be three-dimensional: tons, miles, and trips. The scalar ton-mile specification has obvious appeal as a physical measure of work, but this is by no means definitive. As we have already noted, previous studies add additional explanatory variables that one can interpret as measures of other dimensions to output. The program had 127,838 concavity constraints and 2,520 monotonicity constraints. The generalization of the output specification reduces the number of constraints relative to the scalar specification. Convergence required 22 iterations and approximately 10 hours. At convergence, only 359 concavity constraints and 92 monotonicity constraints were binding. One third (207 of 630) of the observations have a perfect fit.

In Figure 9, we examine the possibility that the ton-miles specification may be too restrictive. We graph the average cost correspondences for two paths in output space, a "low ton" and a "high ton" path, with corresponding levels of ton-miles. The two paths are shown in Figure 10, along with a scatter plot of the actual tons and miles data. Trips are fixed at the level of the sample average, as are prices. The evidence is much less clear here, relative to the comparison of the translog and nonparametric fits earlier. There is substantial agreement between these two average cost fits. Nevertheless, the fits indicate that at low ton-miles the higher tons is more costly than higher miles. Given the relatively fixed costs of loading each truck, this is a sensible outcome.

5 Conclusion

The immediate econometric application of least squares estimation subject only to monotonicity and concavity constraints is limited by the lack of a statistical theory. The consistency of such estimators has been established, but a general approximate distribution theory remains to be found. There

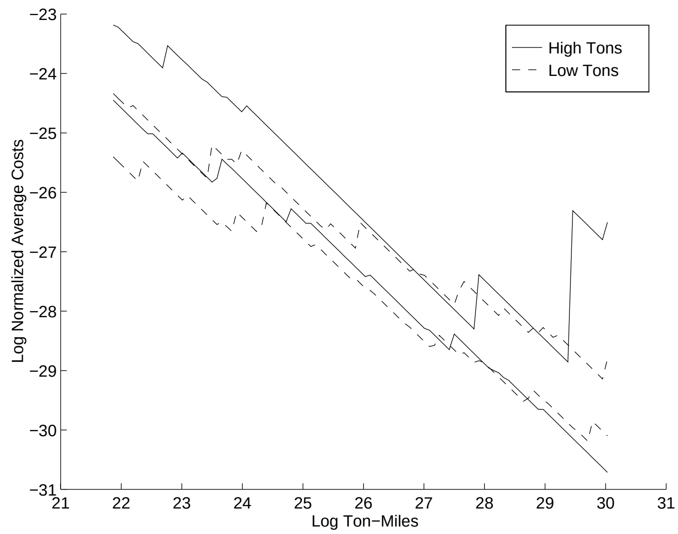

Figure 9: Average Costs for 1976 Trucking

are at least two directions in which this computational research may link with a distribution theory.

First of all, it seems fruitful to combine the local smoothing approaches of nonparametric estimation with restricted estimation. Local smoothing will surely provide convergence rates comparable to those for unrestricted estimators. There are several ways to combine smoothing with restricted estimation, all of which use the methods developed here. The simplest may be to apply the restricted least squares program to the smoothed regression, rather than to the data as we do here. An attractive alternative is to change the least squares objective function (6) to a smoothed version:

$$\hat{y} = \arg\min_{z \in \mathbf{C}} \sum_{n=1}^{N} \sum_{i=1}^{N} K_h \left(p_n - p_i, q_n - q_i\right) \left\|y_i - z(p_n, q_n)\right\|^2 \qquad (12)$$

where $K_h(\cdot)$ is a multivariate kernel density. This is equivalent to a weighted least squares problem which is easily accommodated by our approach.

Secondly, the computation of Bayesian posterior moments (Geweke (1995)) may be enhanced by our methods. The least squares fit may provide a good approximation to the mode of the posterior of a Bayesian non-parametric analysis of these estimation problems. In that case, the central tendency of the posterior can be located with our technique, providing a useful start-

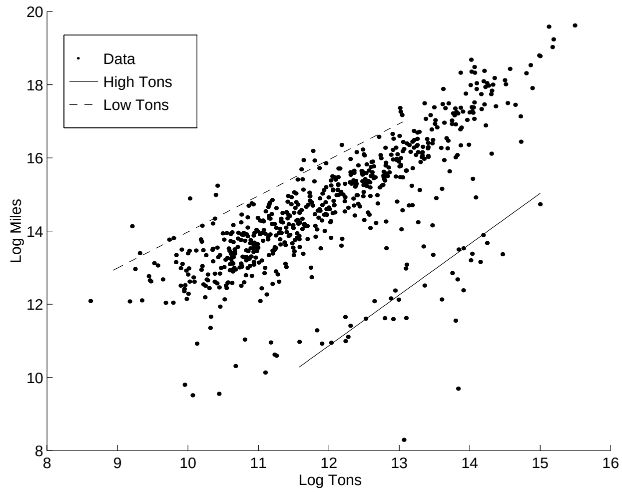

Figure 10: Path through Output of Trucking Firms

ing point for such posterior simulators as the Gibbs sampler. The latter is natural for restricted least squares problems where the support of the posterior distribution is truncated by the restrictions. See Geweke (1995) for an introduction.

References

Berndt, E. R. (1991), 'The practice of econometrics: Classic and contemporary'.

Christensen, L. R. & Greene, W. H. (1976), 'Economies of scale in u. s. electric power generation', *Journal of Political Economy* **84**(4), 655–676.

Diewert, W. E. (1973), 'Functional forms for profit and transformation functions', *Journal of Economic Theory* **6**, 284–316.

Diewert, W. E. & Wales, T. J. (1987), 'Flexible functional forms and global curvature conditions', *Econometrica* **55**(1), 43–68.

Dykstra, R. L. (1983), 'An algorithm for restricted least squares', *Journal of the American Statistical Association* **78**(364), 837–842.

Geweke, J. (1995), Posterior simulators in econometrics. Paper prepared for invited symposium, Seventh World Congress of the Econometric Society, Tokyo.

Gill, P. E., Murray, W., Saunders, M. A. & Writhgt, M. H. (1982), The design and implementation of a quadratic programming algorithm, Technical Report SOL 82-7, Department of Operations Research, Stanford University.

Goldman, S. M. & Ruud, P. A. (1993*a*), Nonparametric multivariate regression subject to monotonicity and convexity constraints. mimeograph.

Goldman, S. R. & Ruud, P. A. (1993*b*), Nonparametric multivariate regression subject to constraint, Technical Report 93-213, Department of Economics, University of California, Berkeley.

Hanson, D. L. & Pledger, G. (1976), 'Consistency in concave regression', *Annals of Statistics* **4**(6), 1038–1050.

Hildreth, C. (1954), 'Point estimates of ordinates of concave functions', *Journal of the American Statistical Association* **49**, 598–619.

Lau, L. J. (1978), *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2, North-Holland, Amsterdam, chapter Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints.

Matzkin, R. L. (1992), Computation and operational properties of nonparametric concavity-restricted estimators, Technical report, Northwestern University.

Murtagh, B. & Saunders, M. (1967), MINOS: large scale nonlinear programming system (for problems with linear constraints), Technical Report SOL 77-9, Department of Operations Research, Stanford University.

Varian, H. R. (1984), 'The nonparametric approach to production analysis', *Econometrica* **52**(2), 579–597.

Varian, H. R. (1985), 'Non-parametric analysis of optimizing behavior with measurement error', *Journal of Econometrics* **30**(1/2), 445–458.

von Neumann, J. (1950), *Functional Operators*, Vol. II, Princeton University Press, Princeton, N.J. This is a reprint of lecture notes first distributed in 1933.

Yazhen (1992), Find Out, PhD thesis, Department of Statistics, University of California at Berkeley.

Ying, J. S. (1990), 'The inefficiency of regulating a competitive industry: productivity gains in trucking following reform', *Review of Economics and Statistics* **72**(2), 191–201.