IMPOSING ECONOMIC CONSTRAINTS IN NONPARAMETRIC REGRESSION: SURVEY, IMPLEMENTATION AND EXTENSION

DANIEL J. HENDERSON AND CHRISTOPHER F. PARMETER

Abstract. Economic conditions such as convexity, homogeneity, homotheticity, and monotonicity are all important assumptions or consequences of assumptions of economic functionals to be estimated. Recent research has seen a renewed interest in imposing constraints in nonparametric regression. We survey the available methods in the literature, but focus on a particular estimator by Hall and Huang (2001) which is easily generalized to other nonparametric settings. We discuss its computational implementation in the face of linear constraints and how it can be extended to handle nonlinear constraints. Finally, we include a small simulation study to showcase the method.

1. Introduction

While imposing economic constraints, such as monotonicity and convexity, in nonparametric econometric models has a storied past, recent research has seen a renewed interest in alternative estimation methods for this problem. Economic conditions such as convexity, homogeneity, homotheticity, and monotonicity are all important assumptions or consequences of assumptions of economic functionals to be estimated. Nonparametric methods are desirable from an applied standpoint since economic theory rarely provides closed form solutions to the structural equations of interest. However, the ability to restrict a model to mimic the economic theory underlying its creation is also a key feature of choosing an estimation method so that appropriate tests of model assumptions may be generated. Thus, having a nonparametric model that is easily constrained at a researcher's disposal is warranted.[1]

In empirical studies on games, such as auctions, monotonicity of players strategies is a key assumption used to derive the equilibrium solution. This monotonicity assumption thus carries over to the estimated equilibrium strategy. And while parametric models of auction have monotonicity 'built-in', their nonparametric counterparts impose no such condition. Thus, using a nonparametric estimator of auctions that allows monotonicity to be imposed is more competitive against parametric alternatives than an estimator that ignores this condition. Recently, Henderson, List, Millimet, Parmeter and Price (2008) have shown that random samples from equilibrium bid distributions can produce non-monotonic nonparametric estimates for small samples. This suggests that being

Date: November 26, 2008.

Key words and phrases. Constraint Weighted Bootstrapping, Reordering, Hessian, Concavity, Identification.

The research on this project has benefitted from the comments of participants in seminars at Cornell Univeristy, UNLV, UC Merced, and UC Riverside as well as participants at the the fifth annual Advances in Econometrics Conference held at Louisiana State University.

[1]An additional benefit of imposing constraints in a nonparametric framework is that it may provide nonparametric identification, see Matzkin (1994).

able to construct an estimator that is monotonic from the onset is important for analyzing auction data.

Analogously, convexity is a theoretically required condition for either a production or cost function and the ability to impose this constraint in a nonparametric setting is thus desirable given that very few models of production yield reduced form parametric models. Cost functions are concave in input prices and outputs, non-decreasing and homogeneous of degree 1 in input prices. Thus, estimating a cost function requires the imposition of three distinct economic conditions. To our knowledge, applied nonparametric studies that estimate cost functions do not impose these conditions. Thus, at the very least there is a loss of efficiency since these constraints are not imposed on the estimator. Moreover, since the constraints are not imposed, it is impossible to test whether these conditions are valid or not.

In general, a wide variety of constrained nonparametric estimation strategies have been proposed to incorporate economic theory within the estimation procedure. While many of these estimators are designed myopically for the issue at hand, a small but burgeoning literature has focused on estimators which can handle many arbitrary economic constraints simultaneously. Our first goal is to summarize past and recent contributions to this literature.[2] We then focus on a particular estimator by Hall and Huang (2001). The method of Hall and Huang (2001) is easily generalized to other nonparametric settings and as such we discuss its computational implementation. Of note is the recent contribution of Racine and Parmeter (2008) who extend the methodology of the monotonically constrained estimator of Hall and Huang (2001) to handle arbitrary constraints.[3] In their paper, while they mention the ability of the method to handle general constraints, their identification and examples all focus on linear (defined in the appropriate sense) restrictions. We augment their discussion by providing identification results in the face of nonlinear constraints since this knowledge will be useful for generalizations outside of the standard regression context. Finally, we showcase the method with a small simulation study.

The rest of this paper proceeds as follows: Section 2 reviews the literature on constrained nonparametric regression. Section 3 explains sequential quadratic programming which is used in the Hall and Huang (2001) setup. Section 4 discusses imposing general, nonlinear constraints, specifically concavity, using constraint weighted bootstrapping. It further shows how concavity can be implemented computationally. A small simulated exercise in included to showcase the result. Finally, Section 5 presents some concluding remarks.

2. Existing Constrained Estimators

Consider the standard nonparametric regression model

$$y_i = m(x_i) + \sigma(x_i)\varepsilon_i, \quad \text{for } i = 1, \dots, n, \tag{1}$$

[2]We know of one survey paper on these methods (Delecroix and Thomas-Agnan 2000). However, many of the estimators we plan to discuss were published after this article.

[3]We should also recognize Henderson et al. (2008) who extended the Hall and Huang (2001) methodology to develop a monotonically constrained auction estimator.

where y_i is the dependent variable, $m(\cdot)$ is the conditional mean function with argument x_i, x_i is a $k \times 1$ vector of covariates that has a two times continuously differentiable density f with compact support, $\sigma(\cdot)$ is the conditional volatility function and ε_i is a random variable with zero mean and unit variance. Our goal is to estimate the unknown conditional mean subject to economic constraints, concavity for example, in a smooth framework.

Imposing arbitrary constraints on nonparametric regression surfaces, while not new to econometrics, has not received as much attention as other aspects of nonparametric estimation (i.e., bandwidth selection). Indeed, one can divide the literature on imposing constraints in nonparametric estimation frameworks into two broad classes:

(1) Developing a nonparametric estimator to satisfy a particular constraint. Here the class of monotonically restricted estimators is a main example.
(2) Developing a nonparametric estimator that satisfies a class of constraints (either smooth or interpolated).

Our goal is to highlight the variety of existing methods and document which are appealing for imposing general constraints on kernel smoothed approaches to nonparametric regression.

2.1. **Isotonic Regression.** The first constrained nonparametric estimators were non-smooth and fell under the heading of 'isotonic regression', initially proposed by Brunk (1955). Brunk's (1955) estimator was a minmax estimator that was designed to impose monotonicity on a regression function with a single covariate, while Hansen, Pledger and Wright (1973) extended the estimator to two dimensions and provided results on consistency of the estimator. To explain the estimator of Brunk, let $\mathcal{C}_B$ be the discrete cone of restrictions in R^n:

$$\left\{(z_1, z_2, \dots, z_n) : z_1 \leq z_2 \leq \cdots \leq z_n\right\}.$$

We let y_i^* be a solution to the minimization problem

$$\min_{(y_1^*,\dots,y_n^*)\in\mathcal{C}_B} \sum_{i=1}^{n} (y_i - y_i^*)^2.$$

This minimization problem has a unique solution that is expressed succinctly by a min-max formula.

Use $X_{(1)}, \dots, X_{(n)}$ to denote the order statistics of X and $y_{[i]}$ the corresponding observation of $X_{(i)}$. Then our 'isotonized' fitted values can be represented as

$$y_i^* = \min_{s\geq i}\max_{t\leq i} \sum_{j=s}^{t} y_{[j]}/(t-s+1), \tag{2}$$

or

$$y_i^* = \max_{s\leq i}\min_{t\geq i} \sum_{j=s}^{t} y_{[j]}/(t-s+1). \tag{3}$$

In Brunk's (1955) approach there is no attempt to smooth the estimation results to values of x between the observation points. A simple approach would be to extend flatly between the values of x_i but this has been criticized for the presence of too many flat spots and a slow rate of convergence.[4]

Interestingly, Hildreth (1954) introduced a related method as that in Brunk (1955), but geared towards estimating a regression function that is restricted to be concave. His procedure amounts to conducting least squares subject to discretized concavity restrictions. Similar to Brunk (1955), let $\mathcal{C}_H$ be the discrete cone of restrictions in R^n:

$$\left\{(z_1, z_2, \ldots, z_n) : \frac{z_{i+1} - z_i}{x_{i+1} - x_i} \geq \frac{z_{i+2} - z_{i+1}}{x_{i+2} - x_{i+1}}, i = 1, \ldots, n-2\right\}$$

then y_i^* is a solution of.

$$\min_{(y_1^*, \ldots, y_n^*) \in \mathcal{C}_H} \sum_{i=1}^{n} (y_i - y_i^*)^2. \tag{4}$$

An iterative procedure is required to solve the minimization as no closed form solution exists. However, unlike the monotonically constrained estimator of Brunk (1955), the concave restricted estimator of Hildreth (1954) extends between observation points linearly, thus falling into the classification of a least-squares spline estimator.

While both of these estimators construct restricted regression estimates found on simple concepts, they are not 'smooth' in the traditional sense. Mukerjee (1988) and Mammen (1991a) smooth the classic isotonic regression estimator of Brunk (1955). An alternative way to characterize their estimators is to say that they forced the traditional Nadaraya-Watson regression smoother to satisfy a monotonicity constraint. The key insight was to use a two step estimator that consisted of a smoothing step and an isotonizing step. Mukerjee (1988) proved that one could preserve the isotonization in the first step by using a log-concave kernel in the smoothing of the second step. Thus, after one uses either (2) or (3) to isotonize the regressand, a smooth, nonparametric estimate of the unknown conditional mean is constructed as

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K((x - X_{(i)})/h_n) y_i^*}{\sum_{i=1}^{n} K((x - X_{(i)})/h_n)}, \tag{5}$$

where h_n is the bandwidth. One does not need to use a special kernel as a second order Gaussian kernel is log concave, thus making this method easy to implement. Mammen (1991a) proved that asymptotically the order of the steps is irrelevant. No equivalent estimator exists for the concave variant introduced by Hildreth (1954) and as such the generalizability of smoothing isotonic type estimators is unknown. Moreover, multivariate extensions to the traditional isotonic regression estimator are difficult to implement and often not available in closed form solutions.

[4] Slower than conventional nonparametric rates.

2.2. **Constrained Spline/Series Estimation.** Both spline and series based functions provide the researcher with a flexible set of basis functions with which to construct a regression model that is linear in parameters, which is intuitively appealing. Early methods using splines or series based methods, designed to impose general economic constraints, include Gallant (1981, 1982) and Gallant and Golub (1984) who introduced the Fourier Flexible Form estimator (FFF), whose coefficients could be restricted to impose concavity, homotheticity and heterogeneity in a nonparametric setting.[5] Constrained spline smoothers were proposed by Dierckx (1980), Holm and Frisen (1985), Ramsay (1988), and Mammen (1991b), to name a few early approaches.

In what follows we describe the basic setup for constrained least squares spline estimation.[6] We define our spline space to be $\mathcal{S}$ which has dimension p.[7] Our least squares spline estimate is a function m which represents a linear combination of spline functions from $\mathcal{S}$ that solves:

$$\min_{s\in\mathcal{S}} \sum_{i=1}^{n} (y_i - m(x_i))^2. \tag{6}$$

To impose constraints we note that positivity of either the first or second derivative at a given point $\tilde{x}$ of the function $m(\cdot)$ can be written equivalently as positivity of a linear combination of the associated parameters with respect to the chosen basis. Thus, monotonicity or concavity can be readily imposed on a discretized grid of points where each point adds additional *linear* constraints on the spline coordinates with the associated basis. It is a natural step to include these linear constraints directly into the least squares spline problem.

Similar to isotonic regression, the literature appears to have focused on concavity first Dierckx (1980) and then monotonicity Ramsay (1988). In what will seen to be a common theme in constrained nonparametric regression, Dierckx (1980) used a quadratic program to enforce *local* concavity or convexity of a spline function. His function estimate, using normalized B-splines (Schumaker 1981) with basis N_j, is

$$\hat{m}(x) = \sum_{j=-3}^{k} c_j^* N_j(x).$$

Here k denotes the total number of knots. The values c_j^* solve the quadratic program

$$\min_{\sum_{j=-3}^{k} d_{j,l} c_j e_j \le 0} \sum_{i=1}^{n} \left(y_i - \sum_{j=-3}^{k} c_j N_j(x_i) \right)^2 . \tag{7}$$

The e_j in equation (7) determine the type of constraint being imposed on the function locally. That is, $e_j = 1$ if the function is locally convex at knot ℓ. $e_j = 0$ if the function is unrestricted at the ℓ^{th}

[5]Monotonicity is not easily imposed in this setting.

[6]For a more detailed treatment of either series or spline based estimation we refer the reader to Eubank (1988) and Li and Racine (2007, chapt. 15).

[7]Unlike kernel smoothing where smoothing is dictated by a bandwidth, in series and spline based estimation, the smoothing is controlled by the dimension of the series or spline space.

knot and $e_j = -1$ if the function is locally concave at knot ℓ. The numbers $d_{j,l}$ are derived from the second derivatives of the basis splines at each of the knots have have simple formula. We have

$$\begin{aligned} d_{j,l} =& 0 && \text{if} \quad j \le l-4 \quad \text{or} \quad j \ge 4 \\ d_{l-3,l} =& \frac{6}{(t_{l+1}-t_{l-2})(t_{l+1}-t_{l-1})} \\ d_{l-1,l} =& \frac{6}{(t_{l+2}-t_{l-1})(t_{l+1}-t_{l-1})} \\ d_{l-2,l} =& -(d_{l-3,l}+d_{l-1,l}), \end{aligned}$$

where t_l refers to the l^{th} point under consideration. Ramsay (1988) developed a similar monotonically constrained spline estimator using I-splines. I-splines have a direct link to the B-splines used by Dierckx (1980). An I-spline of order M is an indefinite integral of a corresponding B-spline of the same order. Ramsay (1988) used I-splines because he was able to establish that they had the property that each individual I-spline is monotonic and that any linear combination of I-splines with positive coefficients is also monotonic. This made it easy to construct the associated monotonic spline estimator. Both of the aforementioned estimators can also be placed in the smoothing spline domain as well.

Yatchew and Bos (1997) develop a series based estimator that can handle general constraints. This estimator is constructed my minimizing the sum of squared errors of a nonparametric function relative to an appropriate Sobolev norm. The basis functions that make up the series estimation are determined from a set of differential equations that provide 'representors'. Representors of function evaluation consist of two functions spliced together, where each of these functions is a linear combination of trigonometric functions. In essence, one can 'represent' any function in Sobolev space through this process (Yatchew and Bos 1977, Appendix 2). Let R be an $n \times n$ 'representor' matrix whose columns (equivalently rows) equal the representors of the function, evaluated at the observations $x_1, \ldots, x_n$.[8] Then, arbitrary constrained estimation of a nonparametric function

$$\min_{f \in \mathcal{F}} \quad n^{-1} \sum_{i=1}^{n} (y_i - m(x_i))^2 \quad \text{s.t. } \|m\|^2_{Sob} \le L,$$

can be recast as

$$\min_{c} \quad n^{-1} \sum_{i=1}^{n} (y_i - Rc)^2 \quad \text{s.t. } c'Rc \le L, c'R^{(1)}c \le L^{(1)}, c'R^{(2)}c \le L^{(2)}, \ldots, c'R^{(k)}c \le L^{(k)}. \tag{8}$$

Here L denotes the upper bound on the squared Sobolev norm of our constrained function, c is an $n \times 1$ vector of coefficients and $\mathcal{F}$ is the constrained function space which we are searching over. Since we are interested in constraints that relate directly to the derivatives of the nonparametric function we are estimating, $R^{(1)}, \ldots, R^{(k)}$ represent the appropriate derivatives of the original representor matrix and $L^{(1)}, \ldots, L^{(k)}$ are the corresponding bounds. For example, if one wished to impose monotonicity, $L^{(1)} = 0$ and $R^{(1)}$ represents the representor matrix with each of the representors

[8]For more on the construction of representor matrices see Wahba (1990) or Yatchew and Bos (1997).

first order differentiated with respect to the corresponding column's variable (i.e., the fifth column of $R^{(1)}$ corresponds to the fifth covariate so the representors are first order differentiated with respect to that variable). Again, this is a quadratic programming problem with a quadratic constraint.[9]

Beresteanu (2004) introduced a spline based procedure that can handle multivariate data and impose multiple, general, derivative constraints. His estimator is solved via quadratic programming over an equidistant grid created on the covariate space. These points are then interpolated to create a globally constrained estimator. He employed his method to impose monotonicity and supermodularity of a cost function for the telephone industry. His estimation setup is similar to the approaches described above and involves setting up a set of appropriately defined constraint matrices for the shape constraint(s) desired and solving for a set of coefficients, then interpolating these points to construct the nonparametric function which satisfies the constraints over the appropriate interval. In essence, since Beresteanu (2004) is constructing his estimator first based on a grid of points and then interpolating, this estimation procedure can be viewed as a two-step series based equivalent of the isotonic regression discussed earlier Mukerjee(1988).

2.3. **The Matzkin Approach.** The seminal work of Matzkin (1991, 1992, 1993, 1994, 1999) considered identification and estimation of general nonparametric problems with arbitrary economic constraints. One of her pioneering insights was that when nonparametric identification was not possible, imposing shape constraints tied to economic theory could provide nonparametric identification in certain estimation settings. Her work laid the foundations for a general operating theory of constrained nonparametric estimation. Her methods focused on standard economic constraints (monotonicity, concavity, homogeneity, etc.) but facilitated in more general settings than regression. Primarily, her work focused on binary-threshold crossing models and polychotomous choice models, although her definition of sub-gradients equally carried over to a regression context. One can suitably recast her estimation method in the regression context as nonparametric constrained least squares.

For example, to impose concavity on a regression function she created 'subgradients', T^j, which were defined for any convex function $m : X \to \mathbb{R}^k$ where $X \subset \mathbb{R}$ is a convex set and $x \in X$ any vector $T \in \mathbb{R}^k$ such that $\forall y \in X \; m(y) \geq m(x) + T(y - x)$.[10] We use the notation T^j to denote that the subgradients are calculated for the observations. Matzkin (1994) showed how to use the subgradients to impose concavity and monotonicity simultaneously. Using the Hildreth (1954) constraints for concavity of a regression surface, Matzkin (1994) rewrites them as

$$m(x_i) \leq m(x_j) + T^j(x_i - x_j), \quad i, j = 1, \ldots, n.$$

She solves the minimization problem in (4) but the minimization is over $m(x_i)$ $\forall i$ and T^j $\forall j$. To impose monotonicity one would add the additional constraint that $T^j > 0$ $\forall j$. Algorithms to

[9]See the work of Yatchew and Härdle (2006) for an empirical application of constrained nonparametric regression using the series based method of Yatchew and Bos (1997). Yatchew and Härdle (2006) focus on nonparametric estimation of an option pricing model where the unknown function must satisfy monotonicity and convexity as well as the density of state prices being a true density.

[10]When $m(x)$ is differentiable at x the gradient of x is the *unique* subgradient of m at x.

solve the constrained optimization problem were first developed for the regression setup by Dykstra (1983), Goldman and Ruud (1992) and Ruud (1995) and for general functions by Matzkin (1999), who used a random search routine regardless of the function to be minimized.

Implementation of these constrained methods is of the two-step variety (see Matzkin 1999). First, for the specified constraints, a feasible solution consisting of a finite number of points is determined through optimization of some criterion function (in Matzkin's choice framework setups this is a pseudo-likelihood function). Second, the feasible points are interpolated or smoothed to construct the nonparametric surface that satisfies the constraints. These methods can be viewed in the same spirit as that of Mukerjee (1988), but for a more general class of problems.

2.4. **Rearrangement.** Recent work on imposing monotonicity on a nonparametric regression function, known as rearrangement, is detailed in Dette, Neumeyer and Pilz (1996) and Chernozhukov, Fernandez-Val and Galichon (2007). The estimator of Dette et al. (2006) combines density and regression techniques to construct a monotonic estimator. The appeal of 'rearrangement' is that no constrained optimization is required to obtain a monotonically constrained estimator, making it computationally efficient compared to the previously described methods. Their estimator actually estimates the inverse of a monotonic function, which can then be inverted to obtain an estimate of the function of interest.

To derive this estimator let M denote a natural number that dictates the number of equi-spaced grid points to evaluate the function. Then, their estimator is defined as

$$\widehat{m}^{-1}(x) = \int_{-\infty}^{x} \frac{1}{Mh} \sum_{j=1}^{M} K\left(\frac{\hat{m}(j/M) - u}{h}\right) du, \tag{9}$$

where $\widehat{m}(x)$ is any unconstrained nonparametric regression function estimate (kernel smoothed, local polynomial, series, splines, neural network, etc.). The intuition behind this estimator is simple; the connection rests on the properties of transformed random variables.

Note that $m(x_i)$ is a transformation of the random variable x_i. The estimator

$$\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{m(x_i) - u}{h}\right),$$

represents the classical kernel density of the random variable $u = m(x_1)$ which has density

$$g(u) = f(x_1)|(m^{-1})'(x_1)|.$$

The integration in (9) is that of a probability density function and as such a CDF is constructed, which is always *monotonically increasing*. The equi-space grid is used for the estimation since the evaluation points are then treated as though they came from a uniform density, making $f(j/M) = I[a, b]$, where a and b denote the lower and upper bounds of the support of X, respectively. Thus, the integration in this case amounts to integrating $|(m^{-1})'(x_1)|$ over its domain, which gives us $m^{-1}(x_1)$. Once this has been obtained, it is a simple matter to reflect this estimate across the $y = x$ line in Cartesian 2-space to obtain our monotonically restricted regression estimator. Chernozhukov et al. (2007) discuss implementation of this estimator in a multivariate setting and show that the constrained estimator *always* improves over an original estimate whenever the original estimate is not monotonic.

The name rearrangement comes from the fact that the point estimates are rearranged so that they are in increasing order (monotonic). This happens because the kernel density estimate of the first stage regression estimates sorts the data from low to high to construct the density, which is then integrated. This sorting, or rearranging, is how the monotonic estimate is produced. It works because monotonicity as a property is nothing more than a special ordering and the kernel density estimator is 'unaware' that the points it is smoothing over to construct a density are from an estimate of a regression function as opposed to raw data.

One issue with this estimator is that while it is intuitive, computationally simple and easy to implement with existing software, it requires the selection of two 'bandwidths'.[11] Additionally, the intuition underlying the ease of implementation does not readily extend itself to general constraints on nonparametric regression surfaces. No such transformation is obtainable to impose concavity using the same insights, for example.

2.5. **Data Sharpening.** Data sharpening derives from the work of Friedman, Tukey and Tukey (1980) and later in Hall and Choi (1999). These methods are designed to admit a wide range of constraints and are closely linked to biased-bootstrap methods (Hall and Presnell 1999). Data sharpening is inherently different than biased-bootstrapping and constraint weighted bootstrapping (to be discussed later) as it alters the data, but keeps the weights associated with each point fixed, whereas biased-bootstrapping and constraint weighted bootstrapping change the weights associated with each point, but keep the points fixed. Both of these methods, however, can be thought of as data tuning methods which in some sense alter the underlying empirical distribution to achieve the desired outcome. We discuss the method of Braun and Hall (2001) in what follows.

Let our original data be $\{x_1, \ldots, x_n\}$ and our sharpened data be $\{z_1, \ldots, z_n\}$. Define the distance between original and sharpened points as $D(x_i, z_i) \geq 0$. We choose $\mathcal{Z} = \{z_1, \ldots, z_n\}$, our set of sharpened data to minimize

$$D(\mathcal{X}, \mathcal{Z}) = \sum_{i=1}^{n} D(x_i, z_i),$$

subject to our constraints of interest. Once the sharpened data have been obtained we apply our method of interest, in this setting nonparametric regression, to the sharpened data.

More formally, our kernel regression (local constant, say) estimator is

$$\widehat{m}(x|\mathcal{X}, \mathcal{Y}) = \frac{\sum_{i=1}^{n} K\left((x_i - x)/h\right) y_i}{\sum_{i=1}^{n} K\left((x_i - x)/h\right)} = \sum_{i=1}^{n} A_i(x) y_i.$$

[11]We use the word bandwidth loosely here as the first stage does not have to involve kernel regression. One could use series in which case the selection would be over the number of terms. Or, if one uses splines then the number of knots would have to be selected in the first stage.

We want to impose an arbitrary constraint on the function, monotonicity for example, by 'sharpening' the ys. Thus, we minimize

$$D(\mathcal{Y}, \mathcal{Q}) = \sum_{i=1}^{n} D(y_i, q_i), \tag{10}$$

for a preselected distance function, subject to the constraints

$$\widehat{m}'(x|\mathcal{X}, \mathcal{Q}) = \sum_{i=1}^{n} A_i'(x) q_i > 0. \tag{11}$$

Notice the conditioning set for which the estimator is defined over has changed from $\mathcal{Y}$ to $\mathcal{Q}$. Thus, we *construct* our restricted estimator while simultaneously minimizing our criterion function. If one chose $D(r,t) = (r-t)^2$, we would have a standard quadratic programming problem provided the constraints were linear (which they are in our monotonicity example). Given that the data are smoothed, compared to rearrangement, the corresponding constrained curve is as smooth as the unconstrained curve. This is true even though the response variables are moved around. The rearranged curve will have ambiguous low order kinks where the non-monotonic portion of the curve is 'forced' to be monotonic resulting in a curve that is less smooth than its unconstrained counterpart.

2.6. **Constraint weighted bootstrapping.** Hall and Huang (2001) suggests an alternative smooth, monotonic nonparametric estimator that admits any number of covariates. Start again with the standard local constant least squares estimator

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left((x_i - x)/h\right) y_i}{\sum_{i=1}^{n} K\left((x_i - x)/h\right)} = \frac{1}{n} \sum_{i=1}^{n} A_i(x) y_i, \tag{12}$$

where $K_i(x) = nK\left((x_i - x)/h\right) / \sum_{i=1}^{n} K\left((x_i - x)/h\right)$. Even though we are choosing to use the local constant least squares framework, this setup can be immediately extended to other types of kernel and local polynomial estimation routines. As it stands, the regression estimator in (12) is not guaranteed to produce a monotonic estimator. Hall and Huang's (2001) insight was to introduce observation specific weights p_i instead of the $1/n$ that appears in (12). These weights can then be manipulated so that the estimator satisfies monotonicity. To be clear,

$$\widehat{m}(x|p) = \sum_{j=1}^{n} p_i A_i(x) y_i,$$

is the constraint weighted bootstrapping estimator. It is still not monotonic until we properly restrict the weights.

In the unconstrained setting we have $p = (p_1, \ldots, p_n) = (1/n, \ldots, 1/n)$ which represents weights drawn from a uniform distribution. If the bandwidth chosen produces an estimate that is already monotonic, the weights should be set equal to the uniform weights. However, if the function by itself is not monotonic then the weights are diverted away from the uniform case to create a monotonic estimate. In order to decide how to manipulate the weights a distance metric is introduced based on power divergence (Cressie and Read 1984):

$$D_\rho(p) = \frac{1}{\rho(1-\rho)}[n - \sum_{i=1}^{n}(np_i)^\rho], \quad -\infty < \rho < \infty. \tag{13}$$

where $\rho \neq 0, 1$. One needs to take limits for $\rho = 0$ or 1. They are given as

$$D_0(p) = -\sum_{i=1}^{n} \log(np_i); \quad D_1(p) = \sum_{i=1}^{n} p_i \log(np_i),$$

If one uses $\rho = 1/2$, then this corresponds to Hellinger distance. This metric is minimized for a selected ρ subject to the constraint that

$$\widehat{m}'(\cdot \mid p) = \sum_{j=1}^{n} p_i A_i'(\cdot) y_i \geq \varepsilon,$$

on a grid of selected points. Here $\varepsilon \geq 0$ can be used to guarantee either weak or strict monotonicity. A nice feature of this estimator is that the kernel and bandwidth are chosen before the weights are selected. This means that the user can choose their desired kernel estimator and bandwidths selector to construct their nonparametric estimator and then constraint it to be monotonic. This leaves the door open to straightforward modification of the estimator.

2.7. **Summary of Methods.** While our discussion of existing methods has indicated a number of choices for the user, there does not exist one clear cut method for imposing arbitrary constraints on a regression surface. Each of the methods discussed has computational or theoretical drawbacks when considered against the set of all available methods.[12] The method of Hall and Huang (2001) is, in our opinion, the method most easily generalized to other nonparametric settings (quantile regression, conditional density estimation, partly linear estimation, etc.) and as such we discuss its computational implementation in the face on nonlinear constraints since this knowledge will be useful for generalizations outside of the standard regression context. Additionally, the Hall and Huang (2001) method is firmly entrenched in a kernel smoothing framework and as noted by Hall and Huang (2001, p. 625), the use of splines does not hold the same attraction for users of kernel methods. Further, recent developments that permit the kernel smoothing of categorical and continuous covariates (Li and Racine 2007) can dominate spline methods. Non-smooth methods, either the fully non-smooth methods of Allon, Beenstock, Hackman, Passy and Shapiro (2007)[13] or the litany of two-step methods discussed above are not appealing either. While rearrangement is a serious competitor, its inability to produce the same insights for multiple nonlinear constraints,

[12]Dette and Pilz (2006) conduct a Monte Carlo comparison of smooth isotonic regression, rearrangement, and the method of Hall and Huang (2001) for the constraint of monotonicity and find that their estimator has desirable finite sample performance compared to the other methods.

[13]We chose not to discuss concave programming methods as they do not directly link to a nonparametric regression setup.

omits it from our discussion's focus. Data-sharpening is also a serious competitor and its merits for the task at hand should not be overlooked. A future research agenda in this literature should focus on the exact finite and asymptotic performance of each of these methods simultaneously.

3. Sequential quadratic programming

We believe that the Hall and Huang (2001) estimator is promising because of its relatively simple extension to all types of kernel estimators as well as its relatively straightforward implementation. Obtaining the weights in their article requires the use of a quadratic programming routine. Although the steps to constructing a constrained nonparametric estimator seem straight forward, implementing these types of programs are often not discussed in detail in econometric papers. In this section we outline sequential quadratic programming (SQP).

Consider the inequality constrained problem

$$\min D(z) \text{ subject to } r_i(z) = 0,\ i \in \mathcal{E},\ \text{ and } c_j(z) \geq 0,\ j \in \mathcal{I}. \tag{14}$$

Where $D : \mathbb{R}^{qo} \to \mathbb{R}$, $r_i : \mathbb{R}^{qo} \to \mathbb{R}^{q_1}$ and $c_j : \mathbb{R}^{qo} \to \mathbb{R}^{q_2}$ can all be nonlinear. We do require that all the functions are smooth in the z argument, however. The idea behind SQP is to convert the nonlinear programming problem in 14 into a conventional quadratic programming (QP) problem. To do this we need to 'linearize' our constraints and 'quadracize' our objective function. Before doing this we introduce some additional concepts.

The Lagrangian of our problem is defined as

$$\mathcal{L}(z, \lambda_r, \lambda_c) = D(z) - \lambda_r' r_i(z) - \lambda_c' c_j(z). \tag{15}$$

Also, we define $B_r(z)' = [\nabla r_1(z), \nabla r_2(z), \ldots, \nabla r_n(z)]$ and $B_c(z)' = [\nabla c_1(z), \nabla c_2(z), \ldots, \nabla c_n(z)]$. Now pick an initial z, z_0, and an initial set of vectors of Lagrange multipliers, $\lambda_{r,0}$ and $\lambda_{c,0}$. Lastly, define $\nabla^2 \mathcal{L}_{zz}(z, \lambda_r, \lambda_c) = \nabla^2 D(z) - \nabla B_r(z)'\lambda_r - \nabla B_c(z)'\lambda_c$. We are now ready to describe how to solve our SQP problem.

Our QP at step 0 is

$$\min D(z_0) + \nabla D(z_0)'q + \frac{1}{2} q' \nabla^2_{zz} \mathcal{L}(z_0, \lambda_{r,0}, \lambda_{c,0}) q, \tag{16}$$

subject to

$$B_r(z_0)q + r(z_0) = 0 \text{ and } B_c(z)q + c(z_0) \geq 0. \tag{17}$$

The solution of this standard quadratic program, q_0, $\ell_{r,0}$, and $\ell_{c,0}$ can be used to update z_0, $\lambda_{r,0}$, and $\lambda_{c,0}$ as follows: $z_1 = z_0 + q_0$, $\lambda_{r,1} = \ell_{r,0}$, and $\lambda_{c,1} = \ell_{c,0}$. These updated values can then be plugged back into the SQP to repeat the whole process until convergence. SQP requires nothing more than repeated evaluation of the levels, first and second order derivatives of the objective and constraint functions. It is a simple matter to determine these derivatives, thus this simplification process requires nothing more than taking derivatives of a set of functions.

4. Imposing nonlinear constraints

Racine and Parmeter (2008) note that there is nothing special about monotonicity for the method of Hall and Huang (2001) to work. Any constraint that is desired, could, in principle, be imposed on the regression surface. We discuss a further generalization of Racine and Parmeter (2008) that can handle general nonlinear constraints and discuss in detail the computational method of sequential quadratic programming required to implement nonparametric regression in this setting.

To begin, note that the monotonic constraint imposed in Hall and Huang (2001) can be written in the more general form:

$$\sum_{i=1}^{n} p_i \left[\sum_{\mathbf{s} \in \mathbf{S}} \alpha_{\mathbf{s}} A_i^{(\mathbf{s})}(x) \right] y_i - c(x) \geq 0, \tag{18}$$

where the inner sum is taken over all vectors $\mathbf{S}$ that correspond to our constraints of interest (monotonicity, say) and $\alpha_{\mathbf{s}}$ are a set of constants used to generate various constraints. In what follows we shall presume, without loss of generality, that for all $\mathbf{s}$, $\alpha_{\mathbf{s}} \geq 0$. $\mathbf{S}$ indexes the order of the derivative associated with the kernel portion of the regression estimator. In our example of monotonicity, $\mathbf{s} = e_j$ is a k-vector (since we have $x \in \mathbb{R}^k$) with 1 in the j^{th} position and zeros everywhere else, $\alpha_{\mathbf{s}} = 1 \; \forall \mathbf{s} \in \mathbf{S}$ and $c(x) = 0$.[14] Racine and Parmeter (2008) provide existence and uniqueness for a set of weights for constraints of the form (18). They call these constraints linear since they are linear with respect to the weights, $p_i \; \forall i$.[15,16]

4.1. Existence and uniqueness of a solution.

Relaxing the assumption that the p_i's enter linearly requires additional assumptions for identification. Specifically, when the following assumptions hold:

(1) The constraint Jacobians, $B_r(z)$ and $B_c(z)$, have full row rank,
(2) The matrix $\nabla^2_{zz}\mathcal{L}(z, \lambda_r, \lambda_c)$ is positive definite on the tangent space of constraints,

our SQP has a unique solution that satisfies the constraints. Essentially, this result comes from the fact that one could have used Newton's method to solve the constrained optimization and the result here is obtained from the associated iterate from running Newton's method instead. These two assumptions are enough to guarantee that a unique solution holds if one were to use Newton's

[14] The notation $A^{(\mathbf{s})}$ refers to the order of the derivative of our weight function with respect to its argument.

[15] Note the subtle difference between the data sharpening methods discussed previously and the constraint weighted bootstrapping methods here. When one chooses to sharpen the data the actual data values are being transformed while the weighting is held constant. Here, the exact opposite occurs, the data is held fixed while the weights are changed. At the end of the day however, the two estimators can be viewed as 'visually' equivalent. That is, both estimators can be looked at as $\widehat{m}(x) = \sum_{i=1}^{n} A_i(x) y_i^*$,where y_i^* corresponds to either the sharpened values or $p_i y_i$ obtained from the constraint weighted bootstrapping approach. The difference between the methods is how y_i^* is arrived at. An interesting topic for future research would be to compare the performance of these two methods across a variety of constraints.

[16] Additionally, to make the constrained optimization computationally simple, they use the L_2 norm with respect to the uniform weights $(1/n)$, as opposed to the power divergence metric. This condenses the problem into a standard quadratic programming problem which can be solved using existing packages in almost all standard econometric software.

method instead of the one we outlined. However, Nocedal and Wright (2000, pgs. 531-532) show that these two procedures, in this setting, are equivalent. For more on existence of a *local* solution we direct the interested reader to Robinson (1974).

Additionally, since we have converted our general nonlinear programming problem into a QP problem, the conditions required for existence of a solution in QP problems are exactly the conditions we need to hold, at each iteration, to guarantee a solution exists in this setting. Thus, the results established in Racine and Parmeter (2008), carry over to our setting, provided our nonlinear constraints are first order differentiable in the ps and satisfy our assumptions listed above, which are easily checked. Moreover, if the forcing matrix ($\nabla^2_{zz}\mathcal{L}(q, \lambda_r, \lambda_c)$) in the quadratic portion of our quadricized objective function is positive semidefinite and if our solution q^* satisfies the set of linearized equality/inequality constraints then q^* is the unique, global solution to the problem (Nocedal and Wright 2000, Theorem 16.4). Positive semi-definiteness guarantees that our objective function is convex which is what yields a global solution. We note that this only shows uniqueness but does not guarantee a solution will even exist.

However, it should be noted that because the constraint weights are restricted to be nonnegative and sum to one, this implies that it may be difficult to impose a constraint that is 'far away' from being satisfied. In essence, the constraints imposed on the problem may be inconsistent in the sense that a weight that is nonnegative or greater than one is *needed* to satisfy the constraints of interest, the nonlinear ones. However, the conditions needed to determine how far away is 'far away' are not investigated here. Our conjecture is that the distance from an observation and the underlying function is dependent on the error process that perturbs the data generating process.

In essence the weights act as vertical scaling factors and if the amount of scaling is restricted then it can be difficult to find a solution. Hall and Presnell (1999) note the difficulty in finding the appropriately sharpened points using essentially the same technique described here in roughly 10% of their simulations. They advocated for an approach similar to a simulated annealing setup that always was able to arrive at a solution although that procedure was computational more intensive than SQP. An alternative, not followed here, would be to dispense with the power divergence metric and all constraints on the weights if no solution is found in the SQP format. In this setting one could use the L_2 norm of Racine and Parmeter (2008) and linearize (provided the nonlinear constraints are differentiable) the nonlinear constraints, again engaging in an iterative procedure to determine the optimal set of weights which can be shown to always exist in this setting.

4.2. **Concavity.** While we discuss general constrained estimation in the face of arbitrary nonlinear constraints, to cement our ideas we focus on the specific example of concavity. Concavity is a common assumption used in the characterization of production functions. Concavity of the production function implies diminishing marginal productivity of each input.[17] This assumption is widely

[17]Quasi-concavity does not imply diminishing marginal productivity to factor inputs. However, under constant returns to scale, quasi-concavity does guarantee diminishing marginal products. This is because quasi-concavity combined with constant returns to scale yields concavity. That being said, a major issue with constant returns to

agreed upon by economists and failure to impose may lead to conclusions which are economically infeasible.

In the case of a single factor, a twice continuously differentiable function $m(x)$ is said to be concave if $m''(x) \leq 0 \ \forall x \in \mathcal{S}(x)$. Extending this result to the case of multiple x's is relatively straight forward. Concavity implies that the Hessian matrix

$$H(m(x)) = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1k} \\ m_{21} & m_{22} & \cdots & m_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kk} \end{bmatrix}$$

must be negative semi-definite.[18] In other words, all the lth ($l = 1, 2, \ldots, k$) order principal minors of H are less than or equal to zero if l is odd and greater than or equal to zero if l is even. The notation m_{lk} is shorthand for $\partial^2 m(x)/\partial x_l \partial x_k$. One could, instead, chose to impose concavity via the constraints given in Hildreth (1954), however, many formal definitions of concavity are linked to the Hessian and as such we enforce concavity using this.

Following Hall and Huang (2001), we have the following constrained nonlinear programming problem:

$$\min D_\rho(p) \ s.t. \ H(m(x|p)) \text{ is negative semi-definite } \forall x \in \mathcal{S}(x), \ p_i \geq 0 \ \forall i, \ \text{and} \ \sum_{i=1}^n p_i = 1. \tag{19}$$

To solve this or any other constrained optimization problem in the spirit of Hall and Huang (2001) one needs to use sequential quadratic programming.

4.3. **SQP for imposing concavity.** In this sub-section we discuss how to impose concavity using SQP. If we use the power divergence measure of Cressie and Read (1984):

$$D_\rho(p) = \frac{1}{\rho(1-\rho)} \left\{ n - \sum_{i=1}^n (np_i)^\rho \right\},$$

for $-\infty < \rho < \infty$ and $\rho \neq 0$, 1, as our objective function to minimize, then we have the following set of functions that need to be estimated prior to solving our QP at any iteration (ℓ^{th}):

Our Objective Function: $D_\rho(p_\ell) = \frac{1}{\rho(1-\rho)} \left\{ n - \sum_{i=1}^n (np_{i,\ell})^\rho \right\}$.

First Partial of Objective Function: $\nabla D_\rho(p_\ell) = vec\left[\frac{-n}{1-\rho}(np_{i,\ell})^{\rho-1}\right]$.

Second Partial of Objective Function: $\nabla^2 D_\rho(p_\ell) = diag\left[n^2(np_{i,\ell})^{\rho-2}\right]$.

Our equality constrained function, $r(z)$**:** $\sum_{i=1}^n p_{i,\ell} - 1$.

First Partial of $r(z)$**:** $B_r(p_\ell) = [1, 1, \ldots, 1]$, an n-vector of ones.

Second Partial of $r(z)$**:** $\nabla B_r(p_\ell)$ which is an $n \times n$ matrix of zeros.

---

scale is that it implies that both the average and marginal productivities of inputs are independent of the scale of production. In other words, they depend only on the relative proportion of inputs.

[18]Alternatively we can examine the eigenvalues of the hessian matrix to make sure that they are all non-positive.

Additionally, we have to calculate our inequality constrained functions as well as their first and second partial derivatives, which can be broken into two pieces. First, we focus directly on the linear inequality constraints, $p_i > 0 \; \forall i$. For this we have

First Partial: $B_{c,1}(p_\ell) = [e_1, e_2, \ldots, e_n]$, where e_j is an $n-$vector of zeros with a 1 in the j^{th} spot.

Second Partial: $\nabla B_{c,1}(p_\ell)$ which is an $n \times n$ matrix of zeros.

Additionally, we have to calculate the first and second derivatives of the determinants of the principal minors of our Hessian matrix *for each* point we wish to impose concavity. In a local constant setting, the Hessian matrix is calculated as follows. Our first order partial derivatives of our local constant smoother are

$$\frac{\partial \hat{m}(x|p)}{\partial x_j} = \sum_{i=1}^{n} p_i y_i \frac{\partial A_i(x)}{\partial x_j} = \tilde{m}_j(x|p) - \hat{m}(x|p)\ddot{m}_j(x|p), \tag{20}$$

where $\tilde{m}_j(x|p)$ is the regression of $p_i y_i \left[(x_{ji} - x_j) / h_\ell^2 \right]$ on $x_1, \ldots, x_k$ and $\ddot{m}_j(x)$ is the regression of $p_i \left[(x_{ji} - x_j) / h_j^2 \right]$ on $x_1, \ldots, x_k$.

To determine the second order partial derivatives of our smooth regression function we need only to calculate the partial derivatives of the three functions in equation 20. Algebra reveals that

$$\begin{aligned}
\frac{\partial \hat{m}(x|p)}{\partial x_s} &= \hat{m}_s^*(x|p) - \hat{m}(x|p)\ddot{m}_s(x|p) \\
\frac{\partial \hat{m}_j^*(x|p)}{\partial x_s} &= \hat{m}_{j,s}^*(x|p) - \hat{m}_j^*(x|p)\ddot{m}_s(x|p) \\
\frac{\partial \ddot{m}_j(x|p)}{\partial x_s} &= \breve{m}_{j,s}(x|p) - \hat{m}_j(x|p)\ddot{m}_s(x|p) \\
\frac{\partial \hat{m}_j^*(x|p)}{\partial x_j} &= -\hat{m}(x|p)/h_j^2 + \breve{m}_j(x) - \hat{m}_j^*(x|p)\ddot{m}_j(x|p) \\
\frac{\partial \ddot{m}_j(x|p)}{\partial x_j} &= \dot{m}_j(x|p) + \breve{m}_j(x|p) - \ddot{m}_j(x|p)^2,
\end{aligned}$$

so that our second order and cross partials of the LCLS estimator are

$$\begin{aligned}
\frac{\partial \hat{m}^2(x|p)}{\partial x_j \partial x_s} =& \hat{m}_{j,s}^*(x|p) - \hat{m}_j^*(x|p)\ddot{m}_s(x|p) - \hat{m}(x|p)\breve{m}_{j,s}(x|p) \\
&+ 2\hat{m}(x|p)\ddot{m}_j(x|p)\ddot{m}_s(x|p) - \hat{m}_s^*(x|p)\ddot{m}_j x|p)
\end{aligned} \tag{21}$$

$$\begin{aligned}
\frac{\partial \hat{m}^2(x|p)}{\partial^2 x_j} =& -\hat{m}(x)/h_j^2 + \breve{m}_j(x|p) - 2\ddot{m}_j(x|p)\hat{m}_j^*(x|p) \\
&+ 2\hat{m}(x|p)\ddot{m}_j(x|p)^2 - \hat{m}(x|p)\breve{m}_j(x|p) - \hat{m}(x|p)\dot{m}_j(x|p).
\end{aligned} \tag{22}$$

Here $\breve{m}_{j,s}(x|p)$ is the regression of $\left[(x_{ji} - x_j) / h_j^2 \right] \left[(x_{si} - x_s) / h_s^2 \right] p_i$ on $x_1, \ldots, x_k$, $\breve{m}_j(x|p)$ is the regression of $\left[(x_{ji} - x_j) / h_j^2 \right] p_i y_i$ on $x_1, \ldots, x_k$ and $\dot{m}_j(x|p)$ is the regression of $\left[(x_{ji} - x_j) / h_j^2 \right]^2 p_i$ on $x_1, \ldots, x_k$.

It is immediately clear that calculating the first and second derivatives of the Hessian matrix using a closed form solution is difficult for more than two or three covariates. We suggest using numerical techniques in the user's preferred software to calculate the first and second derivatives of the Hessian matrix to then pass to the SQP. For k covariates, if one imposes concavity for each of the n points then this requires construction of n $k \times k$ Hessian matrices. There are k determinants of principal minors to be calculated for each Hessian representation, resulting in nk constraints to go with the $n+1$ constraints placed on the weights. This results in a total of $n(k+1)+1$ total constraints.

4.4. **Simulated Examples.** This sub-section uses Monte Carlo simulations to examine the finite sample performance of our constrained estimator. Following the focus on concavity, we choose to perform our simulations imposing concavity in models which should be concave. We consider the two following data generating processes:

$$y = \sqrt{x} + u \tag{23}$$

$$y = \ln(x) + u \tag{24}$$

where x is generated as a uniform distribution from zero to one and u is generated as normal with mean zero and variance equal to 0.1. Note that each of these data generating processes produce a theoretically consistent concave function. However, both the unknown error and finite sample biases may cause the kernel estimate to exhibit ranges of non-concavities.

In each Monte Carlo replication, we consider $N = 100$ observations. The number of replications is set equal to 100. Using this generated data we employ both the constrained and unconstrained estimation procedures for each replication. We use local-constant least-squares and a Gaussian kernel. The weights (p) are determined using $\rho = 0.5$ (Hellinger distance) and are found using the sequential quadratic programming routine SQPSolve in the programming language GAUSS 8.0. While our problem is not a quadratic programming problem, this type of solver uses a modified quadratic program to find the step length for moving in the direction of a minimum.

The simulation results are given in Figures 1 and 2, for cases 23 and 24 respectively. The curves correspond to the 95th percentile of the distance metric. Specifically, in Figure 1 $D_\rho(p) = 0.2412$ and $D_\rho(p) = 0.2465$ is Figure 2. The solid line in each figure is the unconstrained local-constant least-squares estimator and the dashed line is the constrained local-constant least-squares estimator. It is obvious that the unconstrained estimators show regions where the second derivative is positive. Our constrained estimator corrects for these non-concavities by changing the probability weights. We see that the dashed estimator is visually concave in each figure.

Notice that the dashed and solid curves do not differ much. The purpose of this estimator is not to drastically change the behavior of the estimator, but to correct for small sample biases which give results that may be economically inconsistent. This correction allows for 'proper' inference across the range of x.

5. Conclusion

This chapter has surveyed the existing literature on imposing constraints in nonparametric regression, described each method and discussed computational implementation. This survey included recent research that has not been discussed previously in the literature. We also described a novel method to impose general nonlinear constraints in nonparametric regression that can be implemented using only a standard quadratic programming solver. We illustrated this method with a small simulation study.

Overall future research should determine the relevant merits of each of the methods described here to whittle the set of potential methods down to one or two that can be easily and successfully used in applied nonparametric settings. Additionally, we feel that our description of the available methods should help to further research in extending these ideas to additional nonparametric settings, most notably in the estimation of quantile functions (Li and Racine 2008), conditional densities, treatment effects (Li, Racine and Wooldridge 2008), and structural estimators (Henderson et al. 2008).

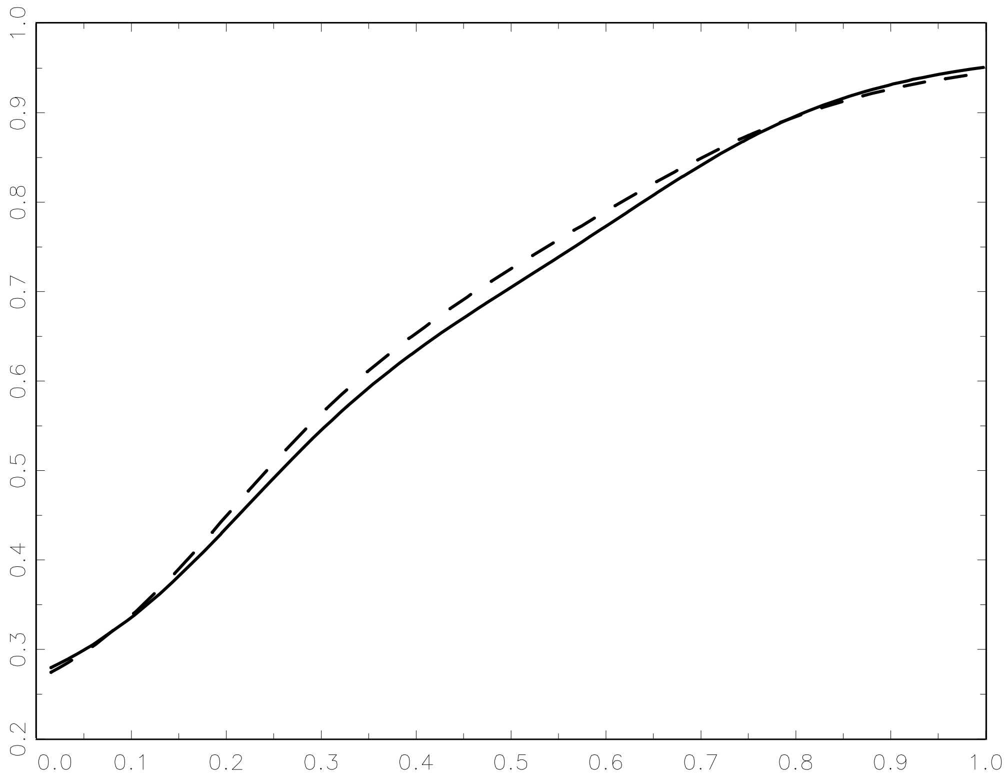

FIGURE 1. Unconstrained and constrained estimator of $y = \sqrt{x} + u$.

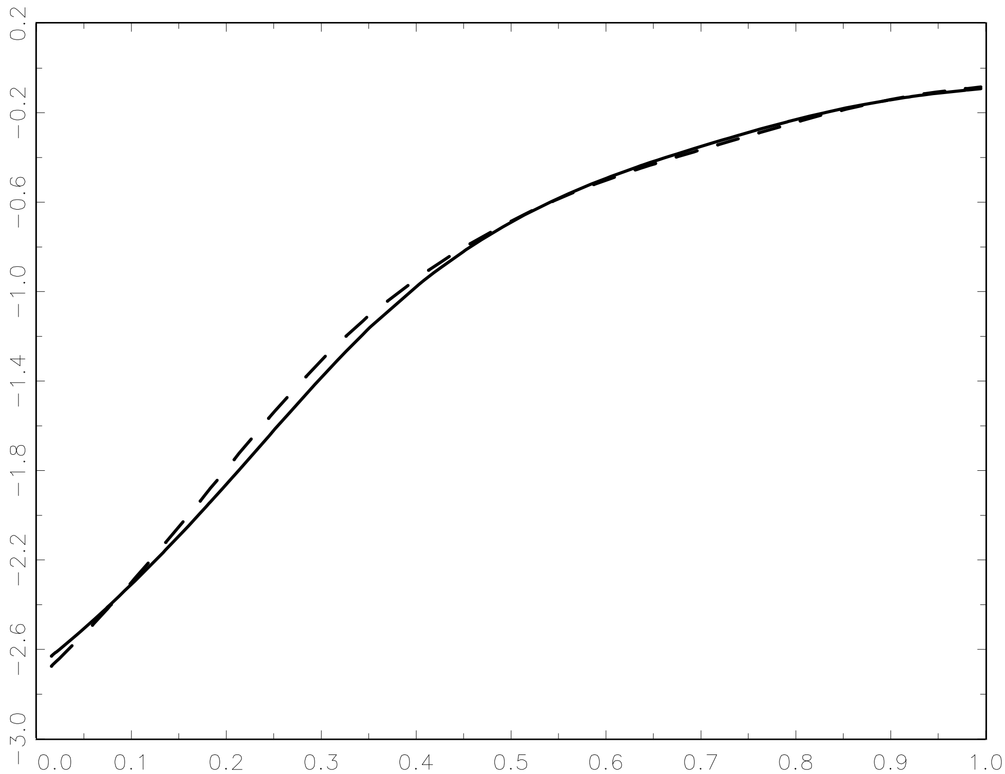

FIGURE 2. Unconstrained and constrained estimation of $y = \ln(x) + u$.

References

Allon, G., Beenstock, M., Hackman, S., Passy, U. & Shapiro, A. (2007), 'Nonparametric estimation of concave production technologies by entropic methods', *Journal of Applied Econometrics* **22**, 795–816.

Beresteanu, A. (2004), Nonparametric estimation of regression functions under restricitons on partial derivatives. *Mimeo*, Duke University.

Braun, W. J. & Hall, P. (2001), 'Data sharpening for nonparametric inference subject to constraints', *Journal of Computational and Graphical Statistics* **10**, 786–806.

Brunk, H. D. (1955), 'Maxmimum likelihood estimates of monotone parameters', *Annals of Mathematical Statistics* **26**, 607–616.

Chernozhukov, V., Fernandez-Val, I. & Galichon, A. (2007), Improving estimates of monotone functions by rearrangement. *Mimeo.*

Choi, E. & Hall, P. (1999), 'Data sharpening as a prelude to density estimation', *Biometrika* **86**, 941–947.

Cressie, N. A. C. & Read, T. R. C. (1984), 'Multinomial goodness-of-fit tests', *Journal of the Royal Statistical Society, Series B* **46**, 440–464.

Dette, H., Neumeyer, N. & Pilz, K. F. (2006), 'A simple nonparametric estimator of a strictly monotone regression function', *Bernoulli* **12**(3), 469–490.

Dette, H. & Pilz, K. F. (2006), 'A comparative study of monotone nonparametric kernel estimates', *Journal of Statistical Computation and Simulation* **76**(1), 41–56.

Dierckx, P. (1980), 'An algorithm for cubic spline fitting with convexity constraints', *Computing* **24**, 349–371.

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

Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, Dekker, New York.

Friedman, J., Tukey, J. W. & Tukey, P. (1980), Approaches to analysis of data that concentrate near intermediate-dimensional manifolds, *in* E. Diday et al., ed., 'Data Analysis and Informatics', North-Holland: Amsterdam.

Gallant, A. R. (1981), 'On the bias in flexible functional forms and an essential unbiased form: The fourier flexible form', *Journal of Econometrics* **15**, 211–245.

Gallant, A. R. (1982), 'Unbiased determination of production technologies', *Journal of Econometrics* **20**, 285–323.

Gallant, A. R. & Golub, G. H. (1984), 'Imposing curvature restrictions on flexible functional forms', *Journal of Econometrics* **26**, 295–321.

Goldman, S. & Ruud, P. (1992), Nonparametric multivariate regression subject to constraint, Technical report, University of California, Berkeley, Department of Economics.

Hall, P. & Huang, H. (2001), 'Nonparametric kernel regression subject to monotonicity constraints', *The Annals of Statistics* **29**(3), 624–647.

Hall, P. & Presnell, B. (1999), 'Intentionally biased bootstrap methods', *Journal of the Royal Statistical Society Series B* **61**, 143–158.

Hansen, D. L., Pledger, G. & Wright, F. T. (1973), 'On consistency in monotonic regression', *Annals of Statistics* **1**(3), 401–421.

Henderson, D. J., List, J. L., Millimet, D. L., Parmeter, C. F. & Price, M. K. (2008), Imposing monotonicity nonparametrically in first price auctions. Virginia Tech AAEC working paper.

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

Holm, S. & Frisen, M. (1985), Nonparametric regression with simple curve characteristics, Technical Report 4, Department of Statistics, University of Goteborg, Goteborg, Sweden.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Li, Q. & Racine, J. S. (2008), 'Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data', *Journal of Business & Economic Statistics* **26**(4), 423–434.

Li, Q., Racine, J. S. & Wooldridge, J. M. (2008), 'Estimating average treatment effects with continuous and discrete covariates: The case of swan-ganz catherization', *American Economic Review* **98**(2), 357–362.

Mammen, E. (1991*a*), 'Estimating a smooth monotone regression function', *Annals of Statistics* **19**(2), 724–740.

Mammen, E. (1991*b*), 'Nonparametric regression under qualitative smoothness assumptions', *Annals of Statistics* **19**(2), 741–759.

Matzkin, R. L. (1991), 'Semiparametric estimation of monotone and concave utility functions for polychotomous choice models', *Econometrica* **59**, 1315–1327.

Matzkin, R. L. (1992), 'Nonparametric and distribution-free estimation of the binary choice and the threshold-crossing models', *Econometrica* **60**, 239–270.

Matzkin, R. L. (1993), 'Nonparametric identification and estimation of polychotomous choice models', *Journal of Econometrics* **58**, 137–168.

Matzkin, R. L. (1994), Restrictions of economic thoery in nonparametric methods, *in* D. L. McFadden & R. F. Engle, eds, 'Handbook of Econometrics', Vol. 4, North-Holland: Amsterdam.

Matzkin, R. L. (1999), Computation of nonparametric concavity restricted estimators. *Mimeo.*

Mukerjee, H. (1988), 'Monotone nonparametric regression', *Annals of Statistics* **16**, 741–750.

Nocedal, J. & Wright, S. J. (2000), *Numerical Optimization*, 2nd edn, Springer.

Racine, J. S. & Parmeter, C. F. (2008), Constrained nonparametric kernel regression: Estimation and inference. Working Paper.

Ramsay, J. O. (1988), 'Monotone regression splines in action (with comments)', *Statistical Science* **3**, 425–461.

Robinson, S. M. (1974), 'Perturbed kuhn-tucker points and rates of convergence for a class of nonlinear-programming algorithms', *Mathematical Programming* **7**, 1–16.

Ruud, P. A. (1995), Restricted least squares subject to monotonicity and concavity restraints. Presented at the 7^{th} World Congress of the Econometric Society.

Schumaker, L. (1981), *Spline Functions: Basic T*, Wiley, New York.

Wahba, G. (1990), *Spline models for observational data*, CBMS-NSF Conference Series in Applied Mathematics 59, SIAM, Philadelphia, PA.

Yatchew, A. & Bos, L. (1997), 'Nonparametric regression and testing in economic models', *Journal of Quantitative Economics* **13**, 81–131.

Yatchew, A. & Härdle, W. (2006), 'Nonparametric state price density estimation using constrained least squares and the bootstrap', *Journal of Econometrics* **133**, 579–599.