CONSTRAINED NONPARAMETRIC KERNEL REGRESSION: ESTIMATION AND INFERENCE

JEFFREY S. RACINE AND CHRISTOPHER F. PARMETER

Abstract. Restricted kernel regression methods have recently received much well-deserved attention. Powerful methods have been proposed for imposing monotonicity on the resulting estimate, a condition often dictated by theoretical concerns; see Hall, Huang, Gifford & Gijbels (2001) and Hall & Huang (2001), among others. However, to the best of our knowledge, there does not exist a simple yet general approach for constraining a nonparametric regression that allows practitioners to impose any manner and mix of constraints on the resulting estimate. In this paper we generalize Hall & Huang's (2001) approach in order to allow for equality or inequality constraints on a nonparametric regression model and its derivatives of any order. The proposed approach is straightforward, both conceptually and in practice. A testing framework is provided allowing researchers to thereby impose and test the validity of the restrictions. Illustrative Monte Carlo results are presented, and an application is considered.

JEL Classification: C12 (Hypothesis testing), C13 (Estimation), C14 (Semiparametric and nonparametric methods)

1. Introduction and Overview

Kernel regression methods can be found in a range of application domains, and continue to grow in popularity. Their appeal stems from the fact that they are robust to functional misspecification that can otherwise undermine conventional parametric regression methods. However, one frequently levied complaint towards kernel regression methods is that, unlike their parametric counterparts, there does not exist a simple yet general method for imposing arbitrary constraints. One consequence of this is that when people wish to impose arbitrary constraints on a nonparametric estimate they must often leave the kernel framework and migrate towards, say, a series framework in which it is relatively straightforward to impose such constraints, or they resort to non-smooth convex programming methods.

Date: August 26, 2008.
Key words and phrases. Restrictions, Equality, Inequality, Smooth, Testing.
We would like to thank but not implicate Daniel Wikström for inspiring conversations and Li-Shan Huang and Peter Hall for their insightful comments and suggestions. All errors remain, of course, our own. Racine would like to gratefully acknowledge support from Natural Sciences and Engineering Research Council of Canada (NSERC:www.nserc.ca), the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca).

One particular constraint that has received much attention in kernel regression settings is that of monotonicity. In the statistics literature, the development of monotonic estimators dates back to the likelihood framework of Brunk (1955). This technique later came to be known as 'isotonic regression' and, while nonparametric in nature (min/max), produced curves that were not smooth. Notable contributions to the development of this method include Hanson, Pledger & Wright (1973) who demonstrated consistency in two dimensions (Brunk (1955) focused solely on the univariate setting) and Dykstra (1983), Goldman & Ruud (1992) and Ruud (1995) who developed efficient computational algorithms for the general class of restricted estimators to which isotonic regression belongs.[1] Mukerjee (1988) and Mammen (1991) developed methods for kernel-based isotonic regression and both techniques consist of a smoothing step using kernels (as opposed to interpolation) and an isonotonization step which imposes monotonicity.[2] A more recent alternative to these kernel-based isotonic methods employs constrained smoothing splines. The literature on constrained smoothing splines is vast and includes the work of Ramsay (1988), Kelly & Rice (1990), Li, Naik & Swetits (1996), Turlach (1997) and Mammen & Thomas-Agnam (1999), to name but a few.

Recent work on imposing monotonicity on a nonparametric regression function includes Pelckmans, Espinoza, Brabanter, Suykens & Moor (2005), Dette, Neumeyer & Pilz (2006) and Chernozhukov, Fernandez-Val & Galichon (2007). Each of these approaches is nonparametric in nature with the last two being kernel-based. Dette et al. (2006) and Chernozhukov et al. (2007) use a method known as 'rearrangement' which produces a monotonically constrained estimator derived from the probability integral transformation lemma. Essentially one calculates the CDF of the density of regression estimates to construct an estimate of the inverse of the monotonic function which is inverted to provide the final estimate. Pelckmans et al. (2005) construct a monotone function based on least squares using the Chebychev norm with a Tikhonov regularization scheme. This method involves solving a standard quadratic program and is comparable to the spline-based methods mentioned above. Braun & Hall (2001) propose a method closely related to rearrangement which they call 'data sharpening' that also involves rearranging the positions of data values,

[1]An excellent overview of isotonic regression can be found in Robertson, Wright & Dykstra (1988).
[2]The order of these two steps are irrelevant asymptotically.

controlled by minimizing a measure of the total distance that the data are moved, subject to a constraint. Braun & Hall (2001) consider applying the method to render a density estimator unimodal and to monotonize a nonparametric regression; see also Hall & Kang (2005).

One of the most promising (and extensible) approaches for imposing monotonicity on a nonparametric regression model is that of Hall & Huang (2001) who proposed a novel approach towards imposing monotonicity constraints on a quite general class of kernel smoothers. Their monotonically constrained estimator is constructed by introducing probability weights for each response data point which can dampen or magnify the impact of any observation thereby imposing monotonicity.[3] The weights are global with respect to the sample and are chosen by minimizing a preselected version of the power divergence metric of Cressie & Read (1984). The introduction of the weights in effect transforms the response variable in order to achieve monotonicity of the underlying regression function. Though this method delivers a smooth monotonically constrained nonparametric kernel estimator, unfortunately, probability weights and power divergence metrics are of limited utility when imposing arbitrary constraints of the type we consider herein. But a straightforward generalization of Hall & Huang's (2001) method will allow one to impose arbitrary constraints, which we outline in the proceeding section.

Imposing arbitrary constraints on nonparametric surfaces, while not new to econometrics, has not received anywhere near as much attention as has imposing monotonicity, at least not in the kernel regression framework. Indeed, the existing literature dealing with constraints in a nonparametric framework appears to fall into three broad categories:

(i) Those that develop nonparametric estimators which satisfy a particular constraint (e.g., monotonically constrained estimators).
(ii) Those that develop nonparametric estimators which can satisfy arbitrary constraints (e.g., constrained smoothing splines).
(iii) Those that develop tests of the validity of constraints (e.g., concavity).

Tests developed in (iii) can be further subdivided into statistical and nonstatistical tests. The nonstatistical tests 'check' for violations of economic theory, such as indifference curves crossing or isoquants having the wrong slope; see Hanoch & Rothschild (1972) and Varian (1985). The

[3]See Dette & Pilz (2006) for a Monte Carlo comparison of smooth isotonic regression, rearrangement, and the method of Hall & Huang (2001).

statistical tests develop a metric to determine if the constraints are satisfied and develop the asymptotic properties of the proposed metric. These metrics are constructed from measures of fit for the unrestricted and restricted models and do not focus on pure 'economic' violations.

Early nonparametric methods designed to impose general economic constraints include Gallant (1981), Gallant (1982), and Gallant & Golub (1984) who introduced the Fourier Flexible Form estimator (FFF) which is a series-based estimator whose coefficients can be easily restricted thereby imposing concavity, homotheticity and homogeneity in a nonparametric setting.[4]

The seminal work of Matzkin (1991), Matzkin (1992), Matzkin (1993) and Matzkin (1994) considers identification and estimation of general nonparametric problems with arbitrary economic constraints and is perhaps most closely related to the methods proposed herein. One of Matzkin's key insights is that when nonparametric identification was not possible, imposing shape constraints tied to economic theory may provide nonparametric identification. This work lays the foundation for a general operating theory of constrained nonparametric estimation. Her methods focus on standard economic constraints (monotonicity, concavity, homogeneity, etc.) but can be generalized to allow for arbitrary constraints on the function of interest. While the methods are completely general, she focuses mainly on the development of economically constrained estimators for the binary and polychotomous choice models.

Implementation of Matzkin's constrained methods is of the two-step variety; see Matzkin (1999) for details. First, for the specified constraints, a feasible solution consisting of a finite number of points is determined through optimization of some criterion function (in the choice framework this is a pseudo-likelihood function). Second, the feasible points are interpolated or smoothed to construct the nonparametric surface that satisfies the constraints. The nonparametric least squares approach of Ruud (1995) is similar in spirit to the work of Matzkin, but focuses primarily on monotonicity and concavity.[5]

[4]We note that monotonicity is not easily imposed in this setting.

[5]While Matzkin's methods are novel and have contributed greatly to issues related to econometric identification, their use for constrained estimation in applied settings appears to be scarce and is likely due to the perceived complexity of the proposed approach. For instance, statements such as those found in Chen & Randall (1997, p. 324) who note that "However, for those who desire the properties of a the distribution-free model, the empirical implementation can be difficult. [...] To estimate the model using Matzkin's method, a large constrained optimization needs to be solved." underscore the perceived complexity of Matzkin's approach. It should be noted that Matzkin has employed her methodology in an applied setting (see Briesch, Chintagunta & Matzkin (2002)) and her web page presents a detailed outline of both the methods and a working procedure for their use in economic applications (Matzkin (1999)).

Yatchew & Bos (1997) develop a series-based estimator that can handle general constraints. This estimator is constructed by minimizing the sum of squared errors of a nonparametric function relative to an appropriate Sobolev norm. The basis functions that make up the series estimator are determined from a set of differential equations that provide 'representors'. Yatchew & Bos (1997) begin by describing general nonparametric estimation and then show how to constrain the function space in order to satisfy given constraints. They also develop a conditional moment test to study the statistical validity of the constraints. Given that Matzkin's early work did not focus on developing tests of economic constraints, Yatchew & Bos (1997) represents one of the first studies to simultaneously consider estimation and testing of economic constraints in a nonparametric (series) setting.

Contemporary work involving the estimation of smooth, nonparametric regression surfaces subject to derivative constraints includes Beresteanu (2004) and Yatchew & Härdle (2006). Beresteanu (2004) introduced a spline-based procedure that can handle multivariate data while imposing 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. Beresteanu (2004) also suggests testing the constraints using an L_2 distance measure between the unrestricted and restricted function estimates. Thus, his work presents a general framework for constraining and testing a nonparametric regression function in a series framework, similar to the earlier work of Yatchew & Bos (1997). He employed his method to impose monotonicity and supermodularity of a cost function for the telephone industry.

The work of Yatchew & Härdle (2006) focuses on nonparametric estimation of an option pricing model where the unknown function must satisfy monotonicity and convexity along with the density of state prices being a true density.[6] Their approach uses the techniques developed by Yatchew & Bos (1997). They too develop a test of their restrictions, but, unlike Beresteanu (2004), their test uses the residuals from the constrained estimate to determine if the covariates 'explain' anything else, and if they do the constraints are rejected.

Contemporary work involving the estimation of nonsmooth, constrained nonparametric regression surfaces includes Allon, Beenstock, Hackman, Passy & Shapiro (2007) who focused on imposing

[6]This paper is closely related to our idea of imposing general derivative constraints as their approach focuses on the first three derivatives of the regression function.

economic constraints for cost and production functions. Allon et al. (2007) show how to construct an estimator consistent with the nonparametric, nonstatistical testing device developed by Hanoch & Rothschild (1972). Their estimator employs a convex programming framework that can handle general constraints, albeit in a non-smooth setting. A nonstatistical testing device similar to Varian (1985) is discussed as well.

Notwithstanding these recent developments, there does not yet exist a methodology grounded in kernel methods that can impose general constraints and statistically test the validity of these constraints. We bridge this gap by providing a method for imposing general constraints in nonparametric kernel settings delivering a smooth constrained nonparametric estimator and we provide a simple bootstrapping procedure to test the validity of the constraints of interest. Our approach is achieved by modifying and extending the approach of Hall & Huang (2001) resulting in a simple and general multivariate, multi-constraint procedure. As noted by Hall & Huang (2001, p. 625), the use of splines does not hold the same attraction for users of kernel methods, and the fact that Hall & Huang's (2001) method is rooted in a conventional kernel framework naturally appeals to the community of kernel-based researchers. Furthermore, recent developments that permit the kernel smoothing of categorical and continuous covariates can dominate spline methods; see Li & Racine (2007) for some examples. Nonsmooth methods,[7] either the fully nonsmooth methods of Allon et al. (2007) or the interpolated methods of Matzkin (1991) and Matzkin (1992), may fail to appeal to kernel users for the same reasons. As such, to the best of our knowledge, there does not yet exist a simple and easily implementable procedure for imposing and testing the validity of arbitrary constraints on a regression function estimated using kernel methods that is capable of producing smooth constrained estimates.

The rest of this paper proceeds as follows. Section 2 outlines the basic approach. Section 2.1 addresses existence and uniqueness of the solution. Section 2.2 presents a simple test of the validity of the constraints. Section 3 considers a number of simulated applications and examines the finite-sample performance of the proposed test. Section 4 presents an empirical application involving technical efficiency on Indonesian rice farms, and Section 5 presents some concluding remarks. Appendix A presents details on the implementation for the specific case of monotonicity

[7]When we use the term nonsmooth we are referring to methods that either do not smooth the nonparametric function *or* smooth the constrained function *after* the constraints have been imposed.

and concavity which may be of interest to some readers, while Appendix B presents R code (R Development Core Team (2008)) to replicate the simulated illustration presented in Section 3.1.

2. Methodology

In what follows we let $\{X_i, Y_i\}_{i=1}^n$ denote sample pairs of explanatory and response variables, and our goal is to estimate the unknown average response $g(x) = E(Y|X = x)$ subject to constraints on $g^{(\mathbf{s})}(x)$ where $\mathbf{s}$ is a k-vector corresponding to the dimension of x. In what follows, the elements of $\mathbf{s}$ represent the order of the partial derivative corresponding to each element of x. Thus $\mathbf{s} = (0, 0, \ldots, 0)$ represents the function itself, while $\mathbf{s} = (1, 0, \ldots, 0)$ represents $\partial g(x)/\partial x_1$. In general, for $\mathbf{s} = (s_1, s_2, \ldots, s_k)$ we have

$$g^{(\mathbf{s})}(x) = \frac{\partial^{s_1} g(x)}{\partial x_1^{s_1}} \cdots \frac{\partial^{s_k} g(x)}{\partial x_k^{s_k}}. \tag{1}$$

We consider the class of kernel regression smoothers that can be written as linear combinations of the response Y_i, i.e.,

$$\hat{g}(x) = \sum_{i=1}^{n} A_i(x) Y_i. \tag{2}$$

This class includes the Nadaraya-Watson estimator (Nadaraya (1965),Watson (1964)), the Priestley-Chao estimator (Priestley & Chao (1972)), and the local polynomial estimator (Fan (1992)), among others.

We presume that the reader may wish to impose constraints on the estimate $\hat{g}(x)$ of the form $l(x) \le g^{(\mathbf{s})}(x) \le u(x)$ for arbitrary $u(\cdot)$, $l(\cdot)$, and $\mathbf{s}$. For some applications, $\mathbf{s} = (0, \ldots, 0, 1, 0, \ldots, 0)$ would be of particular interest, say for example when the partial derivative represents a budget share and therefore must lie in $[0, 1]$. Or, $\mathbf{s} = (0, 0, \ldots, 0)$ might be of interest when an outcome must be bounded (i.e., $g(x)$ could be a probability, and, hence must lie in $[0, 1]$, but this could be violated when using, say, a local linear smoother). Or, $l(\cdot) = u(\cdot)$ might be required (i.e., equality rather than inequality constraints) such as when imposing adding up constraints, say, when the sum of the budget shares must equal one, or when imposing homogeneity of a particular degree, by way of example. The approach we describe is quite general. It is firmly embedded in a conventional multivariate kernel framework, and admits arbitrary combinations of constraints (i.e., for any s or combination thereof) subject to the obvious caveat that the constraints must be internally consistent.

Following Hall & Huang (2001), we consider a generalization of $\hat{g}(x)$ defined in (2) given by

$$\hat{g}(x|p) = \sum_{i=1}^{n} p_i A_i(x) Y_i, \tag{3}$$

and for what follows $\hat{g}^{(\mathbf{s})}(x|p) = \sum_{i=1}^{n} p_i A_i^{(\mathbf{s})}(x) Y_i$ where $A_i^{(\mathbf{s})}(x) = \frac{\partial^{s_1} A_i(x)}{\partial x_1^{s_1}} \cdots \frac{\partial^{s_k} A_i(x)}{\partial x_k^{s_k}}$ for continuous x. Again, in our notation $\mathbf{s}$ represents a $k \times 1$ vector of nonnegative integers that indicate the order of the partial derivative of the weighting function of the kernel smoother.

By way of example, using (3) to generate an unrestricted Nadaraya-Watson estimator we would set $p_i = 1/n$, $i = 1, \ldots, n$, and set

$$A_i(x) = \frac{n K_\gamma(X_i, x)}{\sum_{j=1}^{n} K_\gamma(X_j, x)}, \tag{4}$$

where $K_\gamma(\cdot)$ is a generalized product kernel that admits both continuous and categorical data, and γ is a vector of bandwidths; see Racine & Li (2004) for details. When $p_i \neq 1/n$ for some i, then we would have a constrained Nadaraya-Watson estimator. Note that one uses the same bandwidths for the constrained and unconstrained estimator hence bandwidth selection proceeds using standard methods, i.e., cross-validation on the sample data. Note also that the unconstrained and constrained estimators are identical for those observations for which $p_i = 1/n$.

Let p_u be an n-vector with elements $1/n$ and let p be the vector of weights to be selected. In order to impose our constraints, we choose $p = \hat{p}$ to minimize the distance from p to the uniform weights $p_i = 1/n \; \forall i$ as proposed by Hall & Huang (2001). This is appealing intuitively since the unconstrained estimator is that for which $p_i = 1/n \; \forall i$, as noted above. Whereas Hall & Huang (2001) consider probability weights (i.e., $0 \leq p_i \leq 1$, $\sum_i p_i = 1$) and distance measures suitable for probability weights (i.e., Hellinger), we shall need to relax the constraint that $0 \leq p_i \leq 1$ and will instead allow for both positive and negative weights (while retaining $\sum_i p_i = 1$), and shall also therefore require alternative distance measures. To appreciate why this is necessary, suppose one simply wished to constrain a surface that is uniformly positive to have negative regions. This could be accomplished by allowing some of the weights to be negative, however probability weights would fail to produce a feasible solution (they are non-negative), hence our need to relax this condition.

We shall also have to forgo the power divergence metric of Cressie & Read (1984) which was used by Hall & Huang (2001) since it is only valid for probability weights. For what follows we select the well-worn L_2 metric $D(p) = (p_u - p)'(p_u - p)$ which has a number of appealing features in this context, as will be seen. Our problem therefore boils down to selecting those weights p that minimize $D(p)$ subject to $l(x) \le g^{(\mathbf{s})}(x) \le u(x)$ (and perhaps additional constraints of a similar form), which can be cast as a general nonlinear programming problem. For the illustrative constraints we consider below we have (in)equalities that are linear in p,[8] which can be solved using standard quadratic programming methods and off-the-shelf software. For example, in the R language (R Development Core Team (2008)) it is solved using the quadprog package, in GAUSS it is solved using the qprog command, and in MATLAB the quadprog command. Even when n is quite large the solution is computationally fast using any of these packages. Code in the R language (R Development Core Team (2008)) is available from the authors upon request; see Appendix B for an example. For (in)equalities that are nonlinear in p we can convert the nonlinear programming problem into a standard quadratic programming problem that can again be solved using off-the-shelf software albeit with modification.

2.1. **Existence and Uniqueness of a Solution.** Hall & Huang (2001) demonstrate that a vector of weights always exists that satisfy monotonicity constraints when the regressand is assumed to be positive for all observations. This assumption is too restrictive for the approach at hand. In what follows we shall focus on linear (in p) restrictions which are quite general.[9] See Appendix A for an implementation with constraints that are nonlinear in p in addition to constraints that are linear in p.

Our restrictions have the form:

$$\sum_{i=1}^{n} p_i \left[\sum_{\mathbf{s} \in \mathbf{S}} \alpha_{\mathbf{s}} A_i^{(\mathbf{s})}(x) \right] Y_i - c(x) \ge 0, \tag{5}$$

[8]Common economic constraints that satisfy (in)equalities that are linear in p include monotonicity, supermodularity, additive separability, homogeneity of degree k, diminishing marginal returns/products, general bounding of any order derivative, necessary conditions for concavity, etc.

[9]See Appendix A for an example of how to implement our method with constraints that are nonlinear in p and Henderson & Parmeter (2008) for a more general discussion of imposing arbitrary nonlinear constraints on a nonparametric regression surface, albeit with probability weights and the power divergence metric of Cressie & Read (1984).

where the inner sum is taken over all vectors $\mathbf{S}$ that correspond to our constraints and $\alpha_{\mathbf{s}}$ is a set of constants used to generate various constraints. In what follows we shall presume, without loss of generality, that for all $\mathbf{s}$, $\alpha_{\mathbf{s}} \geq 0$.

In order to economize on notation, we define $\psi_i(x) = \left[\sum_{\mathbf{s}\in\mathbf{S}} \alpha_{\mathbf{s}} A_i^{(\mathbf{s})}(x)\right] Y_i$. If for some sequence j_n in $\{1,\dots,n\}$, $\operatorname{sgn}\psi_{j_n}(x) = 1\ \forall x \in [\mathbf{a},\mathbf{b}]$ and for another sequence l_n in $\{1,\dots,n\}$, $\operatorname{sgn}\psi_{l_n}(x) = -1$ $\forall x \in [\mathbf{a},\mathbf{b}]$, then for those observations that switch signs, p_i may be set equal to zero, while $p_{j_n} > 0$ and $p_{l_n} < 0$ are sufficient to ensure existence of a solution of ps satisfying the constraints.

When no such sequences exist, existence of a weight vector will require further assumptions. For example, if one was willing to assume that i) a sequence $\{i_1,\dots,i_k\} \subseteq \{1,\dots,n\}$ exists such that for each k, $\psi_{i_k}(x)$ is strictly positive and continuous on $(\mathbf{L}_{i_k},\mathbf{U}_{i_k})$, ii) every $x \in [\mathbf{a},\mathbf{b}]$ is contained in at least one interval $(\mathbf{L}_{i_k},\mathbf{U}_{i_k})$, iii) for $1 \leq i \leq n$, $\psi_{i_k}(x)$ is continuous on $[-\infty,\infty]$, then there exists a vector $p = (p_1,\dots,p_n)$ such that the constraints are satisfied for all $x \in [\mathbf{a},\mathbf{b}]$. This result is a trivial extension of the induction argument given in Hall & Huang (2001, Theorem 4.1) which we therefore will not reproduce here.

Moreover, since the forcing matrix (I_n) in the quadratic portion of our L_2 norm, $p'I_np$, is positive semidefinite, if our solution p^* satisfies the set of linear equality/inequality constraints then p^* is the unique, global solution to the problem (Nocedal & Wright (2000, Theorem 16.4)). Positive semi-definiteness guarantees that our objective function is convex which is what yields a global solution.[10]

2.2. **Testing Constraint Validity.** As noted above, there exists a literature on testing restrictions in nonparametric settings including Abrevaya & Jiang (2005), who test for curvature restrictions and survey the literature, Epstein & Yatchew (1985), who develop a nonparametric test of the utility maximization hypothesis and homotheticity, Yatchew & Bos (1997), who develop a conditional moment test for a broad range of smoothness constraints, Ghosal, Sen & van der Vaart (2000), who develop a test for monotonicity, Beresteanu (2004), who as mentioned above discusses using a conditional mean type test for general constraints, and Yatchew & Härdle (2006), who employ a residual-based test to check for monotonicity and convexity. The tests of Yatchew & Bos (1997)

[10]When the forcing matrix is not convex, multiple solutions may exist and these types of problems are referred to as 'indefinite quadratic programs'.

and Beresteanu (2004) are the closest in spirit to the method we adopt below, having the ability to test general smoothness constraints. One could easily use the same test statistic as Yatchew & Bos (1997) and Beresteanu (2004) but replace the series estimator with a kernel estimator if desired. Aside from the test of Yatchew & Bos (1997), most existing tests check for specific constraints. This is limiting in the current setting as our main focus is on a smooth, arbitrarily restricted estimator.

We adopt a testing approach similar to that proposed by Hall et al. (2001) which is predicated on the objective function $D(\hat{p})$. This approach involves estimating the constrained regression function $\hat{g}(x|p)$ based on the sample realizations $\{Y_i, X_i\}$ and then rejecting H_0 if the observed value of $D(\hat{p})$ is too large. We use a resampling approach for generating the null distribution of $D(\hat{p})$ which involves generating resamples for y drawn from the constrained model via *iid* residual resampling (i.e., conditional on the sample $\{X_i\}$), which we denote $\{Y_i^*, X_i\}$. These resamples are generated under H_0, hence we recompute $\hat{g}(x|p)$ for the bootstrap sample $\{Y_i^*, X_i\}$ which we denote $\hat{g}(x|p^*)$ which then yields $D(p^*)$. We then repeat this process B times. Finally, we compute the empirical P value, P_B, which is simply the proportion of the B bootstrap resamples $D(p^*)$ that exceed $D(\hat{p})$, i.e.,

$$P_B = 1 - \hat{F}(D(\hat{p})) = \frac{1}{B}\sum_{j=1}^{B} I(D(p^*) > D(\hat{p})),$$

where $I(\cdot)$ is the indicator function and $\hat{F}(D(\hat{p}))$ is the empirical distribution function (EDF) of the bootstrap statistics. Then one rejects the null hypothesis if P_B is less than α, the level of the test. For an alternative approach involving kernel smoothing of $F(\cdot)$, see Racine & MacKinnon (2007*a*).

Before proceeding further, we note that there exist three situations that can occur in practice:

(i) Impose non-binding constraints (they are 'correct' de facto)
(ii) Impose binding constraints that are correct
(iii) Impose binding constraints that are incorrect

We shall only consider (ii) and (iii) in the Monte Carlo simulations in Section 3 below since, as noted by Hall et al. (2001, p 609), "For those datasets with $D(\hat{p}) = 0$, no further bootstrapping is necessary [. . .] and so the conclusion (for that dataset) must be to not reject H_0." The implication in the current paper is simply that imposing non-binding constraints does not alter the estimator and the unconstrained weights will be $\hat{p}_i = 1/n\ \forall i$ hence $D(\hat{p}) = 0$ and the statistic is degenerate.

Of course, in practice this simply means that we presume people are imposing constraints that bind, which is a reasonable presumption. In order to demonstrate the flexibility of the constrained estimator, in Section 3 below we consider testing for two types of restrictions. In the first case we impose the restriction that the regression function $g(x)$ is equal to a known parametric form $g(x,\beta)$, while in the second case we test whether the first partial derivative is constant and equal to the value one for all x (testing whether the first partial equals zero would of course be a test of significance).

We now demonstrate the flexibility and simplicity of the approach by first imposing a range of constraints on a simulated dataset using a large number of observations thereby showcasing the feasibility of this approach in substantive applied settings, and then consider some Monte Carlo experiments that examine the finite-sample performance of the proposed test.

3. Simulated Illustrations and Finite-Sample Properties of the Proposed Test

For what follows we shall simulate data from a nonlinear multivariate relationship and then consider imposing a range of restrictions by way of example. We consider a 3D surface defined by

$$Y_i = \frac{\sin\left(\sqrt{X_{i1}^2 + X_{i2}^2}\right)}{\sqrt{X_{i1}^2 + X_{i2}^2}} + \epsilon_i, \qquad i = 1, \dots, n, \tag{6}$$

where x_1 and x_2 are independent draws from the uniform [-5,5]. We draw $n = 10,000$ observations from this data generating process (DGP) with $\epsilon \sim N(0, \sigma^2)$ and $\sigma = 0.1$. As we will demonstrate the method by imposing restrictions on the surface and also on its first and second partial derivatives, we shall use the local quadratic estimator for what follows as it delivers consistent estimates of the regression function and its first and second partial derivatives. Figure 1 presents the unrestricted regression estimate whose bandwidths were chosen via least squares cross-validation.[11]

3.1. A Simulated Illustration: Restricting $\hat{g}^{(0)}(x)$.

Next, we arbitrarily impose the constraint that the regression function lies in the range [0,0.5]. A plot of the restricted surface appears in Figure 2.

[11]In all of the restricted illustrations to follow we use the same cross-validated bandwidths.

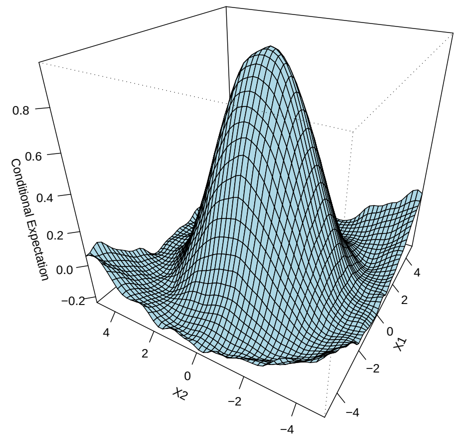

FIGURE 1. Unrestricted nonparametric estimate of (6), $n = 10,000$.

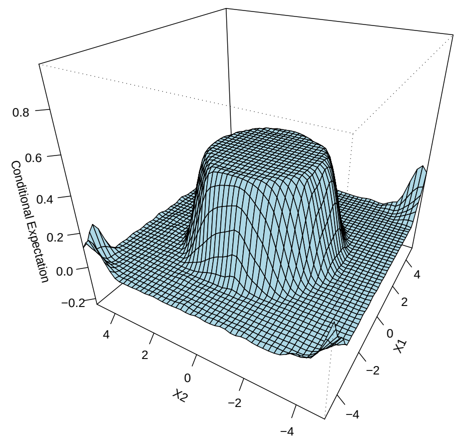

FIGURE 2. Restricted nonparametric estimate of (6) where the restriction is defined over $\hat{g}^{(\mathbf{s})}(x|p)$, $\mathbf{s} = (0,0)$, $(0 \le \hat{g}(x|p) \le 0.5)$, $n = 10,000$.

Figures 1 and 2 clearly reveal that the regression surface for the restricted model is both smooth and satisfies the constraints.

3.2. **A Simulated Illustration: Restricting** $\hat{g}^{(1)}(x)$**.** We consider the same DGP given above, but now we arbitrarily impose the constraint that the first derivatives with respect to both x_1 and x_2 lie in the range [-0.1,0.1].[12] A plot of the restricted surface appears in Figure 3.

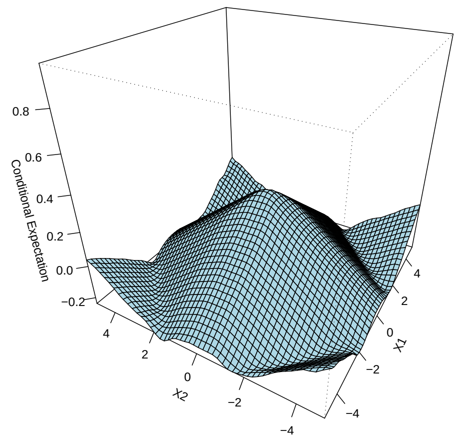

FIGURE 3. Restricted nonparametric estimate of (6) where the restriction is defined over $g^{(\mathbf{s})}(x)$, $\mathbf{s} \in \{(1,0),(0,1)\}$, $(-0.1 \leq \partial\hat{g}(x|p)/\partial x_1 \leq 0.1,\ -0.1 \leq \partial\hat{g}(x|p)/\partial x_2 \leq 0.1)$, $n = 10,000$.

Figure 3 clearly reveals that the regression surface for the restricted model possesses derivatives that satisfy the constraints everywhere and is smooth.

3.3. **A Simulated Illustration: Restricting** $\hat{g}^{(2)}(x)$**.** We consider the same DGP given above, but now we arbitrarily impose the constraint that the second derivatives with respect to both x_1 and x_2 are positive (negative), which is a necessary (but not sufficient) condition for concavity and convexity; see Appendix A for details on imposing concavity or convexity using our approach. As can be seen from figures 4 and 5 the shape of the restricted function changes drastically depending on the curvature restrictions placed upon it.

We could as easily impose restrictions defined perhaps jointly on, say, both $\hat{g}(x)$ and $\hat{g}^{(1)}(x)$, or perhaps on cross-partial derivatives if so desired. We hope that these illustrative applications

[12] $\mathbf{s} = (1,0)$ and $\mathbf{t} = (0,1)$.

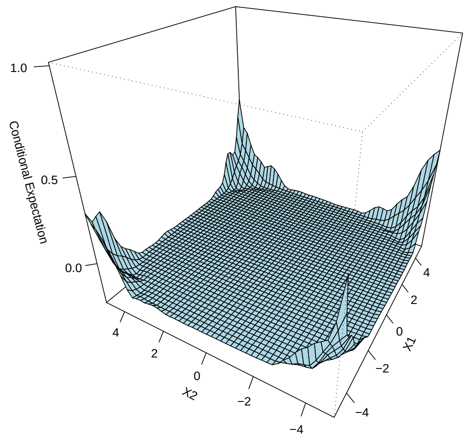

FIGURE 4. Restricted nonparametric estimate of (6) where the restriction is defined over $\hat{g}^{(\mathbf{s})}(x)$, $\mathbf{s} \in \{(2,0),(0,2)\}$ $(\partial\hat{g}^2(x|p)/\partial x_1^2 \geq 0,\ \partial\hat{g}^2(x|p)/\partial x_2^2 \geq 0)$, $n = 10,000$.

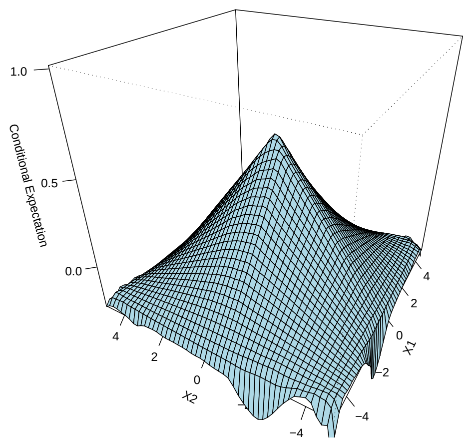

FIGURE 5. Restricted nonparametric estimate of (6) where the restriction is defined over $\hat{g}^{(\mathbf{s})}(x)$, $\mathbf{s} \in \{(2,0),(0,2)\}$ $(\partial\hat{g}^2(x|p)/\partial x_1^2 \leq 0,\ \partial\hat{g}^2(x|p)/\partial x_2^2 \leq 0)$, $n = 10,000$.

reassure the reader that the method we propose is powerful, fully general, and can be applied in large-sample settings.

3.4. **Finite-Sample Performance: Testing for Parametric Functional Form.** We consider testing the restriction that the nonparametric model $g(x)$ is equivalent to a specific parametric functional form (i.e., we impose an equality restriction on $\hat{g}(x)$, namely that $\hat{g}(x)$ equals $x'\hat{\beta}$ where $x'\hat{\beta}$ is the parametric model), by way of illustration. We consider the following DGP:

$$Y_i = g(X_{i1}, X_{i2}) + \epsilon_i = 1 + X_{i1}^2 + X_{i2} + \epsilon_i,$$

where X_{ij}, $j = 1, 2$ are uniform $[-2, 2]$ and $\epsilon \sim N(0, 1/2)$.

We then impose the restriction that $g(x)$ is of a particular parametric form, and test whether this restriction is valid. When we generate data from this DGP and impose the correct model as a restriction (i.e., that given by the DGP, say, $\beta_0 + \beta_1 x_{i1}^2 + \beta_2 x_{i2}$) we can assess the test's size, while when we generate data from this DGP and impose an incorrect model that is in fact linear in variables we can assess the test's power.

We conduct $M = 1,000$ Monte Carlo replications from our DGP, and consider $B = 99$ bootstrap replications; see Racine & MacKinnon (2007*b*) for details on determining the appropriate number of bootstrap replications. Results are presented in Table 1 in the form of empirical rejection frequencies for $\alpha = (0.10, 0.05, 0.01)$ for samples of size $n = 25, 50, 75, 100, 200$.

TABLE 1. Test for correct parametric functional form. Values represent the empirical rejection frequencies for the $M = 1,000$ Monte Carlo replications.

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
		Size	
25	0.100	0.049	0.010
50	0.074	0.043	0.011
75	0.086	0.034	0.008
100	0.069	0.031	0.006
200	0.093	0.044	0.007
		Power	
25	0.391	0.246	0.112
50	0.820	0.665	0.356
75	0.887	0.802	0.590
100	0.923	0.849	0.669
200	0.987	0.970	0.903

Table 1 indicates that the test appears to be correctly sized while power increases with n.

3.5. **Finite-Sample Performance: Testing an Equality Restriction on a Partial Derivative.** For this example we consider a simple linear DGP given by

$$Y_i = g(X_i) + \epsilon_i = \beta_1 X_i + \epsilon_i, \tag{7}$$

where X_i is uniform $[-2, 2]$ and $\epsilon \sim N(0, 1)$.

We consider testing the equality restriction $H_0 : g^{(1)}(x) = 1$ where we take the first order derivative (i.e., $g^{(1)}(x) = dg(x)/dx_1$), and let β_1 vary from 1 through 2 in increments of 0.1. Note that the test of significance would be a test of the hypothesis that $g^{(1)}(x) = 0$ almost everywhere rather than $g^{(1)}(x) = 1$ which we consider, so clearly we could also perform a test of significance in the current framework. The utility of the proposed approach lies in its flexibility as we could as easily test the hypothesis that $g^{(1)}(x) = \xi(x)$ where $\xi(x)$ is an arbitrary function. Significance testing in nonparametric settings has been considered by a number of authors; see Racine (1997) and Racine, Hart & Li (2006) for alternative approaches to testing significance in a nonparametric setting.

When $\beta_1 = 1.0$ we can assess size while when $\beta_1 \neq 1.0$ we can assess power. We construct power curves based on $M = 1,000$ Monte Carlo replications, and we compute $B = 99$ bootstrap replications. The power curves corresponding to $\alpha = 0.05$ appear in Figure 6.

Figure 6 reveals that for small sample sizes (e.g., $n = 25$) there appears to be a small size distortion, however, the distortion appears to fall rather quickly as n increases. Furthermore, power increases with n. Given that the sample sizes considered here would typically be much smaller than those used by practitioners adopting nonparametric smoothing methods, we expect that the proposed test would possess reasonable size in empirical applications.

4. Application: Imposing Constant Returns to Scale for Indonesian Rice Farmers

We consider a production dataset that has been studied by Horrace & Schmidt (2000) who analyzed technical efficiency for Indonesian rice farms. We examine the issue of returns to scale, focusing on one growing season's worth of data for the year 1977, acknowledged to be a particularly wet season. Farmers were selected from six villages of the production area of the Cimanuk River Basin in West Java, and there were 171 farms in total. Output is measured as kilograms of

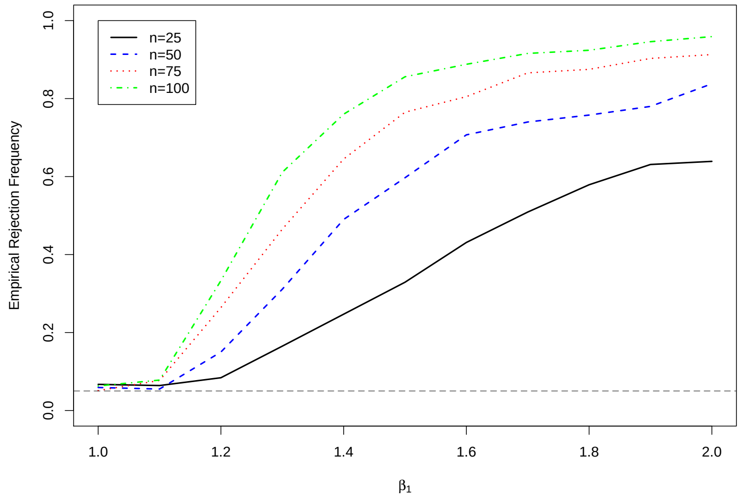

Figure 6. Power curves for $\alpha = 0.05$ for sample sizes $n = (25, 50, 75, 100)$ based upon the DGP given in (7). The dashed horizontal line represents the test's nominal level (α).

rice produced, and inputs included seed (kg), urea (kg), trisodium phosphate (TSP) (kg), labour (hours), and land (hectares). Table 2 presents some summary statistics for the data. Of interest here is whether or not the technology exhibits constant returns to scale (i.e., whether or not the sum of the partial derivatives equals one). We use log transformations throughout.

Table 2. Summary Statistics for the Data

Variable	Mean	StdDev
log(rice)	6.9170	0.9144
log(seed)	2.4534	0.9295
log(urea)	4.0144	1.1039
log(TSP)	2.7470	1.4093
log(labor)	5.6835	0.8588
log(land)	-1.1490	0.9073

We estimate the production function using a nonparametric local linear estimator with cross-validated bandwidth selection. Figure 7 presents the unrestricted and restricted partial derivative sums for each observation (i.e., farm), where the restriction is that the sum of the partial derivatives equals one. The horizontal line represents the restricted partial derivative sum (1.00) and the points represent the unrestricted sums for each farm. An examination of Figure 7 reveals that the estimated returns to scale lie in the interval $[0.98, 1.045]$.

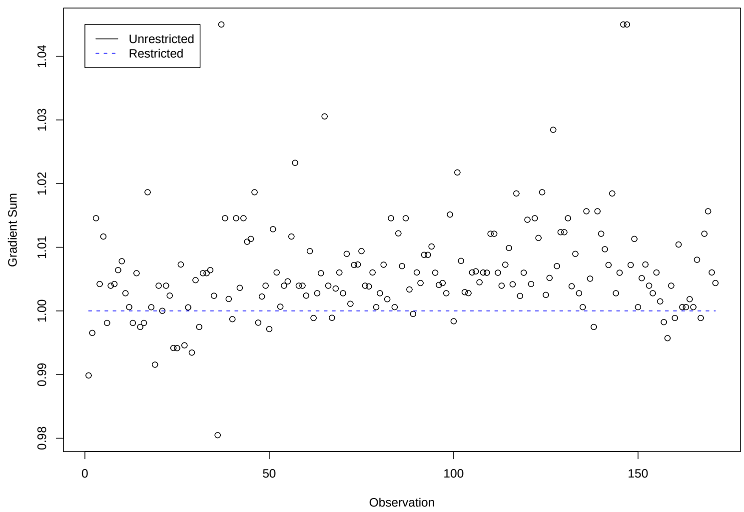

FIGURE 7. The sum of the partial derivatives for observation i (i.e., each farm) appear on the vertical axis, and each observation (farm) appears on the horizontal axis.

Figures 8 and 9 present the unrestricted and restricted partial mean plots, respectively.[13] Notice the change in the partial mean plot of log(urea) across the restricted and unrestricted models. It is clear that the bulk of the restricted weights are targeting this input's influence on returns to scale. The remaining partial mean plots are unchanged visually across the unrestricted and restricted models.

In order to test whether the restriction is valid we apply the test outlined in Section 2.2. We conducted $B = 99$ bootstrap replications and test the null that the technology exhibits constant returns to scale. The empirical P value is $P_B = 0.131$, hence we fail to reject the null at all conventional levels. We are encouraged by this fully nonparametric application particularly as it involves a fairly large number of regressors (five) and a fairly small number of observations ($n = 171$).

[13] A 'partial mean plot' is simply a 2D plot of the outcome y versus one covariate x_j when all other covariates are held constant at their respective medians/modes.

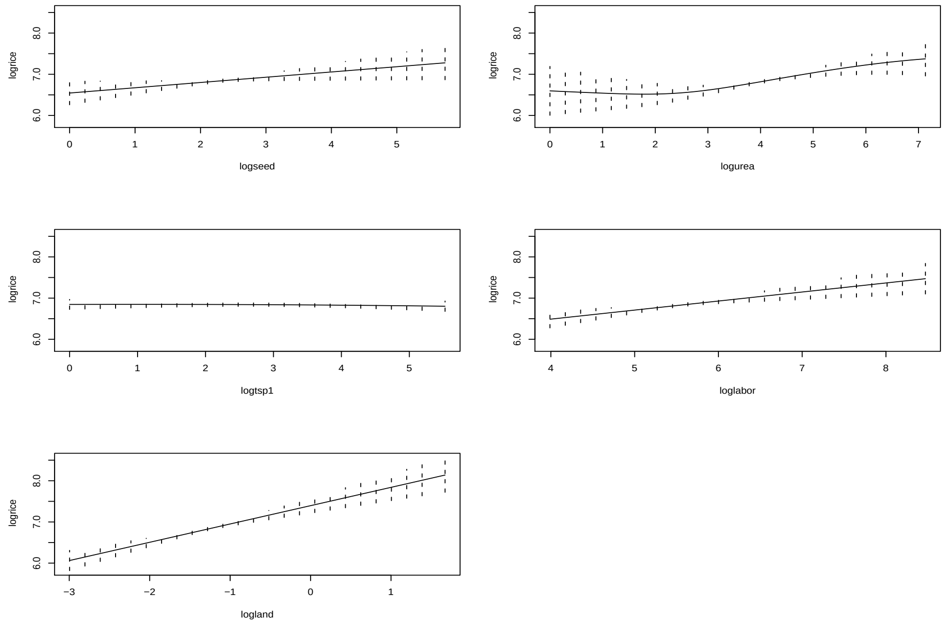

Figure 8. Partial mean plots for the unrestricted production function.

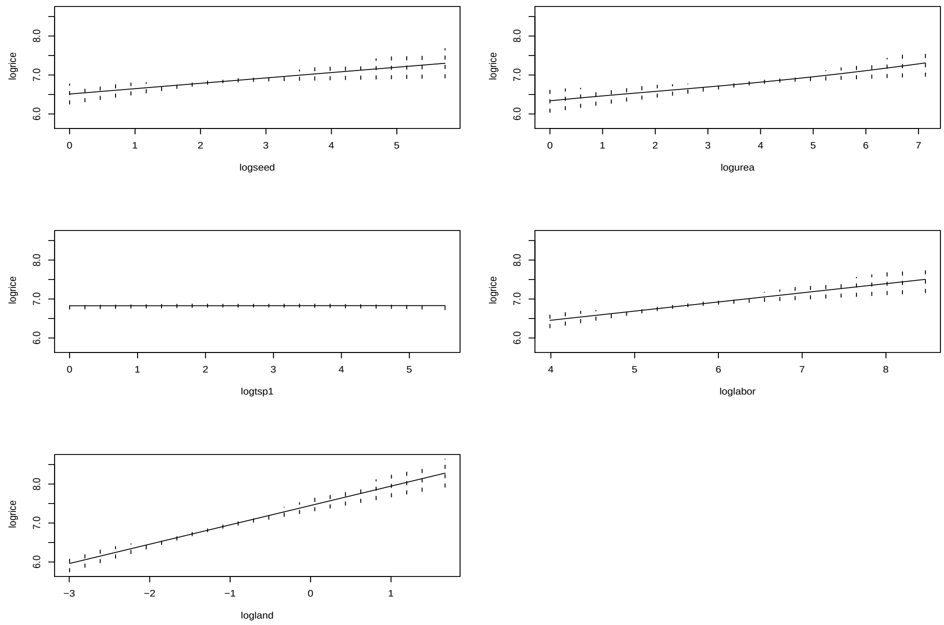

Figure 9. Partial mean plots for the restricted production function.

5. Concluding Remarks

We present a framework for imposing and testing the validity of arbitrary constraints on the sth partial derivatives of a nonparametric kernel regression function, namely, $l(x) \leq g^{(\mathbf{s})}(x) \leq$

$u(x)$, $s = 0, 1, \ldots$. The proposed approach nests special cases such as imposing monotonicity, concavity (convexity) and so forth while delivering a seamless framework for general restricted nonparametric kernel estimation and inference. Illustrative simulated examples are presented, finite-sample performance of the proposed test is examined via Monte Carlo simulations, and an illustrative application is undertaken. An open implementation in the R language (R Development Core Team (2008)) is available from the authors.

One interesting extension of this methodology would be to the cost system setup popular in production econometrics (Kumbhakar & Lovell (2001)). There, the derivatives of the cost function are estimated along with the function itself in a system framework. Currently, Hall & Yatchew (2007) have proposed a method for estimating the cost function based upon integrating the share equations, resulting in an improvement in the rate of convergence relative to direct nonparametric estimation of the cost function. It would be interesting to determine the merits of restricting the first order partial derivatives of the cost function using the approach described here to estimate the cost function in a single equation framework. We also note that the procedure we outline is valid for a range of kernel estimators in addition to those discussed herein. Semiparametric models such as the partially linear, single index, smooth coefficient, and additively separable models could utilize this approach towards constrained estimation. Nonparametric unconditional and conditional density and distribution estimators, as well as survival and hazard functions, smooth conditional quantiles and structural nonparametric estimators including auction methods could also benefit from the framework developed here. We leave this as a subject for future research.

References

Abrevaya, J. & Jiang, W. (2005), 'A nonparametric approach to measuring and testing curvature', *Journal of Business and Economic Statistics* **23**, 1–19.

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 restrictions on partial derivatives. *Mimeo.*

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

Briesch, R. A., Chintagunta, P. K. & Matzkin, R. L. (2002), 'Semiparametric estimation of brand choice behavior', *Journal of the American Statistical Association* **97**, 973–982.

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

Chen, H. Z. & Randall, A. (1997), 'Semi-nonparametric estimation of binary response models with an application to natural resource valuation', *Journal of Econometrics* **76**, 323–340.

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

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.

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

Epstein, L. G. & Yatchew, A. J. (1985), 'Nonparametric hypothesis testing procedures and applications to demand analysis', *Journal of Econometrics* **30**, 149–169.

Fan, J. (1992), 'Design-adaptive nonparametric regression', *Journal of the American Statistical Association* **87**(420), 998–1004.

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.

Ghosal, S., Sen, A. & van der Vaart, A. W. (2000), 'Testing monotonicity of regression', *Annals of Statistics* **28**(4), 1054–1082.

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., Huang, H., Gifford, J. & Gijbels, I. (2001), 'Nonparametric estimation of hazard rate under the constraint of monotonicity', *Journal of Computational and Graphical Statistics* **10**(3), 592–614.

Hall, P. & Kang, K. H. (2005), 'Unimodal kernel density estimation by data sharpening', *Statistica Sinica* **15**, 73–98.

Hall, P. & Yatchew, A. J. (2007), 'Nonparametric estimation when data on derivatives are available', *Annals of Statistics* **35**(1), 300–323.

Hanoch, G. & Rothschild, M. (1972), 'Testing the assumptions of production theory: A nonparametric approach', *Journal of Political Economy* **80**, 256–275.

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

Henderson, D. J. & Parmeter, C. F. (2008), Imposing nonlinear constraints in a smooth nonparametric framework. Virginia Tech Working Paper 07/08.

Horrace, W. & Schmidt, P. (2000), 'Multiple comparisons with the best, with economic applications', *Journal of Applied Econometrics* **15**, 1–26.

Kelly, C. & Rice, J. (1990), 'Monotone smoothing with application to dose response curves and the assessment of synergism', *Biometrics* **46**, 1071–1085.

Kumbhakar, S. C. & Lovell, C. A. K. (2001), *Stochastic Frontier Analysis*, Cambridge University Press.

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

Li, W., Naik, D. & Swetits, J. (1996), 'A data smoothing technique for piecewise convex/concave curves', *SIAM Journal on Scientific Computing* **17**, 517–537.

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

Mammen, E. & Thomas-Agnam, C. (1999), 'Smoothing splines and shape restrictions', *Scandinavian Journal of Statistics* **26**, 239–252.

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 theory 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.

Nadaraya, E. A. (1965), 'On nonparametric estimates of density functions and regression curves', *Theory of Applied Probability* **10**, 186–190.

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

Pelckmans, K., Espinoza, M., Brabanter, J. D., Suykens, J. A. K. & Moor, B. D. (2005), 'Primal-dual monotone kernel regression', *Neural Processing Letters* **22**, 171–182.

Priestley, M. B. & Chao, M. T. (1972), 'Nonparametric function fitting', *Journal of the Royal Statistical Society* **34**, 385–392.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Racine, J. S. (1997), 'Consistent significance testing for nonparametric regression', *Journal of Business and Economic Statistics* **15**(3), 369–379.

Racine, J. S., Hart, J. D. & Li, Q. (2006), 'Testing the significance of categorical predictor variables in nonparametric regression models', *Econometric Reviews* **25**, 523–544.

Racine, J. S. & Li, Q. (2004), 'Nonparametric estimation of regression functions with both categorical and continuous data', *Journal of Econometrics* **119**(1), 99–130.

Racine, J. S. & MacKinnon, J. G. (2007*a*), 'Inference via kernel smoothing of bootstrap P values', *Computational Statistics and Data Analysis* **51**, 5949–5957.

Racine, J. S. & MacKinnon, J. G. (2007*b*), 'Simulation-based tests that can use any number of simulations', *Communications in Statistics* **36**, 357–365.

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

Robertson, T., Wright, F. & Dykstra, R. (1988), *Order Restricted Statistical Inference*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons.

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

Turlach, B. A. (1997), Constrained smoothing splines revisited. *Mimeo*, Australian National University.

Varian, H. R. (1985), 'Nonparametric analysis of optimizing behavior with measurement error', *Journal of Econometrics* **30**, 445–458.

Watson, G. S. (1964), 'Smooth regression analysis', *Sankhya* **26:15**, 175–184.

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.

Appendix A. The Quadratic Program for Joint Monotonicity and Concavity

The method outlined in this paper requires the solution of a standard quadratic programming problem when the (in)equality constraints are linear in p. When our set of constraints is nonlinear in p, we can modify the problem to still allow for the use of standard off-the-shelf quadratic programming methods, which is computationally appealing. This appendix spells out in greater detail how to implement an appropriate quadratic program to solve for a vector of weights that will ensure a regression function is both monotonic (a constraint that is linear in p) and concave (a constraint that is nonlinear in p). For a more general overview of the procedures used to determine a set of weights when a user imposes nonlinear (in p) constraints on a regression function we refer the reader to Henderson & Parmeter (2008), though they restrict attention to probability weights and the power divergence metric of Cressie & Read (1984) whose limitations in the current setting are discussed in Section 1.

Suppose one wished to impose monotonicity and concavity in a two variable regression setting which involves jointly imposing constraints that are linear and nonlinear in p. We wish to minimize $D(p) = (p_u - p)'(p_u - p)$ subject to $\partial\hat{g}(x|p)/\partial x_1 \geq 0$, $\partial\hat{g}(x|p)/\partial x_2 \geq 0$, $H(x)$ (the Hessian of the estimated regression function) being negative semi-definite $\forall\, x \in \mathbb{R}^2$ and $\sum_{i=1}^{n} p_i = 1$. The first two conditions imply monotonicity of the regression function for each covariate, while the third condition gives us concavity of the function. The set of linear constraints for the quadratic program can be represented in matrix form as

$$B^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}, \tag{8}$$

$$C_1^T = \begin{bmatrix} A_1^{(1,0)}(x_1)Y_1 & \cdots & A_1^{(1,0)}(x_n)Y_1 \\ A_2^{(1,0)}(x_1)Y_2 & \cdots & A_2^{(1,0)}(x_n)Y_2 \\ \vdots & \ddots & \vdots \\ A_n^{(1,0)}(x_1)Y_n & \cdots & A_n^{(1,0)}(x_n)Y_n \end{bmatrix}, \tag{9}$$

and

$$C_2^T = \begin{bmatrix} A_1^{(0,1)}(x_1)Y_1 & \cdots & A_1^{(0,1)}(x_n)Y_1 \\ A_2^{(0,1)}(x_1)Y_2 & \cdots & A_2^{(0,1)}(x_n)Y_2 \\ \vdots & \ddots & \vdots \\ A_n^{(0,1)}(x_1)Y_n & \cdots & A_n^{(0,1)}(x_n)Y_n \end{bmatrix}. \tag{10}$$

Solving the quadratic program subject to $B^T p = 1$ and $C_1^T p \geq 0$ and $C_2^T p \geq 0$ will impose the adding up constraint on the weights and monotonicity. However, guaranteeing concavity of the regression function requires a modified approach.

Recall that for a matrix to be negative semi-definite the signs of the determinants of the principal minors must alternate in sign, beginning with a negative or zero value. That is, we need $|H_1^*| \leq 0$, $|H_2^*| \geq 0, \dots, |H_k^*| = |H| \geq 0$ if k is even (≤ 0 if k is odd), where $|\cdot|$ denotes determinant. Aside from the principal minors of order one, the determinant of the remaining principal minor is nonlinear in the ps. In our two variable setting we therefore need to have $\partial^2 g(x|p)/\partial x_1^2 \leq 0$, $\partial^2 g(x|p)/\partial x_2^2 \leq 0$, and $(\partial^2 g(x|p)/\partial x_1^2) \times (\partial^2 g(x|p)/\partial x_2^2) - (\partial^2 g(x|p)/\partial x_2 \partial x_1)^2 \geq 0$. The first two constraints are linear in p and can be written in matrix form as

$$C_3^T = \begin{bmatrix} A_1^{(2,0)}(x_1)Y_1 & \cdots & A_1^{(2,0)}(x_n)Y_1 \\ A_2^{(2,0)}(x_1)Y_2 & \cdots & A_2^{(2,0)}(x_n)Y_2 \\ \vdots & \ddots & \vdots \\ A_n^{(2,0)}(x_1)Y_n & \cdots & A_n^{(2,0)}(x_n)Y_n \end{bmatrix}, \tag{11}$$

and

$$C_4^T = \begin{bmatrix} A_1^{(0,2)}(x_1)Y_1 & \cdots & A_1^{(0,2)}(x_n)Y_1 \\ A_2^{(0,2)}(x_1)Y_2 & \cdots & A_2^{(0,2)}(x_n)Y_2 \\ \vdots & \ddots & \vdots \\ A_n^{(0,2)}(x_1)Y_n & \cdots & A_n^{(0,2)}(x_n)Y_n \end{bmatrix}. \tag{12}$$

The last constraint can to be linearized with respect to p and one could then iterate this procedure using sequential quadratic programming (see Nocedal & Wright (2000, Chapter 18)). Letting $g_{rs}(x|p) = \sum_{i=1}^{n} A_i^{(r,s)}(x) Y_i p_i$, the linearized version of the determinant of the second order cross partial is

$$C_5^T = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ c_{21} & \cdots & c_{2n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}, \tag{13}$$

where $c_{vw} = g_{11}(x_w|p)A_v^{(0,2)}(x_w)Y_v + g_{22}(x_w|p)A_v^{(2,0)}(x_w)Y_v - 2g_{12}(x_w|p)A_v^{(1,1)}(x_w)Y_v$. To solve for the vector of weights consistent with both monotonicity and concavity, the quadratic program would be solved using B and C_1 through C_5 to obtain an initial solution. This solution would then augment the starting value of p to become an updated solution. The process would then be iterated until convergence of the ps occurs. See Henderson & Parmeter (2008) for a more detailed explanation of this process.

Appendix B. R code to replicate the example in Section 3.1

We provide R code (R Development Core Team (2008)) to replicate the example in Section 3.1. Ignoring the code that generates the data for this example, the approach requires only 12 simple commands involving straightforward code and a call to a short routine that follows which generates the weights necessary for solving the quadratic programming problem (the rest of the code is used to generate the estimation and evaluation data). To allow the user to test the code on a trivial dataset we have changed the number of observations to $n = 250$ and evaluate on a grid of size 25x25 (instead of 10,000 and 50x50 used in Section 3.1).

```
library(np)
library(quadprog)

n <- 250
n.eval <- 25
x.min <- -5
x.max <- 5
lower <- 0.0
upper <- 0.5

## The following loads a simple function that will return the
## weight matrix multiplied by n

source("Aymat_train_eval.R")

## Generate a draw from the DGP

x1 <- runif(n,x.min,x.max)
x2 <- runif(n,x.min,x.max)
y <- sin(sqrt(x1^2+x2^2))/sqrt(x1^2+x2^2) + rnorm(n,sd=.1)
data <- data.frame(y,x1,x2)
rm(y,x1,x2)

## Create the evaluation data matrix

data.eval <- data.frame(y=0,expand.grid(x1=seq(x.min,x.max,length=n.eval),
                                        x2=seq(x.min,x.max,length=n.eval)))

## Now that we have generated the data, here is the body of the code
## (12 commands excluding comments)

## Generate the cross-validated local linear bandwidth object
## using the np package, then compute the unrestricted model
```

```
## and gradients using the np package

bw <- npregbw(y~x1+x2,regtype="ll",tol=.1,ftol=.1,nmulti=1,data=data)
model.unres <- npreg(bws=bw,data=data,newdata=data.eval,gradients=TRUE)

## Start from uniform weights equal to 1/n, generate p, Dmat, and dvec
## which are fed to the quadprog() function

p <- rep(1/n,n)
Dmat <- diag(1,n,n)
dvec <- as.vector(p)

## Generate the weight matrix

Aymat.res <- Aymat(0,data,data.eval,bw)

## Create Amat which is fed to the quadprog() function. The first line
## contains the adding to one constraint, the next blocks contain the
## lower and upper bound weighting matrices.

Amat <- t(rbind(rep(1,n),Aymat.res,-Aymat.res))

rm(Aymat.res)

## Create bvec (the vector of constraints) which is fed to the
## quadprog() function

bvec <- c(0,(rep(lower,n.eval)-fitted(model.unres)),
          (fitted(model.unres)-rep(upper,n.eval)))

## Solve the quadratic programming problem

QP.output <- solve.QP(Dmat=Dmat,dvec=dvec,Amat=Amat,bvec=bvec,meq=1)

## That's it. Now extract the solution and update the uniform weights

p.updated <- p + QP.output$solution

## Now estimate the restricted model using the np package and you are done.

data.trans <- data.frame(y=p.updated*n*data$y,data[,2:ncol(data)])
model.res <-
npreg(bws=bw,data=data.trans,newdata=data.eval,gradients=TRUE)

## You could then, say, plot the restricted estimate if you wished.

plot(model.res,data=data.trans)
```

Here is the `Aymat` code located in `source("Aymat_train_eval.R")` called by the above example. It returns the weight matrix for the local linear estimator and its derivatives multiplied by n.

```
Aymat <- function(j.reg=1,mydata.train,mydata.eval,bw) {
  y <- mydata.train[,1]
  n.train=nrow(mydata.train)
  n.eval=nrow(mydata.eval)

  X.train <- as.data.frame(mydata.train[,-1])
  names(X.train) <- names(mydata.train)[-1]
  X.eval <- as.data.frame(mydata.eval[,-1])
  names(X.eval) <- names(mydata.eval)[-1]
  X.col.numeric <- sapply(1:ncol(X.train),function(i){is.numeric(X.train[,i])})

  k <- ncol(as.data.frame(X.train[,X.col.numeric]))

  Aymat <- matrix(NA,nrow=n.eval,ncol=n.train)

  iota <- rep(1,n.train)

  for(j in 1:n.eval) {

    evalmat <- as.data.frame(t(matrix(as.numeric(X.eval[j,X.col.numeric]), k,n)))
    names(evalmat) <- names(X.eval)[X.col.numeric]

    W <- as.matrix(data.frame(iota,X.train[,X.col.numeric]-evalmat))

    K <- npksum(txdat=X.eval[j,],
                exdat=X.train,
                bws=bw$bw)$ksum

    Wmat.sum.inv <- solve(npksum(exdat=X.eval[j,],
                                 txdat=X.train,
                                 tydat=W,
                                 weights=W,
                                 bws=bw$bw)$ksum[,,1])

    Aymat[j,] <- sapply(1:n,
                   function(i){(Wmat.sum.inv %*% W[i,]*K[i]*y[i])[(j.reg+1)]})

  }

  return(n.train*Aymat)
}
```