cemmap

centre for microdata methods and practice

Nonparametric Estimation and Inference under Shape Restrictions

Joel Horowitz
Sokbae Lee

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP67/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

NONPARAMETRIC ESTIMATION AND INFERENCE UNDER SHAPE RESTRICTIONS

by

Joel L. Horowitz
Northwestern University
Evanston, IL 60208
U.S.A.

and

Sokbae Lee
Seoul National University
Seoul, Republic of Korea

October 2015

Abstract

Economic theory often provides shape restrictions on functions of interest in applications, such as monotonicity, convexity, non-increasing (non-decreasing) returns to scale, or the Slutsky inequality of consumer theory; but economic theory does not provide finite-dimensional parametric models. This motivates nonparametric estimation under shape restrictions. Nonparametric estimates are often very noisy. Shape restrictions stabilize nonparametric estimates without imposing arbitrary restrictions, such as additivity or a single-index structure, that may be inconsistent with economic theory and the data. This paper explains how to estimate and obtain an asymptotic uniform confidence band for a conditional mean function under possibly nonlinear shape restrictions, such as the Slutsky inequality. The results of Monte Carlo experiments illustrate the finite-sample performance of the method, and an empirical example illustrates its use in an application.

We thank David Jacho-Chávez for providing the data used in this paper. Part of this research was carried out while Joel L. Horowitz was a visitor at the Department of Economics, University College London, and the Centre for Microdata Methods and Practice.

NONPARAMETRIC ESTIMATION AND INFERENCE UNDER SHAPE RESTRICTIONS

1. INTRODUCTION

Let Y be a scalar random variable and X be a scalar random variable or vector. This paper presents a method for nonparametrically estimating and carrying out inference about the conditional mean function

$$g(x) \equiv E(Y \mid X = x)$$

under a shape restriction on g such as monotonicity, convexity, non-increasing (non-decreasing) returns to scale, or the Slustky inequality of consumer theory. Economic theory often provides shape restrictions but does not provide finite-dimensional parametric models. For example, cost functions are monotone increasing, concave in input prices, and may exhibit non-increasing or non-decreasing returns to scale. Demand functions satisfy the Slutsky inequality, which is nonlinear. This motivates nonparametric estimation under shape restrictions. This paper explains how to estimate and form a uniform confidence band for g under shape restrictions that are more complicated than monotonicity or convexity and may be nonlinear.

It is well known that g can be estimated consistently and with the optimal rate of convergence without imposing shape restrictions. Fan and Gijbels (1996) and Härdle (1990), among many others, describe nonparametric estimation and rates of convergence without shape restrictions. Mammen (1991a, 1991b), Mammen and Thomas-Agnan (1999), and Wang and Shen (2013) discuss rates of convergence with shape restrictions. However, fully nonparametric estimates can be noisy and inconsistent with economic theory due to random sampling errors. For example, Blundell, Horowitz, and Parey (2012, 2013) found fully nonparametric estimates of demand functions to be wiggly and non-monotonic. Blundell, Horowitz, and Parey (2012, 2013) also found that imposing the Slutsky restriction reduced random noise and led to well-behaved nonparametric estimates without the need for arbitrary and possibly incorrect parametric or semiparametric assumptions.

Many methods are available for carrying out consistent nonparametric estimation under shape restrictions. See, for example, Hall and Huang (2001, 2002); Hall, Huang, Gifford, and Gijbels (2001); Hall and Presnell (1999); and the references cited in the foregoing paragraph. Asymptotic inference is not difficult if the values of x at which the shape restriction binds or does not bind in the sampled population are known. Liew (1976) illustrates this in the context of inequality constrained estimation of a linear model. Du, Parmeter, and Racine (2013) carry out kernel nonparametric estimation. In applications, however, it is not known where in the sampled population the shape restriction does or does not bind. This greatly complicates inference, because random sampling errors can cause the shape restriction to bind or not bind the estimated and true g at different values of x. A similar problem arises in inference about a finite-dimensional parameter that may be on the boundary of the parameter set (Andrews 1999). Existing results on inference about a shape-restricted, nonparametrically estimated conditional mean function are limited to functions that are assumed to be monotonic or convex. The literature on inference under monotonicity or convexity restrictions is vast. See, among many others, Birke and Dette (2006); Dumbgen (2003); Chernozhukov, Fernandez-Val, and Galichon (2009); Dette, Neumeyer, and Pilz (2006); Groeneboom, Jongbloed, and Wellner (2001); Pal and Woodroofe (2007); and the references therein. Existing results do not treat shape restrictions such as increasing or decreasing returns to scale and the Slutsky inequality that are of particular importance in economics. There is also a large literature on testing the hypothesis that a shape restriction holds. See, for example, Romano, Shaikh, and Wolf (2014); Andrews and Shi (2013); Lee, Song, and Whang (2013); Chernozhukov, Lee, and Rosen (2013); Hall and Yatchew (2005); and the references therein.

This paper is concerned with inference under shape restrictions, such as the Slutsky restriction, that may be nonlinear in a sense that is defined in Section 4. We formulate the estimation problem as minimization of a local quadratic objective function subject to constraints that implement the shape restriction. In general, the shape restriction generates a continuum of constraints. We reduce the number of constraints to a finite value by imposing the shape restriction and estimating g only on a discrete grid of points x in the support of X. The grid becomes finer as the sample size, n, increases, thereby ensuring that, asymptotically, the shape restriction holds everywhere in the support of X. This enables us to obtain a confidence band for g that, asymptotically, is uniform over the support of X and satisfies the shape restriction.

The use of a discrete grid of points x enables us to overcome the problem of not knowing which constraints are binding in the sampled population. Let $\mathcal{S}$ be the set of constraints that bind in the population. This set is unknown. We find a data-based set $\hat{\mathcal{S}}$ of "possibly binding constraints" and carry out estimation under the (possibly false) assumption that $\hat{\mathcal{S}} = \mathcal{S}$. We show that $\hat{\mathcal{S}} = \mathcal{S}$ with probability approaching 1 as $n \to \infty$. Consequently $\mathcal{S}$ can be treated as known asymptotically, and asymptotic inference can be carried out as if $\mathcal{S}$ were known and $\hat{\mathcal{S}} = \mathcal{S}$.

Let $g_0(x)$ and $\hat{g}(x)$, respectively, denote the true conditional mean function and the shape-restricted nonparametric estimator. We show that with suitable scaling, $\hat{g}(x) - g_0(x)$ is asymptotically jointly normally distributed with mean 0 over grid points. Asymptotic normality makes it possible to obtain a confidence band for g_0 that is uniform over grid points. As $n \to \infty$ and the distance between grid points approaches 0, the uniform confidence band over grid points converges to a uniform confidence band over all values of x.

Estimation of $g(x)$ at points x that are not in the grid is unnecessary for forming an asymptotic uniform confidence band for g but may be of interest for other reasons. Estimation of $g(x_{new})$ at a point x_{new} that is not in the grid can be carried out using the methods of this paper by shifting the location of the grid so that x_{new} is a point of the shifted grid. Alternatively, $g(x_{new})$ can be estimated using any of a variety of methods for interpolating $g(x)$ between grid points subject to the shape restrictions. The choice among interpolation methods is arbitrary and, except in special cases, does not yield an estimator that converges in probability as rapidly as an estimator based on the shifted grid.

Section 2 outlines the main steps involved in implementing our method. Section 3 presents the unconstrained and constrained nonparametric estimators of g and defines the grid. Section 4 describes the method for finding the set $\hat{\mathcal{S}}$ of possibly binding constraints. Section 5 explains how to carry out inference about g and form a uniform confidence band for g under shape restrictions. To minimize notational complexity, the discussion in Sections 2-5 assumes that X is a scalar random variable. The extension to higher dimensions is outlined in Section 6. Section 7 presents the results of Monte Carlo experiments and an empirical example that illustrate the numerical performance of our methods. Section 8 presents concluding comments. The proofs of theorems are in the appendix, which is Section 9.

2. A GUIDE TO IMPLEMENTATION

This section outlines the main steps of our method for estimating and obtaining a uniform confidence band for g. We assume here that X is a scalar random variable whose support is $[0,1]$. The extension to a multidimensional X is presented in Section 6.

1. Define a grid $0 < x_1 < x_2 < ... < x_J < 1$ of J equally spaced points on $(0,1)$. A data-based method for choosing J in applications is presented in Section 7.

2. Estimate $g(x_j)$ ($j = 1,...,J$) nonparametrically by using local quadratic estimation with bandwidth h. Let $\tilde{g}(x_j)$ denote the resulting estimate. A method for choosing h in applications is presented in Section 3.1.

3. Use the estimates $\tilde{g}(x_j)$ to find the set $\hat{\mathcal{S}}$ of possibly binding shape constraints. $\hat{\mathcal{S}}$ is given by equation (4.7).

4. Re-estimate $g(x_j)$ ($j = 1,...,J$) nonparametrically using constrained local quadratic estimation under the restriction that the shape constraints in $\hat{S}$ are binding (that is, they are equalities) and ignoring all other shape constraints.

5. Form a uniform confidence band for g using the method of equation (5.8) or (5.10).

3. THE ESTIMATORS OF g

This section describes our methods for estimating g with and without shape restrictions. The unrestricted estimator is used to estimate the set of possibly binding constraints. The shape-restricted estimator is an extension of the unrestricted estimator. Section 3.1 presents the unrestricted estimator. Section 3.2 presents grid and the shape-restricted estimator.

3.1 *The Unrestricted Estimator*

This section presents the unrestricted nonparametric estimator of g that is used throughout the remainder of this paper. Let $\{Y_i, X_i : i = 1,...,n\}$ denote an independent random sample from the distribution of (Y, X). Assume for now that X is a scalar random variable. The extension to a multidimensional X is presented in Section 6. Also assume that the support of X is a compact interval. Without further loss of generality, let this interval be $[0,1]$.

We use local quadratic estimation with bandwidth $h \propto n^{-1/5}$ to obtain the unrestricted nonparametric estimator of g. In applications, the bandwidth can be chosen by using cross-validation or plug-in methods for local constant or local linear estimation. Under our assumptions, local quadratic estimation with $h \propto n^{-1/5}$ provides an estimator of g that is free of asymptotic bias, and the bandwidth can be selected by standard methods. Local constant, local linear, and series estimation methods do not have this property. They require undersmoothing or explicit bias correction to prevent asymptotic bias, and this requires choice of an auxiliary bandwidth (or series length in the case of series estimation). There are no satisfactory data-based methods for choosing the auxiliary bandwidth or series length. Hall and Horowitz (2013) provide numerical illustrations of this problem. Calonico, Cattaneo, and Farrell (2014) present an alternative form of undersmoothing that does not require an auxiliary bandwidth. This method has some desirable theoretical properties but is more complex than the one used here.

The following notation is used to define the unrestricted estimator of $g(x)$ and in the remainder of this paper. Let K denote a probability density function that is supported on $[-1,1]$ and symmetrical about 0. For any $v \in [-1,1]$, let $K_h(v) = K(v/h)$. Let $U = Y - g(X)$ and $\sigma_U^2 = Var(U)$. For any $x \in [0,1]$, let N_x denote the interval $[x-h, x+h]$. Define

$$n_x = \sum_{i=1}^{n} I(|X_i - x| \le h).$$

For $X_{i_1},...,X_{i_{n_x}} \in N_x$, define the $n_x \times 3$ matrix

$$\boldsymbol{X}^{(x)} = \begin{bmatrix} 1 & (X_{i_1} - x) & (X_{i_1} - x)^2 \\ & ..... & \\ 1 & (X_{i_{n_x}} - x) & (X_{i_{n_x}} - x)^2 \end{bmatrix},$$

the $n_x \times n_x$ diagonal matrix

$$\boldsymbol{W}^{(x)} = diag[K_h(X_i - x)]:\ X_i \in N_x],$$

and the 3×3 matrix

$$\boldsymbol{S}_n^{(x)} = n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{X}^{(x)}.$$

Also, for $X_{i_1},...,X_{i_{n_x}} \in N_x$, define the $n_x \times 1$ vectors $\boldsymbol{Y}^{(x)} = (Y_{i_1},...,Y_{i_{n_x}})'$, $\boldsymbol{U}^{(x)} = (U_{i_1},...,U_{i_{n_x}})'$, and $\boldsymbol{g}^{(x)} = [g(X_{i_1}),...,g(X_{i_{n_x}})]'$.

Now let $\boldsymbol{b} = (b_1, b_2, b_3)'$ be a 3×1 vector, and let

$$\tilde{\boldsymbol{b}}(x) \equiv [\tilde{b}_1(x), \tilde{b}_2(x), \tilde{b}_3(x)]' = \arg\min_{\boldsymbol{b}} (\boldsymbol{Y}^{(x)} - \boldsymbol{X}^{(x)}\boldsymbol{b})' \boldsymbol{W}^{(x)} (\boldsymbol{Y}^{(x)} - \boldsymbol{X}^{(x)}\boldsymbol{b})$$

$$= \arg\min_{\boldsymbol{b}} (\boldsymbol{b}' \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{X}^{(x)} \boldsymbol{b} - 2\boldsymbol{Y}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{X}^{(x)} \boldsymbol{b}).$$

The unrestricted estimator of $g(x)$ is $\tilde{g}(x) = \tilde{b}_1$. Standard algebra of least squares estimation shows that

$$\tilde{\boldsymbol{b}}(x) = (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{Y}^{(x)}$$

$$= (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{U}^{(x)} + (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{g}^{(x)}.$$

Therefore,

$$\tilde{g}(x) - g(x) = e_1' (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{U}^{(x)} + e_1' (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{g}^{(x)} - g(x),$$

where $e_1 = (1,0,0)'$.

Now make the following assumptions:

Assumption 1: $\{Y_i, X_i : i = 1,...,n\}$ is an independent random sample from the distribution of (Y, X), where (i) $\text{supp}(X) = [0,1]$. (ii) $Y = g(X) + U$, $E(U \mid X = x) = 0$ and $E(U^2 \mid X = x) = \sigma_U^2$ (a finite constant) for every $x \in [0,1]$. (iii) g satisfies the assumed shape restriction. (iv) $E(|U|^3) < \infty$.

Assumption 2: (i) $g(x)$ is four times continuously differentiable at each $x \in [0,1]$. (ii) The distribution of X has a probability density function with respect to Lebesgue measure, f_X, that is continuously differentiable everywhere in $[0,1]$. (iii) $f_X(x) \ge \delta$ for some $\delta > 0$ and every $x \in [0,1]$.

Assumption 3: (i) K is a bounded probability density function that is supported on $[-1,1]$ and symmetrical about 0; (ii) $h = cn^{-1/5}$ for some finite $c > 0$.

Except for Assumptions 1(iii) and 1(iv), these are standard assumptions in local polynomial nonparametric estimation. Assumption 1(iii) ensures that the shape restricted model is not misspecified. Assumption 1(iv) used in Section 3 to ensure that $P(\hat{\mathcal{S}} = \mathcal{S}) \to 1$, where $\mathcal{S}$ is the unknown set of constraints that bind in the population and $\hat{\mathcal{S}}$ is the data-based set of possibly binding constraints. Assumption 1(iv) is also used in Section 4 to obtain the asymptotic distribution of the constrained estimator of g. Assumption 1(ii) requires U to be homoscedastic. This assumption can be removed at the cost of a much more complex estimation procedure than the one presented here. Assumptions 2 and 3 make the local quadratic estimator undersmoothed, as is necessary to avoid asymptotic bias in the estimator of g. The assumption that g has four continuous derivatives is stronger than needed to obtain the asymptotic distributional results presented in this paper. The results can be obtained under the assumption that g is twice continuously differentiable. However, this requires choosing an undersmoothing bandwidth or an auxiliary bandwidth for explicit bias correction. There are no satisfactory empirical methods for making these choices in applications. The method of Calonico, Cattaneo, and Farrell (2014) permits g to have three derivatives at the cost of greater complexity than the method used here.

The following proposition states the properties of $\tilde{g}(x)$ that are used in this paper.

Proposition 1: Let Assumptions 1(i), 1(ii), 2, and 3 hold. For each $x \in (0,1)$

(3.1) $$(nh)^{1/2}[\tilde{g}(x) - g(x)] = (nh)^{1/2} e_1'(\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{U}^{(x)} + r_n .$$

and

(3.2) $$(nh)^{1/2}[\tilde{g}(x) - g(x)] \to^d N(0, \sigma_{\tilde{g}(x)}^2),$$

where $r_n \le ch^2$ for all x and some constant $c < \infty$, and

$$\sigma^2_{\tilde{g}(x)} = C_K \frac{\sigma^2_U}{f_X(x)},$$

where C_K is a constant that depends on K but not on x, g, or h. ■

Result (3.1) follows from Theorem 3.1 of Fan and Gijbels (1996). See, also, Ruppert and Wand (1994). Result (3.2) is obtained by applying the Lindeberg-Levy central limit theorem to the first term on the right-hand side of (3.1). Fan and Gijbels (1996, p. 62) give the formula for C_K. In applications, $f_X(x)$ and σ^2_U can be replaced with consistent estimators to give the consistent estimator of $\sigma^2_{\tilde{g}(x)}$,

$$\hat{\sigma}^2_{\tilde{g}(x)} = C_K \frac{\hat{\sigma}^2_U}{\hat{f}_X(x)},$$

where $\hat{f}_X(x)$ is a consistent estimator of $f_X(x)$ (e.g., a kernel nonparametric density estimator) and $\hat{\sigma}^2_U$ is a consistent estimator of σ^2_U. When X is a scalar, σ^2_U can be estimated by the method of Rice (1984); Gasser, Sroka, and Jennen-Steinmetz (1986); and Buckley, Eagleson, and Silverman (1988). To construct this estimator, let $X_{(1)} < X_{(2)} < ... < X_{(n)}$ be the ordered sequence of X_i 's. and let $\{Y_{(i)}\}$ be the similarly ordered values of the Y_i 's. The estimator of σ^2_U is

$$\hat{\sigma}^2_U = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} [Y_{(i+1)} - Y_{(i)}]^2 .$$

This estimator is $n^{-1/2}$ -consistent under Assumptions 1 and 2.

3.2 *Shape-Restricted Estimation and the Grid*

A shape restriction on g can be written $(Ag)(x) \le 0$ for every $x \in [0,1]$, where A is an operator. For example, if g is monotone non-increasing, then $Ag = dg/dx$. The shape restriction constitutes infinitely many constraints on g. We represent the shape restriction as a finite number of constraints by imposing it only at a grid of J equally spaced points $0 < x_1 < x_2 < ... < x_J < 1$. J increases as n increases. Because the grid points are equally spaced, the distance between two consecutive grid points is $\Delta = 1/(J+1)$. Let the notation $a_n \ll b_n$ mean that $a_n / b_n \to 0$ as $n \to \infty$. We assume that:

<u>Assumption 4</u>: (i) $J \ll n^{1/5}(\log n)^{-1/4}$. (ii) $h < 1/[2(J+1)]$. (iii) $J \to \infty$ as $n \to \infty$.

Assumptions 4(i) and 3(ii) ensure that Assumption 4(ii) holds for all sufficiently large n. Assumption 4(ii) requires $h < 1/[2(J+1)]$ for any n. Under this assumption, for each $i = 1,...,n$, there is only one j such that $X_i \in N_{x_j}$. Therefore, $\tilde{g}(x_j)$ and $\tilde{g}(x_k)$ are statistically independent if $j \ne k$. Assumption 4(iii) ensures that the distance between grid points decreases as n increases.

Let $g_j = g(x_j)$, and let $\boldsymbol{g}$ be the $J \times 1$ vector $\boldsymbol{g} = (g_1, ..., g_J)'$. With this notation, a shape restriction that is imposed only at grid points can be written

$$A_k(\boldsymbol{g}) \le 0; \; k = 1, ..., \kappa \le J ,$$

where the A_k's are functions. For example, the restriction that g is non-increasing, can be represented as $g_{j+1} - g_j \le 0$ for every $j = 1, ..., J-1$. Thus, $A_k(\boldsymbol{g}) = g_{k+1} - g_k$, and $\kappa = J - 1$. J and κ both increase as n increases. This dependence on n is not represented in the notation but is understood throughout this paper.

We impose shape restrictions on the grid by constraining differences between values of $g(x)$ at different values of x, not by constraining derivatives of g. This is because estimators of derivatives of g converge more slowly than the estimator of g. Consequently, the random sampling error of the constrained estimator of g is larger and the uniform confidence band for g wider if shape restrictions are imposed by constraining derivatives than if they are imposed by constraining differences.

Let $\mathcal{S} = \{k : A_k(\boldsymbol{g}) = 0\}$ denote the set of constraints that bind in the sampled population, and let $|\mathcal{S}|$ denote the number of elements in $\mathcal{S}$. Estimation of $\boldsymbol{g}$ subject to $A_k(\boldsymbol{g}) \le 0$ ($k = 1, ..., \kappa$) is asymptotically equivalent to estimating $\boldsymbol{g}$ subject to $A_k(\boldsymbol{g}) = 0$ for $k \in \mathcal{S}$. In other words, constraints that do not bind in the population can be dropped, and constraints that do bind can be replaced by equalities. In typical applications, the function $A_k(\boldsymbol{g})$ depends only on a few components of $\boldsymbol{g}$. For example, $A_k(\boldsymbol{g}) = g_{k+1} - g_k \le 0$ represents the restriction that g is non-increasing, and $A_k(\boldsymbol{g})$ depends on only two components of $\boldsymbol{g}$. The restriction that g is convex can be represented as $A_k(\boldsymbol{g}) = -g(x_k) + 2g(x_{k+1}) - g(x_{k+2}) \le 0$, and $A_k(\boldsymbol{g})$ depends on only three components of $\boldsymbol{g}$. If $|\mathcal{S}| < \kappa$ (not all constraints are binding), then there may be some components of $\boldsymbol{g}$ that do not affect the value of $A_k(\boldsymbol{g})$ for any $k \in \mathcal{S}$. That is, there may be some g_j's for which $\partial A_k(\boldsymbol{g}) / \partial g_j = 0$ for every $k \in \mathcal{S}$. These g_j's are unconstrained. They can be estimated and inference about them carried out using the unrestricted nonparametric method of Section (3.1). It is necessary to carry out constrained estimation and inference only for g_j's satisfying $\partial A_k(\boldsymbol{g}) / \partial g_j \ne 0$ for some $k \in \mathcal{S}$. These components affect the value of $A_k(\boldsymbol{g})$ for some $k \in \mathcal{S}$ and, therefore, are constrained. Call these g_j's "active components" of $\boldsymbol{g}$. Call the remaining components "inactive." Define

$$\mathcal{C} = \{ j : \; g_j \text{ is an active component of } \boldsymbol{g} \} ,$$

and let $|\mathcal{C}|$ denote the number of elements of $\mathcal{C}$. Let $\boldsymbol{g}^{(a)}$ denote the $|\mathcal{C}|\times 1$ vector of active components of $\boldsymbol{g}$.

We use the following notation to define the shape-restricted local quadratic estimator of $\boldsymbol{g}^{(a)}$. Index the components of $\boldsymbol{g}^{(a)}$ and the grid points corresponding to them by $\ell = 1,...,|\mathcal{C}|$. Define the $n\times 3$ matrix

$$\boldsymbol{X}^{(\ell)} = \begin{bmatrix} 1 & (X_1 - x_\ell) & (X_1 - x_\ell)^2 \\ & ... & \\ 1 & (X_n - x_\ell) & (X_n - x_\ell)^2 \end{bmatrix}; \ \ell = 1,...,|\mathcal{C}|.$$

Let $\boldsymbol{W}^{(\ell)}$ be the $n\times n$ diagonal matrix whose (i,i) component is $K_h(X_i - x_\ell)$, and let $\boldsymbol{Q}$ be the $3|\mathcal{C}|\times 3|\mathcal{C}|$ block diagonal matrix

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X}^{(1)\prime}\boldsymbol{W}^{(1)}\boldsymbol{X}^{(1)} & 0 & ... & 0 \\ 0 & \boldsymbol{X}^{(2)\prime}\boldsymbol{W}^{(2)}\boldsymbol{X}^{(2)} & ... & 0 \\ & ... & ... & \\ 0 & 0 & ... & \boldsymbol{X}^{(|\mathcal{C}|)\prime}\boldsymbol{W}^{(|\mathcal{C}|)}\boldsymbol{X}^{(|\mathcal{C}|)} \end{bmatrix}.$$

Finally, define the $3|\mathcal{C}|\times 1$ vector

$$\boldsymbol{d} = [(\boldsymbol{X}^{(1)\prime}\boldsymbol{W}^{(1)}\boldsymbol{Y})' \quad ... \quad (\boldsymbol{X}^{(|\mathcal{C}|)\prime}\boldsymbol{W}^{(|\mathcal{C}|)}\boldsymbol{Y})']',$$

and let $\boldsymbol{b}$ the $3|\mathcal{C}|\times 1$ vector $(b_{11}, b_{21}, b_{31}, ..., b_{1|\mathcal{C}|}, b_{2|\mathcal{C}|}, b_{3|\mathcal{C}|})'$.

If $\mathcal{S}$ were known, the shape restricted local quadratic estimator of g_ℓ ($\ell \in \mathcal{C}$) would be $\tilde{b}_{1\ell}$, where $\tilde{b}_{1\ell}$ is the $(3\ell - 2)$ component of the $3|\mathcal{C}|\times 1$ vector

$$\tilde{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} (0.5\boldsymbol{b}'\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{d}'\boldsymbol{b}).$$

subject to:

$$A_k(\boldsymbol{g}) = 0; \quad k \in \mathcal{S}.$$

However, $\mathcal{S}$ is unknown in applications. Therefore, we replace it with the estimate $\hat{\mathcal{S}}$ that is described in Section 4. Redefine the active components of $\boldsymbol{g}$ as the g_j 's satisfying $\partial A_k(\boldsymbol{g})/\partial g_j \neq 0$ for some $k \in \hat{\mathcal{S}}$. In the definitions of $\boldsymbol{Q}$ and $\boldsymbol{d}$, replace $\mathcal{C}$ with

$$\hat{\mathcal{C}} = \{j: \ g_j \text{ is a redefined active component of } \boldsymbol{g}\}$$

and replace $|\mathcal{C}|$ with $|\hat{\mathcal{C}}|$, which is the number of elements of $\hat{\mathcal{C}}$. We estimate an active component g_ℓ by $\hat{g}_\ell = \hat{b}_{1\ell}$, which is the $(3\ell - 2)$ component of the vector

(3.3) $$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} (0.5\boldsymbol{b}'\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{d}'\boldsymbol{b}).$$

subject to:

$$A_k(\boldsymbol{g}) = 0; \quad k \in \hat{\mathcal{S}}.$$

The estimator of the vector of redefined active components $\boldsymbol{g}^{(a)}$ is $\hat{\boldsymbol{g}}^{(a)} = (\hat{b}_{11}, \hat{b}_{14}, .., \hat{b}_{3|\mathcal{C}|-2})'$.

In summary, we estimate the active components of $\boldsymbol{g}$ by solving (3.3) and the remaining components by the unrestricted method of Section (3.1). Denote the resulting estimator of $\boldsymbol{g}$ by $\hat{\hat{\boldsymbol{g}}}$. Section 4 obtains the asymptotic distribution of $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ and a uniform confidence band for $\boldsymbol{g}$.[1]

4. THE SET OF POSSIBLY BINDING CONSTRAINTS

This section explains how to find the set $\hat{\mathcal{S}}$ of possibly binding constraints. Define the $J \times 1$ vectors $\boldsymbol{g}^{(J)} = [g(x_1), ..., g(x_J)]'$ and $\tilde{\boldsymbol{g}}^{(J)} = [\tilde{g}(x_1), ..., \tilde{g}(x_J)]'$. For each $k = 1, ..., \kappa$, define the set

$$\mathcal{C}(k) = \{ j : \partial A_k(\boldsymbol{g}) / \partial g_j = 0 \text{ for all } \boldsymbol{g} \}.$$

Let $|\mathcal{C}(k)|$ denote the number of components of $\mathcal{C}(k)$, and define $J_0 = \max_{k=1,...,\kappa} |\mathcal{C}(k)|$. Let $\mathcal{M}$ be the set of constraints for which $0 < A_k(\boldsymbol{g}^{(J)}) \le [(\log n)/(nh)]^{1/2}$, and let $|\mathcal{M}|$ be the number of constraints in $\mathcal{M}$. Make the following assumption:

<u>Assumption 5</u>: (i) There is a finite constant q not depending on n such that $|\mathcal{C}(k)| \le q$ for all $k = 1, ..., \kappa$. (ii) The vector $\boldsymbol{g}^{(J)}$ is contained in a compact subset $\mathcal{G}$ of $\mathbb{R}^J$. (iii) A_k is a twice continuously differentiable function of its arguments for each $k = 1, ..., \kappa$. There is a constant $\mathcal{A} < \infty$ such that $|\partial A_k^2(\boldsymbol{g}^{(J)}) / \partial g_{j\ell}| \le \mathcal{A}$ for all $\boldsymbol{g}^{(J)} \in \mathcal{G}$; $j, \ell = 1, ..., J$; and $k = 1, ..., \kappa$. (iv) $|\mathcal{M}| \to 0$ as $n \to \infty$.

Assumption 5(i) is motivated by the observation that with typical shape restrictions, such as monotonicity, convexity, or the Slutsky inequality, $A_k(\boldsymbol{g})$ depends on only a few components of $\boldsymbol{g}$. Assumption 5(iv) holds for typical shape restrictions if Assumption 4 holds. Examples illustrating this are given in the appendix.

Under Assumption 5(iii), a Taylor series expansion gives

---

[1] If n is small, random sampling errors may cause the unrestricted estimator of the inactive components of $\boldsymbol{g}$ to violate the constraints $A_k(\boldsymbol{g}) \le 0$. This problem can be avoided by imposing the constraints on all components of $\boldsymbol{g}$. The problem does not arise if n is large and the assumptions of this paper are satisfied. Therefore, it does not affect the asymptotic distribution of $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ or the asymptotic uniform confidence band for $g(x)$.

$$(4.1)\qquad A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) = \left[\frac{\partial A_k(\boldsymbol{g}^{(J)})}{\partial \boldsymbol{g}}\right]'(\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)}) + O_p\left(\left\|\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)}\right\|^2\right)$$

$$= \left[\frac{\partial A_k(\boldsymbol{g}^{(J)})}{\partial \boldsymbol{g}}\right]'(\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)}) + O_p[J/(nh)].$$

The components of $(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)})$ are independently normally distributed with means of 0 by (3.2). If J were fixed, asymptotic multivariate normality of $(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)})$ and of the $\kappa \times 1$ vector whose k'th component is $(nh)^{1/2}[A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)})]$ would follow from (3.2) and (4.1). The following theorem shows that this result holds even if $J \to \infty$ as $n \to \infty$.

Theorem 4.1: Let Assumptions 1-4 and 5(i)-5(iii) hold. Define the $J \times J$ diagonal matrix $\omega = diag[\sigma^2_{\tilde{g}(x_j)}:\ j = 1,...,J]$. Let Υ be the $\kappa \times \kappa$ matrix whose (k,ℓ) component is

$$\Upsilon_{k\ell} = \left[\frac{\partial A_k(\boldsymbol{g}^{(J)})}{\partial \boldsymbol{g}}\right]' \omega \left[\frac{\partial A_\ell(\boldsymbol{g}^{(J)})}{\partial \boldsymbol{g}}\right].$$

Define the random variables $\boldsymbol{z} \sim N(0,\omega)$ and $\tilde{\boldsymbol{z}} \sim N(0,\Upsilon)$. Then

$$(4.2)\qquad \lim_{n\to\infty} \sup_{\boldsymbol{t}} | P[(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(J)} - \boldsymbol{g}^{(J)}) \le \boldsymbol{t}] - P(\boldsymbol{z} \le \boldsymbol{t}) | = 0$$

and

$$(4.3)\qquad \lim_{n\to\infty} \sup_{\boldsymbol{t}} | P\{(nh)^{1/2}[A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)})] \le \boldsymbol{t}\} - P(\tilde{\boldsymbol{z}} \le \boldsymbol{t}) | = 0. \ \blacksquare$$

Now let $c_n = (\log n)^{1/2}$. Then $1 - \Phi(c_n) = O[(n \log n)^{-1/2}]$, where Φ is the standard normal distribution function. It follows from (4.3) that asymptotically,

$$P\left[\Upsilon_{kk}^{-1/2}(nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) | > c_n\right] = O[(n\log n)^{-1/2}]$$

and

$$(4.4)\qquad P\left[\Upsilon_{kk}^{-1/2}(nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) | > c_n \text{ for any } k = 1,...,\kappa\right] = O[\kappa(n\log n)^{-1/2}] = o(1).$$

If $k \in \mathcal{S}$, then $A_k(\boldsymbol{g}^{(J)}) = 0$. Therefore,

$$(4.5)\qquad P[| A_k(\tilde{\boldsymbol{g}}^{(J)}) | \le \Upsilon_{kk}^{1/2}(nh)^{-1/2} c_n \text{ for every } k \in \mathcal{S}] = 1 - O[\kappa(n\log n)^{-1/2}] \to 1.$$

If $k \notin \mathcal{S}$, $| A_k(\boldsymbol{g}^{(J)}) | > [(\log n)/(nh)^{1/2}]$, and $\Upsilon_{kk}^{-1/2}(nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) | \le c_n$, then for all sufficiently large n,

$$\Upsilon_{kk}^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) | \geq \Upsilon_{kk}^{1/2} | A_k(\boldsymbol{g}^{(J)}) | - \Upsilon_{kk}^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) |$$

$$\geq \Upsilon_{kk}^{1/2} | A_k(\boldsymbol{g}^{(J)}) | - (nh)^{-1/2} c_n > (nh)^{-1/2} c_n .$$

But $|\mathcal{M}|$ is an integer, and $|\mathcal{M}| \to 0$ as $n \to \infty$ by Assumption 5(iv). Therefore, $\Upsilon_{kk}^{1/2} (nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) | > c_n$ for every $k \notin \mathcal{S}$ if n is sufficiently large and $\Upsilon_{kk}^{-1/2} (nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) - A_k(\boldsymbol{g}^{(J)}) | \leq c_n$ for every $k \leq \kappa$. By (4.4), the probability of the latter event is $1 - o(1)$. Therefore,

(4.6) $$P\left[\Upsilon_{kk}^{-1/2} (nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) | \leq c_n \text{ for any } k \notin \mathcal{S} \right] = o(1) .$$

Define

$$\tilde{\mathcal{S}} = \left\{ k : \Upsilon_{kk}^{-1/2} (nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) | \leq c_n \right\} .$$

Then it follows from (4.5) and (4.6) that $P(\tilde{\mathcal{S}} = \mathcal{S}) \to 1$ as $n \to \infty$.

This result continues to hold if Υ_{kk} is replaced by the consistent estimator obtained by replacing $\sigma^2_{\tilde{g}(x_j)}$ with $\hat{\sigma}^2_{\tilde{g}(x_j)}$ in ω and $\boldsymbol{g}^{(J)}$ with $\tilde{\boldsymbol{g}}^{(J)}$ in $\Upsilon_{k\ell}$. Define

(4.7) $$\hat{\mathcal{S}} = \left\{ k : \hat{\Upsilon}_{kk}^{-1/2} (nh)^{1/2} | A_k(\tilde{\boldsymbol{g}}^{(J)}) | \leq c_n \right\} .$$

Then $P(\hat{\mathcal{S}} = \mathcal{S}) \to 1$ as $n \to \infty$. $\hat{\mathcal{S}}$ can be calculated from the data and is the desired set of possibly binding constraints.

5. ASYMPTOTIC DISTRIBUTION OF THE CONSTRAINED ESTIMATOR AND UNIFORM CONFIDENCE BAND

This section shows that $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ is asymptotically multivariate normally distributed with mean 0. This result is used to obtain an asymptotic uniform confidence band for $g(x)$. Because $P(\hat{\mathcal{S}} = \mathcal{S}) \to 1$ as $n \to \infty$, the asymptotic distribution of $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ is the same regardless of whether $\mathcal{S}$ is estimated by $\hat{\mathcal{S}}$ or known. Therefore, it suffices to derive the asymptotic distribution and confidence band under the assumption that $\mathcal{S}$ is known. Accordingly, it is assumed throughout this section that $\mathcal{S}$ is known. Section 5.1 treats the case of linear constraint functions A_k. Section 5.2 extends the results of section 5.1 to nonlinear constraints.

5.1 *Linear Constraints*

If the functions A_k are linear, the inequalities $A_k(\boldsymbol{g}) \le 0$ ($k = 1,...,\kappa$) can be written as $\boldsymbol{Ag} \le \boldsymbol{r}$, where $\boldsymbol{A}$ is a $\kappa \times J$ matrix and $\boldsymbol{r}$ is a $\kappa \times 1$ vector. For example, if g is monotone non-increasing, then $A_k(\boldsymbol{g}) = g(x_{k+1}) - g(x_k)$, $\boldsymbol{r} = 0$, and $\boldsymbol{A}$ is the $(J-1) \times J$ matrix

$$\boldsymbol{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ & & & \ldots & & & \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}.$$

It follows from (3.2) and the Assumption 4 that the estimators of the inactive components of $(nh)^{1/2}(\hat{\boldsymbol{g}} - \boldsymbol{g})$ are asymptotically normally distributed with means of 0 independently of each other and of the active components. Therefore, asymptotic distribution of $(nh)^{1/2}(\hat{\boldsymbol{g}} - \boldsymbol{g})$ will be known after the asymptotic distribution of its of active components has been obtained.

Let $\breve{\kappa} \le \kappa$ the number of rows of $\boldsymbol{A}$ corresponding to constraints affecting active components of $\boldsymbol{g}$. Let $\boldsymbol{A}^{(a)}$ be the $\breve{\kappa} \times |\mathcal{C}|$ matrix consisting of the columns of $\boldsymbol{A}$ corresponding to active components of $\boldsymbol{g}$. The constraints on these components can be written as $\boldsymbol{A}^{(a)}\boldsymbol{g}^{(a)} = \boldsymbol{r}^{(a)}$, where $\boldsymbol{r}^{(a)}$ is a $\breve{\kappa} \times 1$ vector. Let $\breve{\boldsymbol{A}}$ denote the $\kappa \times 3|\mathcal{C}|$ matrix in which column $3\ell - 2$ ($\ell = 1,...,|\mathcal{C}|$) is column ℓ of $\boldsymbol{A}^{(a)}$ and the remaining columns are all zeros. Because it can be assumed as $n \to \infty$ that $\mathcal{S}$ is known, problem (3.3) with linear constraints can be rewritten as

(5.1) $$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}}(0.5\boldsymbol{b}'\boldsymbol{Qb} - \boldsymbol{d}'\boldsymbol{b})$$

subject to:

$$\breve{\boldsymbol{A}}\boldsymbol{b} = \boldsymbol{r}^{(a)}.$$

Note that $\breve{\boldsymbol{A}}$ and $\boldsymbol{r}$ are non-stochastic. Problem (5.1) can be solved analytically using the method of Lagrangian multipliers. The solution is the well-known constrained least squares estimator

$$\hat{\boldsymbol{b}} = \tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}'(\breve{\boldsymbol{A}}\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}')^{+}(\breve{\boldsymbol{A}}\tilde{\boldsymbol{b}} - \boldsymbol{r}^{(a)}),$$

where $\tilde{\boldsymbol{b}}^{(a)}$ is the subvector of the unconstrained local quadratic estimator of Section (3.1) corresponding to active components of $\boldsymbol{g}$ and the superscript + denotes the Moore-Penrose generalized inverse. Only columns $1, 4, \ldots, 3|\mathcal{C}|-2$ of the $\kappa \times 3|\mathcal{C}|$ matrix $\breve{\boldsymbol{A}}$ are non-zero, and components $1, 4, \ldots, 3|\mathcal{C}|-2$ of the $3|\mathcal{C}| \times 1$ vector $\tilde{\boldsymbol{b}}$ correspond to components of $\tilde{\boldsymbol{g}}^{(a)}$, which is the local quadratic estimator of $\boldsymbol{g}^{(a)}$. The submatrix of $\breve{\boldsymbol{A}}$ consisting of columns $1, 4, \ldots, 3|\mathcal{C}|-2$ is $\boldsymbol{A}^{(a)}$. Therefore, $\breve{\boldsymbol{A}}\tilde{\boldsymbol{b}} = \boldsymbol{A}^{(a)}\tilde{\boldsymbol{g}}^{(a)}$,

$$\breve{A}\tilde{\boldsymbol{b}} - \boldsymbol{r}^{(a)} = \boldsymbol{A}^{(a)}\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{r}^{(a)}$$

$$= \boldsymbol{A}^{(a)}(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)}) + \boldsymbol{A}^{(a)}(E\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + (\boldsymbol{A}^{(a)}\boldsymbol{g}^{(a)} - \boldsymbol{r}^{(a)}),$$

and

$$\hat{\boldsymbol{b}} = \tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}'(\breve{\boldsymbol{A}}\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}')^{+}[\boldsymbol{A}^{(a)}(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)}) + \boldsymbol{A}^{(a)}(E\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + (\boldsymbol{A}^{(a)}\boldsymbol{g}^{(a)} - \boldsymbol{r}^{(a)})].$$

Under Assumption 1(iii),

$$\boldsymbol{A}^{(a)}\boldsymbol{g}^{(a)} - \boldsymbol{r}^{(a)} = 0$$

and

$$\boldsymbol{A}^{(a)}(E\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) = O(h^4) = o[(nh)^{-1/2}]$$

for all J. Therefore,

$$\hat{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} = (\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) - [\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}'(\breve{\boldsymbol{A}}\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}')^{+}\boldsymbol{A}^{(a)}]^{(a)}(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)}) + o[(nh)^{-1/2}]$$

$$= \left\{I_{|\mathcal{C}|\times|\mathcal{C}|} - [\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}'(\breve{\boldsymbol{A}}\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}')^{+}\boldsymbol{A}^{(a)}]^{(a)}\right\}(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)}) + o[(nh)^{-1/2}]$$

for all J, where $[\cdot]^{(a)}$ denotes the $|\mathcal{C}|\times|\mathcal{C}|$ submatrix of the $|3\mathcal{C}|\times|3\mathcal{C}|$ matrix $[\cdot]$ consisting of rows and columns 1, 4,…, $3|\mathcal{C}|-2$. Define

$$\Xi = \left\{I_{|\mathcal{C}|\times|\mathcal{C}|} - [\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}'(\breve{\boldsymbol{A}}\boldsymbol{Q}^{-1}\breve{\boldsymbol{A}}')^{+}\boldsymbol{A}^{(a)}]^{(a)}\right\}.$$

Then

$$\hat{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} = \Xi(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)}) + o[(nh)^{-1/2}] \tag{5.2}$$

for all J.

Now rearrange the components of $\boldsymbol{g}$ and $\hat{\hat{\boldsymbol{g}}}$ into the vectors $[\boldsymbol{g}^{(-a)}, \boldsymbol{g}^{(a)}]'$ and $[\hat{\hat{\boldsymbol{g}}}^{(-a)}, \hat{\boldsymbol{g}}^{(a)}]'$, respectively, where the superscript $(-a)$ denotes inactive components of $\boldsymbol{g}$ and $\hat{\hat{\boldsymbol{g}}}$. It follows from Theorem 4.1 and asymptotic negligibility of $(nh)^{1/2}(E\tilde{\boldsymbol{g}}^{(a)} - \tilde{\boldsymbol{g}}^{(a)})$ that $(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(a)} - E\tilde{\boldsymbol{g}}^{(a)})$ are independently asymptotically multivariate normally distributed with diagonal covariance matrices. Therefore, (5.2) implies that $(nh)^{1/2}(\hat{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)})$ and $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ are linear combinations of asymptotic multivariate normals and are, themselves, asymptotically multivariate normal. Consequently, we have the following theorem.

Theorem 5.1: Let Assumptions 1-5 hold. Let Σ_n be the $J \times J$ matrix

$$\Sigma_n = \begin{bmatrix} \Sigma_n^{(-a)} & 0 \\ 0 & \Sigma_n^{(a)} \end{bmatrix},$$

$$\Sigma_n^{(-a)} = diag[\sigma^2_{\tilde{g}(x_j)} :\ g(x_j) \text{ is inactive}],$$

$$\Sigma_n^{(a)} = \Xi\,\Omega\,\Xi',$$

and

$$\Omega = diag[\sigma^2_{\tilde{g}(x_j)} :\ g(x_j) \text{ is active}].$$

Let $\boldsymbol{z}$ be a random vector that is distributed as $N(0,\Sigma_n)$. If the constraints on g are linear, then

(5.3) $$\lim_{n\to\infty} \sup_{\boldsymbol{t}} |P[(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g}) \le \boldsymbol{t}] - P(\boldsymbol{z} \le \boldsymbol{t})| = 0. \ \blacksquare$$

Result (5.3) continues to hold if $\sigma^2_{\tilde{g}(x_j)}$ is replaced in Σ_n with the estimator $\hat{\sigma}^2_{\tilde{g}(x_j)}$ described in Section 3.1. Denote the resulting matrix by $\hat{\Sigma}_n$. Then it follows from (5.3) that a two-sided, asymptotic $1-\alpha$ confidence interval for $g(x_j)$ is

(5.4) $$\hat{\hat{g}}_j - (nh)^{-1/2}\hat{\Sigma}_{n,jj} z_{1-\alpha/2} \le g_j \le \hat{\hat{g}}_j + (nh)^{-1/2}\hat{\Sigma}_{n,jj} z_{1-\alpha/2}$$

where $\hat{\hat{g}}_j$ is the j'th component of $\hat{\hat{\boldsymbol{g}}}$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. It follows from the properties of constrained least squares estimators that $\Sigma_n - \mathrm{cov}(\tilde{\boldsymbol{g}}^{(J)})$ is positive semidefinite. Therefore, the confidence interval (5.4) is no wider and may be narrower than a confidence interval for g_j based on the unconstrained estimator $\tilde{g}_j$.

An asymptotic $1-\alpha$ uniform confidence band for $g(x_j)$ ($j=1,\ldots,J$) is

$$\mathcal{I}_{1-\alpha} = \{g_1,\ldots,g_J :\ -\gamma_{j2} \le \hat{\hat{g}}_j - g_j \le \gamma_{j1};\ \ j=1,\ldots,J\},$$

where γ_{j1} and γ_{j2} are critical values. The coverage probability is

$$P\left\{\bigcap_{j=1}^{J}[-\gamma_{j2} \le \hat{\hat{g}}_j - g_j \le \gamma_{j1}]\right\}$$

or, equivalently,

(5.5) $$P\left\{\bigcap_{j=1}^{J}[\hat{\hat{g}}_j - \gamma_{j1} \le g_j \le \hat{\hat{g}}_j + \gamma_{j2}]\right\}.$$

An asymptotic coverage probability of $1-\alpha$ can be obtained by choosing the γ_{j1}'s and γ_{j2}'s so that

(5.6) $$P\left\{\bigcap_{j=1}^{J}\{-\gamma_{j2} \le Z_j \le \gamma_{j1}\}\right\} = 1-\alpha ,$$

where $Z = (Z_1,...,Z_J)'$ is a random variable with the $N[0,(nh)^{-1}\hat{\Sigma}_n]$ distribution. For a symmetrical confidence band, $\gamma_{j1} = \gamma_{j2} = \gamma_j$, where $\gamma_j > 0$, and

(5.7) $$P\left\{\bigcap_{j=1}^{J}| Z_j | \le \gamma_j\}\right\} = 1-\alpha .$$

The probability in (5.7) can be estimated by Monte Carlo for any γ_j 's by drawing random samples from the $N[0,(nh)^{-1}\hat{\Sigma}_n]$ distribution.

Except in special cases, the boundaries of the confidence band (5.5) do not satisfy the shape restrictions that are assumed to hold for $\boldsymbol{g}$. To obtain a confidence band whose boundaries satisfy the shape restrictions, define $\boldsymbol{e}$ to be a $J \times 1$ vector of 1's, $\boldsymbol{\gamma}_1 = (\gamma_{11},...,\gamma_{J1})'$, and $\boldsymbol{\gamma}_2 = (\gamma_{12},...,\gamma_{J2})'$. The uniform confidence band that has the minimum average width among bands satisfying the shape restrictions can be obtained by solving the nonlinear programming problem

(5.8) $$(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)' = \arg\min_{\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2} \boldsymbol{e}'(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1)$$

subject to

(5.9) $$P\left\{\bigcap_{j=1}^{J}\{-\zeta_{j2} \le Z_j \le \zeta_{j1}\}\right\} = 1-\alpha$$

$$\boldsymbol{A}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{\zeta}_1) \le 0$$

$$\boldsymbol{A}(\hat{\hat{\boldsymbol{g}}} + \boldsymbol{\zeta}_2) \le 0.$$

The resulting confidence band is $\hat{\boldsymbol{g}} - \boldsymbol{\gamma}_1 \le \boldsymbol{g} \le \hat{\boldsymbol{g}} + \boldsymbol{\gamma}_2$. The minimum width symmetrical confidence band that satisfies the shape restrictions can be obtained by setting $\boldsymbol{\gamma}_2 = -\boldsymbol{\gamma}_1$ in problem (5.7).

Solving problem (5.8) is difficult computationally because of the nonlinear constraint (5.9). The nonlinear constraint can be removed and computation simplified at the cost of a wider confidence band. To obtain this band, let $\tilde{\boldsymbol{\gamma}}_1$ and $\tilde{\boldsymbol{\gamma}}_2$ be the vectors of confidence limits obtained in (5.5)-(5.6). A confidence band whose boundaries satisfy the shape restrictions but whose average width may exceed the average width of the band obtained from (5.8) is $\hat{\hat{\boldsymbol{g}}} - \hat{\boldsymbol{\gamma}}_1 \le \boldsymbol{g} \le \hat{\hat{\boldsymbol{g}}} + \hat{\boldsymbol{\gamma}}_2$, where

(5.10) $$(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2)' = \arg\min_{\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2} \boldsymbol{e}'(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1)$$

subject to

$$\zeta_1 \le \tilde{\gamma}_1 \le \tilde{\gamma}_2 \le \zeta_2$$

$$\boldsymbol{A}(\hat{\hat{\boldsymbol{g}}} - \zeta_1) \le 0$$

$$\boldsymbol{A}(\hat{\hat{\boldsymbol{g}}} + \zeta_2) \le 0.$$

The grid points $\{x_j\}$ become dense in $[0,1]$ as $n \to \infty$. Therefore, $\{\gamma_j : j = 1,..., J\}$ and (5.6)-(5.10) provide an asymptotic uniform confidence region for $g(x)$.

5.2 *Nonlinear Constraints*

This section explains how to obtain an asymptotic uniform confidence band for g when one or more of the functions $A_k(\boldsymbol{g})$ specifying the shape constraints is nonlinear. As was explained in the introduction to Section 4, we assume that $\hat{\mathcal{S}} = \mathcal{S}$ in deriving asymptotic uniform confidence band.

As in Section 3.1, let $\tilde{\boldsymbol{g}}$ denote the unconstrained local quadratic estimator of $\boldsymbol{g}$. For $k \in \mathcal{S}$, define the scalar η_k by $A_k(\tilde{\boldsymbol{g}}) = \eta_k$. Let $\boldsymbol{b}^{(1)}$ denote the intercept components of $\boldsymbol{b}$ (that is, components $1, 4, \ldots, 3|\mathcal{C}|-2$). Then unrestricted estimates of the active components of $\boldsymbol{g}$ can be obtained by solving

(5.11) $$\tilde{\boldsymbol{b}}^{(a)} = \arg\min_{\boldsymbol{b}} (0.5\boldsymbol{b}'\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{d}'\boldsymbol{b})$$

subject to:

$$A_k(\boldsymbol{b}^{(1)}) = \eta_k \quad (k \in \mathcal{S}).$$

The constrained estimates are obtained by solving

(5.12) $$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} (0.5\boldsymbol{b}'\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{d}'\boldsymbol{b})$$

subject to:

$$A_k(\boldsymbol{b}^{(1)}) = 0 \quad (k \in \mathcal{S}).$$

But $k \in \mathcal{S}$ implies that $\eta_k \to 0$ as $n \to \infty$. Therefore, finding $\hat{\boldsymbol{g}}$ is equivalent to finding the effect of a small change in the η_k's on the optimal solution to (5.11). This can be done by using the theory of sensitivity analysis in nonlinear programming (Fiacco 1983).

To state the result of the sensitivity analysis, modify the definition of $A_k(\cdot)$ to include all components of $\boldsymbol{b}$ in its arguments. Let $\boldsymbol{A}(\boldsymbol{b})$ be the $|\mathcal{S}|\times 1$ vector whose k'th component is $A_k(\boldsymbol{b})$. Let $\boldsymbol{A}_{\boldsymbol{b}}$ be the $|\mathcal{S}|\times 3|\mathcal{C}|$ matrix whose (k, j) component is $\partial A_k(\boldsymbol{b}_0)/\partial \boldsymbol{b}_j$, where $\boldsymbol{b}_0$ is the $3|\mathcal{C}|\times 1$ vector

$$\boldsymbol{b}_0 = \{[g(x_j), g'(x_j), g''(x_j)]' : j \in \mathcal{C}\}.$$

Let $[\cdot]^{(a)}$ denote the $|\mathcal{C}|\times|\mathcal{C}|$ submatrix of the $|3\mathcal{C}|\times|3\mathcal{C}|$ matrix $[\cdot]$ consisting of rows and columns 1, $4,\ldots,3|\mathcal{C}|-2$. Define

$$\bar{\Xi} = \left\{ I_{|\mathcal{C}|\times|\mathcal{C}|} - [\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}]^{(a)} \right\}.$$

The result is given by the following theorem.

Theorem 5.2: Let Assumptions 1-5 hold. Define $\bar{\Sigma}_n$ as the $J\times J$ matrix whose (j,k) element is

$$\bar{\Sigma}_{n,jk} = \begin{bmatrix} \bar{\Sigma}_n^{(-a)} & 0 \\ 0 & \bar{\Sigma}_n^{(a)} \end{bmatrix},$$

$$\bar{\Sigma}_n^{(-a)} = diag[\sigma^2_{\tilde{g}(x_j)} :\ g(x_j) \text{ is inactive}],$$

$$\Sigma_n^{(a)} = \bar{\Xi}\,\Omega\,\bar{\Xi}',$$

and

$$\Omega = diag[\sigma^2_{\tilde{g}(x_j)} :\ g(x_j) \text{ is active}] \quad .$$

Let $\boldsymbol{z}$ be a random vector that is distributed as $N(0,\bar{\Sigma}_n)$. Then

(5.13) $$\lim_{n\to\infty}\sup_{\boldsymbol{t}} |P[(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}}-\boldsymbol{g})\le \boldsymbol{t}] - P(\boldsymbol{z}\le \boldsymbol{t})| = 0. \ \blacksquare$$

An asymptotic uniform confidence region for $\boldsymbol{g}$ and $g(x)$ can now be constructed as in Section 5.1 by replacing (5.3) with (5.13).

5.3 *Uniformity*

The asymptotic distributional results and confidence bands in Sections 5.1 and 5.2 assume that the constraints are fixed as $n\to\infty$. That is, if $A_k(\boldsymbol{g}) = c_k$ for some $k\le\kappa$ and $c_k<0$, then c_k remains constant as $n\to\infty$. The asymptotic results do not hold uniformly over c_k in a neighborhood of 0. For each $k=1,\ldots,\kappa$ there are sequences $\{c_{nk}\}$ such that $A_k(\boldsymbol{g}) = c_{nk} < 0$, $c_{nk}\to 0$ as $n\to\infty$, and (5.3) and (5.13) do not hold. This problem can be overcome by replacing the constraints $A_k(\boldsymbol{g})\le 0$ with the slightly relaxed versions $A_k(\boldsymbol{g})\le\delta_n$, where $\delta_n = (nh)^{-1/2}\log n$. Then with probability approaching 1 as $n\to\infty$, no relaxed constraint is binding. The asymptotic distribution of $(nh)^{1/2}(\hat{\boldsymbol{g}}-\boldsymbol{g})$ with the relaxed constraints is the same as that of the unconstrained estimator, and there is no issue of uniformity. In finite samples, however, one or more of the relaxed constraints may be binding. Therefore, the finite-sample variability of the constrained estimator $(nh)^{1/2}(\hat{\boldsymbol{g}}-\boldsymbol{g})$ may be less than that of the unconstrained estimator. The results of Blundell, Horowitz, and Parey (2012) illustrate this reduction in finite-sample variability in an empirical setting.

6. MULTIVARIATE EXTENSION

This section outlines the extension of the results of Sections 3-5 to a two-dimensional explanatory vector X, such as price and income in a demand function. Extensions to higher dimensions are possible but are less useful for economics and likely to yield low estimation precision because of the curse of dimensionality. The extension to a two-dimensional X involves mainly notational adjustments to the results of Sections 3-5. Therefore, the results of the extension are presented without proofs.

Denote the data by $\{Y_i, X_{i1}, X_{i2} : i = 1,....,n\}$. Assume that $\text{supp}(X_{i1}, X_{i2}) = [0,1]^2$ for all $i = 1,...,n$. The following notation is used to state the unrestricted estimator of $g(x_1, x_2)$. Let K denote a probability density function that is supported on $[-1,1]^2$ and whose odd moments are all zero. For any $v \equiv (v_1, v_2) \in [-1,1]^2$ and bandwidths h_1 and h_2, let

(6.1) $$K_h(v) = K(v_1 / h_1, v_2 / h_2).$$

For any $x_1, x_2 \in [0,1]^2$, let N_x denote the rectangle $\{\xi, \zeta : |\xi - x_1| \le h_1,\ |\zeta - x_2| \le h_2\}$. Define

$$n_x = \sum_{i=1}^{n} I(|X_{i1} - x_1| \le h_1) I(|X_{i2} - x_2| \le h_2).$$

For $i = i_1, i_2, ..., i_{n_x}$ and $(X_{i_1 1}, X_{i_1 2}), ..., (X_{i_{n_x} 1}, X_{i_{n_x} 2}) \in N_x$, define the 1×6 vector

$$\boldsymbol{X}_i^{(x)} = \left[1\ \ (X_{i1} - x_1)\ \ (X_{i2} - x_2)\ \ (X_{i1} - x_1)^2\ \ (X_{i2} - x_2)^2\ \ (X_{i1} - x_1)(X_{i2} - x_2)\right]$$

and the $n_x \times 6$ matrix

$$\boldsymbol{X}^{(x)} = \begin{pmatrix} \boldsymbol{X}_{i_1}^{(x)} \\ \ldots \\ \boldsymbol{X}_{i_{n_x}}^{(x)} \end{pmatrix}.$$

For $X_i = (X_{i1}, X_{i2})$ and $x = (x_1, x_2)$, define the $n_x \times n_x$ diagonal matrix

$$\boldsymbol{W}^{(x)} = diag[K_h(X_i - x)]:\ X_i \in N_x]$$

and the 6×6 matrix

$$\boldsymbol{S}_n^{(x)} = n_x^{-1} \boldsymbol{X}^{(x)\prime} \boldsymbol{W}^{(x)} \boldsymbol{X}^{(x)}.$$

For $X_{i_1}, ..., X_{i_{n_x}} \in N_x$, define the $n_x \times 1$ vector $\boldsymbol{Y}^{(x)} = (Y_{i_1}, ..., Y_{i_{n_x}})'$ and $\boldsymbol{g}^{(x)} = [g(X_{i_1}), ..., g(X_{i_{n_x}})]'$.

Now let $\boldsymbol{b} = (b_1, ..., b_6)'$ be a 6×1 vector, and let

$$\tilde{\boldsymbol{b}}(x) \equiv [\tilde{b}_1(x), ..., \tilde{b}_6(x)]' = \arg\min_{\boldsymbol{b}} (\boldsymbol{Y}^{(x)} - \boldsymbol{X}^{(x)}\boldsymbol{b})'\boldsymbol{W}^{(x)}(\boldsymbol{Y}^{(x)} - \boldsymbol{X}^{(x)}\boldsymbol{b})$$

$$= \arg\min_{\boldsymbol{b}} (\boldsymbol{b}'\boldsymbol{X}^{(x)\prime}\boldsymbol{W}^{(x)}\boldsymbol{X}^{(x)}\boldsymbol{b} - 2\boldsymbol{Y}^{(x)\prime}\boldsymbol{W}^{(x)}\boldsymbol{X}^{(x)}\boldsymbol{b}).$$

The unrestricted estimator of $g(x)$ is $\tilde{g}(x) = \tilde{b}_1(x)$. Standard algebra of least squares estimation shows that

$$\tilde{\boldsymbol{b}}(x) = (\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime}\boldsymbol{W}^{(x)}\boldsymbol{Y}^{(x)}$$

Therefore,

$$\tilde{g}(x) - g(x) = e_1'(\boldsymbol{S}_n^{(x)})^{-1} n_x^{-1} \boldsymbol{X}^{(x)\prime}\boldsymbol{W}^{(x)}\boldsymbol{Y}^{(x)} - g(x),$$

where $e_1 = (1,0,0,0,0,0)'$.

Now make the following assumptions, which are modifications of Assumptions 1-3:

Assumption 1′: $\{Y_i, X_i : i = 1,...,n\}$ is an independent random sample from the distribution of (Y, X), where (i) $\text{supp}(X) = [0,1]^2$. (ii) $Y = g(X) + U$, $E(U \mid X = x) = 0$ and $E(U^2 \mid X = x) = \sigma_U^2$ (a finite constant) for every $x \in [0,1]^2$. (iii) g satisfies the assumed shape restriction. (iv) $E(|U|^3) < \infty$.

Assumption 2′: (i) $g(x)$ is four times continuously differentiable with respect to any combination of the components of $x \in [0,1]^2$. (ii) The distribution of X has a probability density function with respect to Lebesgue measure, f_X, that is continuously differentiable everywhere in $[0,1]^2$. (iii) $f_X(x) \ge \delta$ for some $\delta > 0$ and every $x \in [0,1]^2$.

Assumption 3′: (i) K is a bounded probability density function that is supported on $[-1,1]^2$, and all odd moments of K are zero; (ii) $h_j = c_j n^{-1/6}$ ($j = 1,2$) for some finite $c_j > 0$.

The bandwidths undersmooth $\tilde{g}$, so there is no asymptotic bias, and can be selected by applying cross-validation or plug-in methods to the local linear estimator of g. The following proposition generalizes Proposition 1 to the case of a bivariate X.

Proposition 1′: Let Assumptions 1′(i), 1′(ii), 2′, and 3′ hold. For each $x \in (0,1)^2$,

$$(nh_1h_2)^{1/2}[\tilde{g}(x) - g(x)] \to^d N(0, \sigma^2_{\tilde{g}(x)}),$$

where $\sigma^2_{\tilde{g}(x)} > 0$ is finite.■

The expression for $\sigma^2_{\tilde{g}(x)}$ is lengthy and is given by Ruppert and Wand (1994, equation (4.7)). It requires a consistent estimator of σ^2_U for a bivariate X. To do this, let $j(i)$ (i = 1, …, n) be a set of indices that is defined through the following recursion:

$$j(1) = \arg\min_{j=2,\dots,n} \left\| X_j - X_1 \right\|$$

and

$$j(i) = \arg\min_{j \neq i, j(1),\dots j(i-1)} \left\| X_j - X_i \right\|; \quad i = 2,\dots,n.$$

The number $j(i)$ is the index of the design point that is nearest to X_i among those whose indices are not $j(1), \dots, j(i - 1)$. Then σ^2_U can be estimated by

$$\hat{\sigma}^2_U = \frac{1}{2n}\sum_{i=1}^{n}(Y_i - Y_{j(i)})^2$$

Under Assumption 1′, $\hat{\sigma}^2_U$ is a $n^{1/2}$-consistent estimator of σ^2_U (Horowitz and Spokoiny 2001).

The grid consists of J^2 equally spaced points $\{x_{jk} = (x_{1j}, x_{2k}): j,k = 1,\dots,J\}$. The shape restriction is

$$A_k(\boldsymbol{g}) \le 0; \ k = 1,\dots,\kappa,$$

where $\boldsymbol{g}$ is the $J^2 \times 1$ vector whose components are $g(x_{1j}, x_{2k})$ ($j,k = 1,\dots,J$). To obtain the bivariate extension of Theorem 5.1 replace Assumption 4 by

Assumption 4′: (i) $J \ll n^{1/6}(\log n)^{-1/4}$. (ii) $h_1, h_2 < 1/[2(J+1)]$. (iii) $J \to \infty$ as $n \to \infty$.

The bivariate extension of Theorem 5.1 is obtained from the scalar version by replacing Assumptions 1-4 with Assumptions 1′-4′, J with J^2, and $\sigma^2_{\tilde{g}(x_j)}$ with its bivariate extension, $\sigma^2_{\tilde{g}(x_{jk})}$.

To define the constrained estimator, index the grid points by $\ell \in \mathcal{C}$. For $\ell \in \mathcal{C}$ let

(6.2) $$\boldsymbol{X}_i^{(\ell)} = \left[1 \ (X_{i1} - x_{\ell 1}) \ \ (X_{i2} - x_{\ell 2}) \ \ (X_{i1} - x_{\ell 1})^2 \ \ (X_{i2} - x_{\ell 2})^2 \ \ (X_{i1} - x_{\ell 1})(X_{i2} - x_{\ell 2})\right].$$

Define the matrix $\boldsymbol{Q}$ and vector $\boldsymbol{d}$ as in Section 2 but with K_h and $\boldsymbol{X}^{(\ell)}$ as in (6.1) and (6.2). Let $\boldsymbol{b}$ be the $6|\mathcal{C}|\times 1$ vector $(b_{11}, b_{21} \dots, b_{61}, \dots, b_{1|\mathcal{C}|}, b_{2|\mathcal{C}|}, \dots, b_{6|\mathcal{C}|})'$. Define

(6.3) $$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}}(0.5\boldsymbol{b}'\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{d}'\boldsymbol{b}).$$

subject to:

$$A_k(\boldsymbol{g}) = 0; \quad k \in \hat{\mathcal{S}}.$$

Problem (6.3) is the same as (3.3) but with K_h and $\boldsymbol{X}^{(\ell)}$ as in (6.1) and (6.2). Define active components of $\boldsymbol{g}$ as in Section 3. The constrained estimator of the active components of $\boldsymbol{g}$ is

$$\hat{\boldsymbol{g}}^{(a)} = (\hat{b}_{11}, \hat{b}_{71}, .., \hat{b}_{6|\mathcal{C}|-5})'.$$

To obtain the bivariate extensions of Theorems 5.1 and 5.2, redefine $|\mathcal{M}|$ as the number of constraints for which $0 < A_k(\tilde{\boldsymbol{g}}^{(J)}) \le [(\log n)/(nh_1h_2)]^{1/2}$, where $\tilde{\boldsymbol{g}}^{(J)}$ is the $J^2 \times 1$ vector $[\tilde{g}(x_{11}),...,\tilde{g}(x_{JJ})]'$. The bivariate extensions of Theorems 5.1 and 5.2 are obtained from the scalar versions of these theorems by replacing Assumptions 1-4 with Assumptions 1′-4′, using the redefined $|\mathcal{M}|$ in Assumption 5(iv), replacing J with J^2, $(nh)^{1/2}$ with $(nh_1h_2)^{1/2}$, and the scalar version of $\sigma^2_{\tilde{g}(x_j)}$ with its bivariate extension. The appendix gives an example in which the redefined $|\mathcal{M}| \to 0$ in the bivariate case.

7. MONTE CARLO EXPERIMENTS AND AN EMPIRICAL EXAMPLE

This section presents the results of Monte Carlo experiments and an empirical example that illustrate the usefulness of the shape-restricted estimator described in Sections 3-6. The empirical example consists of estimating a production function under a shape constraint. The Monte Carlo experiments are designed to mimic the empirical example and illustrate the finite-sample performance of the uniform confidence band based on the shape-restricted estimator.

To describe the model used in the experiments and example, let Y, K, and L, respectively, denote value-added output, capital, and labor. Suppose that

(7.1) $\quad \log Y = f(K,L) + U$,

where U is an unobserved random variable that is independent of K and L and satisfies $E(U) = 0$. Suppose that the function $\exp[f(K,L)]$ satisfies constant or decreasing returns to scale in levels. That is

(7.2) $\quad \exp[f(\lambda K, \lambda L)] \le \lambda \exp[f(K,L)]$

for all $\lambda > 0$. It is customary to use the log transformation in empirical economics, and we follow that convention here. Taking logarithms on both sides of (7.2) yields

(7.3) $\quad f(\lambda K, \lambda L) \le \log\lambda + f(K,L)$

for all $\lambda > 0$. Define $y = \log Y$, $k = \log K$, $\ell = \log L$, $g(k,\ell) = f(e^k, e^\ell)$, and $\tilde{\lambda} = \log\lambda$. Then (7.3) is equivalent to

(7.4) $\quad g(\tilde{\lambda} + k, \tilde{\lambda} + \ell) \le \tilde{\lambda} + g(k,\ell)$.

Model (7.1) is equivalent to

(7.5) $y = g(k, \ell) + U$.

The Monte Carlo experiments and empirical example are based on (7.4) and (7.5).

7.1 *Monte Carlo Experiments*

This section presents the results of a small set of Monte Carlo experiments that illustrate the finite-sample performance of the uniform confidence band for $g(k,\ell)$ in (7.5) using the shape restricted estimator. In the experiments, samples of size $n = 1000$ and $n = 2000$ were generated from the production function

$$g(k,\ell) = \log\left(\left\{[\exp(k)]^{1/2} + [\exp(\ell)]^{1/2}\right\}^{2\tau}\right)$$

for some constant $\tau > 0$. The resulting model is

$$y = \log\left(\left\{[\exp(k)]^{1/2} + [\exp(\ell)]^{1/2}\right\}^{2\tau}\right) + U . \tag{7.6}$$

Model (7.6) is equivalent to the following production function model in levels:

$$Y = \exp[(K^{1/2} + L^{1/2})^{2\tau} + U] .$$

Values of k and ℓ in (7.6) were generated randomly and independently of each other from the $U[0,1]$ distribution. U was sampled independently of (k,ℓ) from the $N(0, 0.01)$ distribution.

We report results for $\tau = 1$ (constant returns to scale) and $\tau = 0.5$ (decreasing returns to scale). In each experiment, the shape restriction is that g satisfies non-increasing returns to scale. Thus, $\tau = 1$ when the shape constraint is binding.

We used the grid

$$\left\{(k_i, \ell_j) = \left(\frac{i}{J+1}, \frac{j}{J+1}\right) : i, j = 1, ..., J\right\}, \tag{7.7}$$

where J is chosen using the method described in the next paragraph. Using this grid, the discrete version of (7.4) is

$$g\left(\frac{i+1}{J+1}, \frac{j+1}{J+1}\right) \le \frac{1}{J+1} + g\left(\frac{i}{J+1}, \frac{j}{J+1}\right)$$

for every $i, j = 1, ..., J-1$.

We used the local quadratic estimator of g described in Section 3 with the uniform kernel function. A baseline bandwidth $h_0 = 0.15$ was determined by auxiliary simulations. Then we set $h = C_h h_0 (n/1000)^{-1/6}$, where $C_h = 0.95, 1,$ or 1.05, depending on the experiment. In practice, as is explained in Section 3, h can be chosen by cross-validation for local linear estimation. We did not use cross-validation in the Monte Carlo experiments because of its computational cost. The size of the grid, J, was $J = \lfloor J(h) \rfloor$, the largest integer not greater than $J(h)$, where

$$J(h) = \min\left[\frac{1}{h}(\log n)^{-1/2}, \frac{1}{2h} - 1.1\right]. \tag{7.8}$$

This choice of J satisfies Assumptions 4´(i) and 4´(ii).

There were 1,000 Monte Carlo replications in each experiment with 10,000 draws used to estimate the distribution of Z. When the estimated set of possibly binding constraints is nonempty, the limiting distribution of Z is degenerate. We used the singular value decomposition (SVD) to deal with singularity.

We present results for symmetrical nominal 95% uniform confidence bands for g obtained from the shape-restricted estimator. We also present results on uniform confidence bands using the unconstrained estimator (that is, the returns to scale constraint was not imposed) and the infeasible constrained estimator in which the true set of binding constraints, $\mathcal{S}$, is used in place of the estimated set $\hat{\mathcal{S}}$. We call this the oracle estimator.

The results of the experiments are shown in Tables 1-2. Column 3 of the tables shows the empirical coverage probabilities of the uniform confidence bands. Column 4 shows the relative average widths of the various confidence bands. These are the ratios of the average widths of the bands to the average width of the band based on the unconstrained estimator. The relative width of the latter band is 1.0 by definition. The relative widths of the bands based on the constrained and oracle estimators are smaller (larger) than 1.0 according to whether the average widths of these bands are smaller (larger) than the average widths of the bands obtained from the unconstrained estimator. The results show that the empirical coverage probabilities of the bands obtained with all estimators are close to the nominal probability of 0.95. The confidence bands based on the constrained estimator are narrower than the bands based on the unconstrained estimator when the constraints are binding ($\tau = 1$) and have almost the same width when the constraints are not binding ($\tau = 0.5$). In this set of experiments, the constrained estimator performs as well as the oracle estimator.

7.2 *Empirical Example*

This section reports the results of estimating a production function for the Chinese chemical industry using the firm-level data of Jacho-Chávez, Lewbel, and Linton (2010). We estimated the production function using data for 1995 and 2001. The dependent variable, y, is the logarithm of value-added real output. The explanatory variables are the logarithm of the net value of real fixed assets, k, and the logarithm of the number of employees, ℓ. As in Jacho-Chávez, Lewbel, and Linton (2010), observations with outliers are removed and both regressors are normalized by their respective medians. As in the Monte Carlo experiments, we used the local quadratic estimator with a uniform kernel function. For each year, the bandwidth, h, was chosen by cross-validation. The grid points were chosen to be within the support of (k,ℓ) in the data. The number of points, J, was determined by (7.8). The sample sizes were $n=1560$ for 1995 and $n=1638$ for 2001. Increasing returns to scale are unlikely in the chemical industry. Accordingly, we carried out unconstrained estimation of g and estimation under the restriction of non-increasing returns to scale.

Table 3 and Figure 1 present the estimation results at several points (k,ℓ) for which the normalized values of k and ℓ are equal. The constrained and unconstrained estimates are similar, as is to be expected in an industry that has non-increasing returns. However, the constrained estimates are more precise than the unconstrained ones. For example, in 1995 the constrained and unconstrained point estimates of $g(2.524, 2.524)$ are the same, but the standard error of the constrained estimate is much less than that of the unconstrained estimate. It can be seen from Figure 1 that the constrained estimates are slightly more precise than the unconstrained ones in the middle of the distribution of (k,ℓ) and much more precise near the boundaries of the support of (k,ℓ).

8. CONCLUSIONS

Economic theory often provides shape restrictions on functions of interest in applications, but it does not provide finite-dimensional parametric models. This motivates nonparametric estimation under shape restrictions. Shape restrictions can stabilize noisy nonparametric estimates without imposing arbitrary restrictions, such as additivity or a single-index structure, that may be inconsistent with economic theory and the data. This paper has explained how to estimate and obtain an asymptotic uniform confidence band for a conditional mean function under a possibly nonlinear shape restriction. There is a large literature in statistics and econometrics on estimating a conditional mean function under linear shape restrictions, such as monotonicity or convexity. To our knowledge, this paper is the first to construct a uniform confidence band under shape restrictions such as non-increasing or non-decreasing returns to scale and the Slutsky inequality of consumer theory. The results of Monte Carlo experiments and an empirical application have illustrated the finite-sample performance and usefulness of our method. The methods of this paper can be extended to conditional quantile functions with shape restrictions, though doing so is complicated technically because of the non-differentiability of the objective function of quantile estimation. Estimation of conditional quantile functions under shape restrictions is a topic for further research.

9. PROOFS OF THEOREMS

9.1 *Examples in which* $|\mathcal{M}| = o(1)$ *as* $n \to \infty$

Scalar case: Let X be a scalar and

$$g(x) = \begin{cases} 0 \text{ if } 0 \le x \le 0.5 \\ (x-0.5)^2 \text{ if } 0.5 < x \le 1. \end{cases}$$

Let there be J equally spaced grid points in $[0,1]$. Assume that J is odd so that $x_{(J+1)/2} = 0.5$ and $x_j > 0.5$ implies that $j > (J+1)/2$. The shape restriction is that g is non-decreasing, so $g(x_{j-1}) - g(x_j) \le 0$. Let $x_j > x_{j-1} \ge 0.5$ be grid points. Then $j \ge (J+3)/2$, and

$$|g(x_j) - g(x_{j-1})| = \frac{1}{(J+1)^2}[2j-1-(J+1)]$$

$$< \frac{2j}{(J+1)^2}.$$

Therefore,

$$0 < |g(x_j) - g(x_{j-1})| \le \left(\frac{\log n}{nh}\right)^{1/2} \tag{9.1}$$

implies that

$$j \le \frac{(J+1)^2}{2}\left(\frac{\log n}{nh}\right)^{1/2}$$

Under Assumptions 3(ii) and 4(i),

$$\frac{(J+1)^2}{2}\left(\frac{\log n}{nh}\right)^{1/2} \to 0$$

as $n \to \infty$. Therefore, there can be no grid points x_j and x_{j-1} satisfying (9.1) if n is sufficiently large, which implies that $|\mathcal{M}| = 0$ if n is sufficiently large.

Bivariate case: Let $X = (X_1, X_2)'$ be two dimensional. Let

$$g(x_1, x_2) = \begin{cases} 0 \text{ if } x_1^2 + x_2^2 \le 0.25 \\ x_1^2 + x_2^2 - 0.25 \text{ if } x_1^2 + x_2^2 > 0.25 \end{cases}$$

Let there be J equally spaced grid points in each of the two dimensions. Assume that J is odd so that $x_{1,(J+1)/2} = x_{2,(J+1)/2} = 0.5$. Therefore, $x_{1,j} > 0.5$ or $x_{2,k} > 0.5$ implies that $j + k > J + 1$. The shape restriction is

$$g(x_{1,j-1},x_{2k})-g(x_{1j},x_{2k})\le 0$$

$$g(x_{1j},x_{2,k-1})-g(x_{1j},x_{2k})\le 0$$

This is the finite-difference analog of the restriction $\partial g(x_1,x_2)/\partial x_1 \ge 0$ and $\partial g(x_1,x_2)/\partial x_2 \ge 0$.

If $g(x_{1j},x_{2k})>0$ and $j\ge 2$, then

$$g(x_{1j},x_{2k})-g(x_{1,j-1},x_{2k})=\frac{1}{(J+1)^2}(2j-1) \text{ if } g(x_{1,j-1},x_{2k})>0$$

$$g(x_{1j},x_{2k})-g(x_{1,j-1},x_{2k})\le\frac{1}{(J+1)^2}(2j-1) \text{ if } g(x_{1,j-1},x_{2k})=0.$$

Therefore,

$$g(x_{1,j},x_{2k})-g(x_{1,j-1},x_{2k})\le\frac{2j}{(J+1)^2}.$$

Similarly, if $g(x_{1j},x_{2k})>0$ and $k\ge 2$, then

$$g(x_{1j},x_{2k})-g(x_{1j},x_{2,k-1})\le\frac{2k}{(J+1)^2}.$$

Now proceed as in the example for a scalar X.

9.2 *Proofs of Theorems 4.1, 5.1, and 5.2*

Proof of Theorem 4.1: By (3.1),

$$(nh)^{1/2}[\tilde{g}(x_j)-g(x_j)]\to^d e_1'(nh/n_{x_j})(\boldsymbol{S}_n^{(x_j)})^{-1}(nh)^{-1/2}\boldsymbol{X}^{(x_j)\prime}\boldsymbol{W}^{(x_j)}\boldsymbol{U}^{(x_j)}.$$

Define

$$\tilde{\boldsymbol{X}}^{(x_j)}=\begin{bmatrix} 1 & h^{-1}(X_{i_1}-x_j) & h^{-2}(X_{i_1}-x_j)^2 \\ & ..... & \\ 1 & h^{-1}(X_{i_{n_x}}-x_j) & h^{-2}(X_{i_{n_x}}-x_j)^2 \end{bmatrix}$$

and

$$\boldsymbol{e}_h=\begin{pmatrix} 1 & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & h^2 \end{pmatrix}.$$

Then,

$$(nh)^{1/2}[\tilde{g}(x_j)-g(x_j)]\to^d e_1'(nh/n_{x_j})(\boldsymbol{S}_n^{(x_j)})^{-1}\boldsymbol{e}_h(nh)^{-1/2}\tilde{\boldsymbol{X}}^{(x_j)\prime}\boldsymbol{W}^{(x_j)}\boldsymbol{U}^{(x_j)}.$$

The 3×1 vector $(nh)^{-1/2}\tilde{\boldsymbol{X}}^{(x_j)\prime}\boldsymbol{W}^{(x_j)}\boldsymbol{U}^{(x_j)}$ is asymptotically trivariate normally distributed with mean 0 by the multivariate extension of the Lindeberg-Levy central limit theorem. Let $\tilde{\Sigma}_j$ denote the covariance matrix of the limiting distribution. Let $\mathcal{B}^J$ denote the set of all convex sets in $\mathbb{R}^J$ and Z_j be a random vector with the $N(0,\tilde{\Sigma}_j)$ distribution. It follows from Theorem 1.1 of Bentkus (2003) (see, also, Corollary 11.1 of DasGupta (2008)) that for some constant $c_1 < \infty$

$$(9.2)\qquad \sup_{B\in\mathcal{B}^3}\left|P[(nh)^{-1/2}\tilde{\boldsymbol{X}}^{(x_j)\prime}\boldsymbol{W}^{(x_j)}\boldsymbol{U}^{(x_j)} \in B] - P(Z_j \in B)\right| \le c_1(nh)^{-1/2}.$$

Let $\boldsymbol{\xi}$ denote the $3J\times 1$ vector whose $3j-2,\ldots,3j$ components are $(nh)^{-1/2}\tilde{\boldsymbol{X}}^{(x_j)\prime}\boldsymbol{W}^{(x_j)}\boldsymbol{U}^{(x_j)}$. Let $\tilde{Z}$ denote a random vector with the $N(0,\tilde{\Sigma})$ distribution, where

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_1 & \ldots & 0 \\ 0 & \ldots & 0 \\ 0 & \ldots & \tilde{\Sigma}_J \end{pmatrix}.$$

Then (9.2) implies that for some $c_2 < \infty$

$$(9.3)\qquad \sup_{B\in\mathcal{B}^{3J}}\left|P(\boldsymbol{\xi} \in B] - P(\tilde{Z} \in B)\right| \le c_2 J(nh)^{-1/2}.$$

Let L_j denote the probability limit of $e_1'(nh/n_{x_j})(\boldsymbol{S}_n^{(x_j)})^{-1}$ as $n\to\infty$. Standard calculations for kernel estimators show that

$$(9.4)\qquad e_1'(nh/n_{x_j})(\boldsymbol{S}_n^{(x_j)})^{-1} = L_j + O_p[(nh)^{-1/2}].$$

The theorem follows by combining (9.3), (9.4), and Assumption 4. Q.E.D.

<u>Proof of Theorem 5.1</u>: By Theorem 4.1, $(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(a)} - \tilde{\boldsymbol{g}}^{(a)})$ and $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}}^{(-a)} - \boldsymbol{g}^{(-a)})$ are asymptotically multivariate normal with means of zero. Therefore, $(nh)^{1/2}(\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g})$ is a linear combination of asymptotic multivariate normals with means of zero and is asymptotically multivariate normally distributed with a mean of 0. The covariance matrix Σ_n follows from (3.2), (5.2), and independence of the components of $\hat{\hat{\boldsymbol{g}}} - \boldsymbol{g}$ from one another. Q.E.D.

The following notation is used in the proof of Theorem 5.2. Define the $(3|\mathcal{C}|+|\mathcal{S}|)\times(3|\mathcal{C}|+|\mathcal{S}|)$ matrix

$$H = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}_{\boldsymbol{b}}' \\ \boldsymbol{A}_{\boldsymbol{b}} & 0_{|\mathcal{S}|\times|\mathcal{S}|} \end{pmatrix}$$

and the $(3|\mathcal{C}|+|\mathcal{S}|)\times|\mathcal{S}|$ matrix

$$V = \begin{pmatrix} 0_{3|\mathcal{C}|\times|\mathcal{S}|} \\ I_{3|\mathcal{S}|\times|\mathcal{S}|} \end{pmatrix}.$$

Then

$$H^{+} = \begin{pmatrix} R_{11} & R_{12} \\ R_{12}' & R_{22} \end{pmatrix},$$

where

$$R_{11} = \boldsymbol{Q}^{-1}[I_{3|\mathcal{C}|\times 3|\mathcal{C}|} - \boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}],$$

$$R_{12} = \boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+},$$

and

$$R_{22} = -(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}.$$

In addition,

$$H^{+}V = \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix}.$$

Proof of Theorem 5.2: Let $\hat{\boldsymbol{b}}$ denote the solution to (5.9), and note that only components of $\boldsymbol{g}$ in $\mathcal{C}$ are affected by the constraints. It follows by corollaries 3.2.4 and 3.2.5 of Fiacco (1983) that

$$\hat{\boldsymbol{b}} - \tilde{\boldsymbol{b}}^{(a)} = -R_{12}\boldsymbol{A}(\tilde{\boldsymbol{b}}^{(a)}) + O\left(\left\|\boldsymbol{A}(\tilde{\boldsymbol{b}}^{(a)}\right\|^2\right),$$

where $\boldsymbol{A}(\cdot) = [A_1(\cdot), ..., A_{|\mathcal{S}|}(\cdot)]'$ and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^{|\mathcal{S}|}$. Therefore,

$$\hat{\boldsymbol{b}} = \tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}(\tilde{\boldsymbol{b}}^{(a)}) + O\left(\left\|A(\tilde{\boldsymbol{b}}^{(a)})\right\|^2\right),$$

Recall that $\boldsymbol{b}_0$ denotes the $3|\mathcal{C}|\times 1$ vector

$$\boldsymbol{b}_0 = \{[g(x_j), g'(x_j), g''(x_j)]' : j \in \mathcal{C}\}$$

and that the modified definition of $A_k(\cdot)$ includes all components of $\boldsymbol{b}$ in its arguments. Then Taylor series expansions yield

$$A_k(\tilde{\boldsymbol{b}}^{(a)}) = A_k(\boldsymbol{b}_0) + \frac{\partial A_k(\breve{\boldsymbol{b}}_k)}{\partial \boldsymbol{b}'}(\tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{b}_0)$$

$$= A_k(\boldsymbol{b}_0) + \frac{\partial A_k(\boldsymbol{b}_0)}{\partial \boldsymbol{b}'}(\tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{b}_0) + \left[\frac{\partial A_k(\boldsymbol{b}_0)}{\partial \boldsymbol{b}'} - \frac{\partial A_k(\breve{\boldsymbol{b}}_k)}{\partial \boldsymbol{b}'}\right](\tilde{\boldsymbol{b}}^{(a)} - \boldsymbol{b}_0)$$

for each $k = 1,...,|\mathcal{S}|$, where $\breve{\boldsymbol{b}}$ is between $\tilde{\boldsymbol{b}}^{(a)}$ and $\boldsymbol{b}_0$. Under Assumption 5, $\partial A_k(\boldsymbol{b}) / \partial \boldsymbol{b}$ is a continuous function of $\boldsymbol{b}$. Moreover columns of $\partial A_k(\boldsymbol{b}) / \partial \boldsymbol{b}$ corresponding to derivatives of g are zero, and $\boldsymbol{A}(\boldsymbol{b}_0) = 0$ because $\boldsymbol{A}$ is defined as the vector giving the binding constraints. Therefore,

$$\boldsymbol{A}(\tilde{\boldsymbol{b}}^{(a)}) = \boldsymbol{A}_{\boldsymbol{b}}(\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + O\left(\left\| \tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} \right\|^2 \right)$$

and

$$\hat{\boldsymbol{b}} - \boldsymbol{b}_0 = [I_{3|\mathcal{C}|\times 3|\mathcal{C}|} - \boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}](\tilde{\boldsymbol{b}} - \boldsymbol{b}_0) + O\left(\left\| \tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} \right\|^2 \right).$$

But $\boldsymbol{A}_{\boldsymbol{b}}(\tilde{\boldsymbol{b}} - \boldsymbol{b}_0) = \boldsymbol{A}_{\boldsymbol{b}}(\tilde{\boldsymbol{b}} - E\tilde{\boldsymbol{b}})$, because the estimator of $\boldsymbol{g}$ is undersmoothed and columns of $\boldsymbol{A}_{\boldsymbol{b}}$ corresponding to derivatives of $\boldsymbol{g}$ are zero. Therefore,

$$\hat{\boldsymbol{g}} - \boldsymbol{g}^{(a)} = [I_{3|\mathcal{C}|\times 3|\mathcal{C}|} - \boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}]^{(a)}(\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + O\left(\left\| \tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} \right\|^2 \right)$$

$$= \{I_{|\mathcal{C}|\times|\mathcal{C}|} - [\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}]^{(a)}\}(\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + O\left(\left\| \tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)} \right\|^2 \right),$$

and

$$(nh)^{1/2}(\hat{\boldsymbol{g}} - \boldsymbol{g}^{(a)}) = \{I_{|\mathcal{C}|\times|\mathcal{C}|} - [\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}'(\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{Q}^{-1}\boldsymbol{A}_{\boldsymbol{b}}')^{+}\boldsymbol{A}_{\boldsymbol{b}}]^{(a)}\}(nh)^{1/2}(\tilde{\boldsymbol{g}}^{(a)} - \boldsymbol{g}^{(a)}) + o_p(1).$$

Now proceed as in the proof of Theorem 5.1. Q.E.D.

REFERENCES

Andrews, D.W.K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67, 1341-1383.

Andrews, D.W.K. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica*, 81, 609-666.

Bentkus, V. (2003). On the dependence of the Berry-Esséen bound on dimension. *Journal of Statistical Planning and Inference,* 113, 385-402.

Birke, M. and H. Dette (2006). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics*, 34, 384-404.

Blundell, R., Horowitz, J.L., and Parey, M. (2012). Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, 3, 29-51.

Blundell, R., Horowitz, J.L., and Parey, M. (2013). Nonparametric estimation of a heterogeneous demand function under the Slutsky inequality restriction. cemmap working paper CWP54/13, Centre for Microdata Methods and Practice, London.

Buckley, M.J., G.K. Eagleson, and B.W. Silverman (1988). The Estimation of Residual Variance in Nonparametric Regression, *Biometrika*, 75, 189-199.

Calonico, S., M.D. Cattaneo, and M.H. Farrell (2014). On the effect of bias estimation on coverage accuracy in nonparametric inference. Working paper, University of Michigan.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96, 559-575.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: estimation and inference. *Econometrica*, 81, 667-737.

Das Gupta, A. (2008). Asymptotic theory of statistics and probability. New York: Springer.

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

Du, P. C. Parmeter, and J.C. Racine (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, 23, 1343-1372.

Dumbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli*, 9, 423-449.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

Fiacco, A.V. (1983). *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. New York: Academic Press.

Gasser, T., L. Sroka, and C. Jennen-Steinmetz (1986). Residual Variance and Residual Pattern in Nonlinear Regression, *Biometrika*, 73, 625-633.

Groeneboom, P., G. Jongbloed, and J.A. Wellner (2001). Estimation of a convex function: characterizations and asymptotic theory. *Annals of Statistics*, 29, 1653-1698.

Hall, P. and J.L. Horowitz (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics*, 41, 1892-1921.

Hall, P. and L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, 29, 624-647.

Hall, P. and L.-S. Huang (2002). Unimodal density estimation using kernel methods. *Statistica Sinica*, 12, 965-990.

Hall, P., L.-S. Huang, J.A. Gifford, and I. Gijbels (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statistics*, 10, 592-614.

Hall, P. and B. Presnell (1999). Density estimation under constraints. *Journal of Computational and Graphical Statistics*, 8, 259-277.

Hall, P. and A. Yatchew (2005). Unified approach to teting functional hypotheses in semiparametric contexts. *Journal of Econometrics*, 127, 225-252.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Jacho-Chávez, D., A. Lewbel, and O. Linton (2010). Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics*, 156, 392-407.

Lee, S., K. Song, and Y.-J. Whang (2013). Testing functional inequalities. *Journal of Econometrics*, 172, 14-32.

Liew, C.K. (1976). Inequality constrained least-squares estimation. *Journal of the American Statistical Association*, 71, 746-751.

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

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

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

Pal, J.K. and Woodroofe M. (2007). Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statistica Sinica*, 17, 1601-1616.

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression, *Annals of Statistics*, 12, 1215-1230.

Romano, J., A. Shaikh, and M. Wolf (2014). A practical two-step method for testing moment inequality models. *Econometrica*, 82, 1979-2002.

Ruppert, D. and M.P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22, 1346-1370.

Wang, X. and J. Shen (2013). Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM Journal on Control and Optimization*, 51, 2753-2787.

Table 1: Results of Monte Carlo Experiments when the Constraints Are Binding

Model: Constant returns to scale; $\tau = 1$
Nominal coverage probability: 0.95

Unconstrained Estimation

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.946	1.0
1000	1.00	0.942	1.0
1000	1.05	0.945	1.0
2000	0.95	0.954	1.0
2000	1.00	0.952	1.0
2000	1.05	0.951	1.0

Constrained Estimation

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.947	0.818
1000	1.00	0.933	0.818
1000	1.05	0.944	0.818
2000	0.95	0.953	0.819
2000	1.00	0.954	0.819
2000	1.05	0.960	0.818

Oracle Estimation

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.949	0.816
1000	1.00	0.935	0.817
1000	1.05	0.944	0.817
2000	0.95	0.958	0.817
2000	1.00	0.959	0.817
2000	1.05	0.962	0.817

Table 2: Results of Monte Carlo Experiments when the Constraints Are Not Binding

Model: Decreasing returns to scale; $\tau = 0.5$
Nominal coverage probability: 0.95

Unconstrained Estimation

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.946	1.0
1000	1.00	0.941	1.0
1000	1.05	0.943	1.0
2000	0.95	0.954	1.0
2000	1.00	0.951	1.0
2000	1.05	0.951	1.0

Constrained Estimation

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.946	0.999
1000	1.00	0.942	1.0
1000	1.05	0.944	1.0
2000	0.95	0.953	1.0
2000	1.00	0.949	1.0
2000	1.05	0.952	1.0

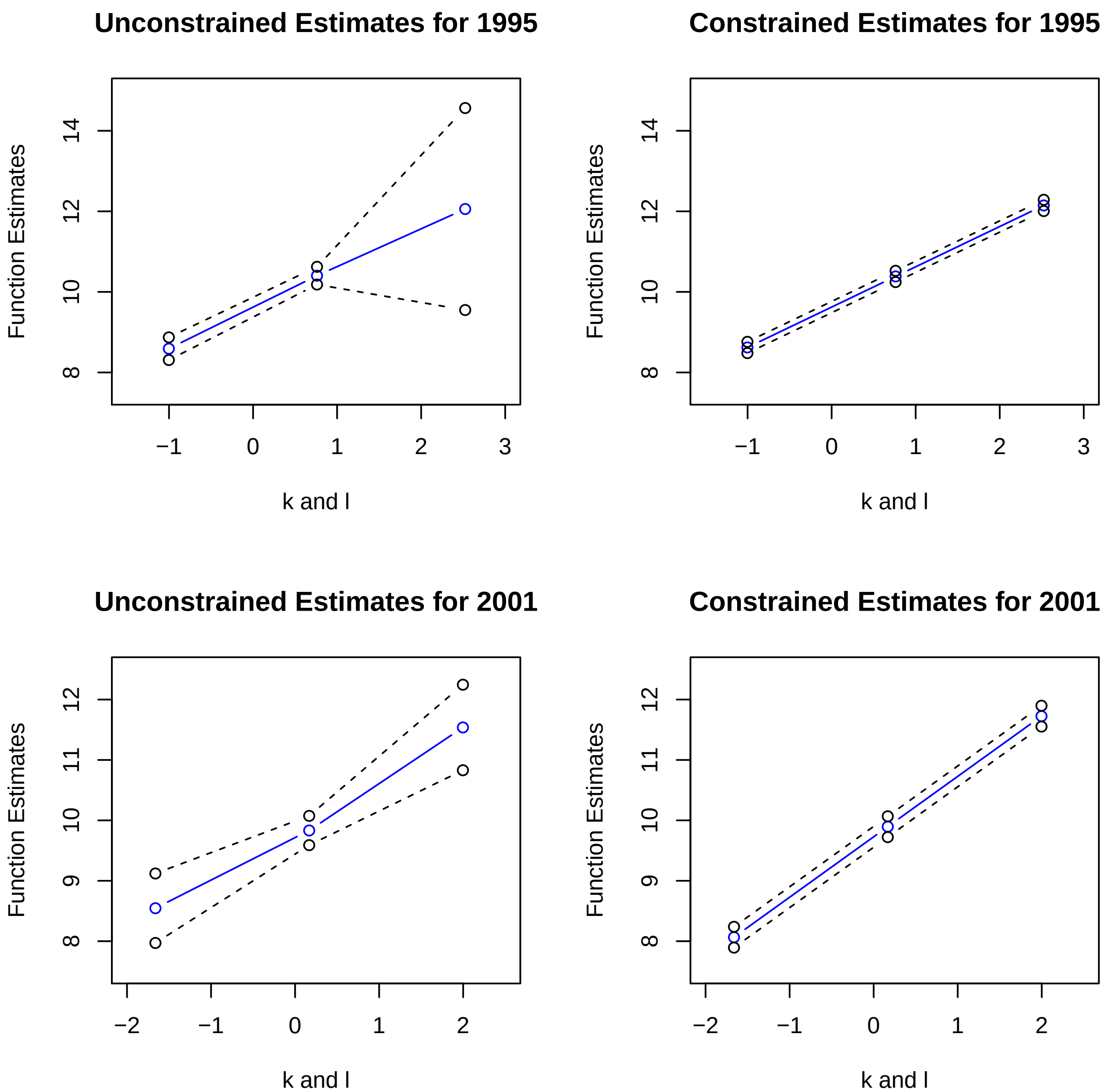
Oracle Estimatio

n	C_h	Empirical Cov. Prob.	Relative Width
1000	0.95	0.948	1.0
1000	1.00	0.942	1.0
1000	1.05	0.944	1.0
2000	0.95	0.953	1.0
2000	1.00	0.949	1.0
2000	1.05	0.952	1.0

Table 3: Results of Estimating a Production Function

Year	Estimation Method	k, ℓ	Function Estimate	Standard Error	Lower 95% Uniform Bound	Upper 95% Uniform Bound
1995	Unconstrained	-1.002	8.591	0.117	8.309	8.872
		0.761	10.403	0.092	10.183	10.623
		2.524	12.058	1.044	9.550	14.565
	Constrained	-1.002	8.621	0.072	8.481	8.761
		0.761	10.384	0.072	10.244	10.524
		2.524	12.147	0.072	12.007	12.287
2001	Unconstrained	-1.662	8.546	0.239	7.971	9.121
		0.168	9.833	0.101	9.590	10.075
		1.997	11.538	0.295	10.830	12.246
	Constrained	-1.662	8.066	0.089	7.894	8.239
		0.168	9.806	0.089	9.723	10.068
		1.997	11.725	0.089	11.553	11.898

Figure 1. Nonparametric Estimates and their Uniform Confidence Bands



Note: The solid lines represent nonparametric estimates, whereas the dashed lines show 95% uniform confidence bands. The circles correspond to the grid. On one hand, the top and bottom panels show estimates for 1995 and 2001, respectively. On the other hand, the left and right panels show unconstrained and constrained estimates, respectively.