TESTS OF ADDITIVE DERIVATIVE CONSTRAINTS

by Thomas M. Stoker

March 1987, Revised October 1988

WP#2079-88

ABSTRACT

This paper proposes nonparametric tests of additive constraints on the first and second derivatives of a model $E(y|x)=g(x)$, where the true function g is unknown. Such constraints are illustrated by the economic restrictions of homogeneity and symmetry, and the functional form restrictions of additivity and linearity. The proposed tests are based on estimates of regression coefficients, that statistically characterize the departures from the constraint exhibited by the data. The coefficients are based on weighted average derivatives, that are reformulated in terms of derivatives of the density of x. Coefficient estimators are proposed that use nonparametric kernel estimators of the density and its derivatives. These statistics are shown to be $\sqrt{N}$ consistent and asymptotically normal, and thus are comparable to estimators based on a (correctly specified) parametric model of $g(x)$.

TESTS OF ADDITIVE DERIVATIVE CONSTRAINTS

1. Introduction

Derivative constraints play an important role in the empirical study of economic behavior. One source of derivative constraints is the standard implications of economic theory for marginal responses. For instance, economic theory implies that production costs are homogeneous in input prices and that demand functions are zero-degree homogeneous in prices and income, which are restrictions on the derivatives of cost and demand functions respectively. The symmetry conditions of optimization provide other examples; for instance, cost minimization implies equality constraints on the derivatives of input quantities with respect to input prices.

Derivative constraints also arise from restrictions used to simplify econometric models. These include include constant returns-to-scale restrictions on production functions and exclusion restrictions on demand or production systems. Such restrictions are valuable for increasing precision in estimation or facilitating applications of econometric models.

Associated with the use of derivative constraints is the necessity of testing their statistical validity. Rejection of a constraint representing a basic implication of economic theory suggests either a revision of model specification, or a reconsideration of the applicability of the theory to the empirical problem at hand. The use of restrictions to simplify empirical models is justified only when the restrictions are not in conflict with the data evidence.

In current practice, derivative constraints are typically tested using a parametric approach. Here a specific functional form of behavioral equations is postulated, and the constraints on behavioral derivatives are related to restrictions on the parameters to be estimated. Tests of the derivative constraints then coincide with standard hypothesis tests of the parametric

restrictions. This approach is limited by the initial specification of the parametric model, which must be held as a maintained assumption which the restrictions are tested against. If the maintained assumption is in conflict with the process generating the observed data, the results are uninterpretable.[1]

The purpose of this paper is to propose a nonparametric approach to testing derivative constraints. To fix ideas, suppose that a behavioral model is represented by $g(x)=E(y|x)$, where y denotes a dependent variable, $x = (x_1,\ldots,x_M)'$ a vector of continuous variables, and the form of g is unknown. A "derivative constraint" refers to a restriction on the derivatives of g that holds for all values of x. In particular, we consider tests of additive constraints of the form

$$G_0(x)g(x) \;+\; \sum_j G_j(x)\,\frac{\partial g(x)}{\partial x_j} \;+\; \sum_{j\leq k} H_{jk}(x)\,\frac{\partial g^2(x)}{\partial x_j\partial x_k} \;=\; D(x) \qquad (H)$$

where $G_0(x)$, $G_j(x)$ and $H_{jk}(x)$, $j\leq k$, $j,k=1,\ldots,M$ and $D(x)$ are known, prespecified functions of x. The constraint (H) is intrinsically linear in $g(x)$ and its derivatives, but is otherwise unrestricted.

One obvious idea is to use a nonparametric smoothing technique to characterize $g(x)$ and its derivatives, and study the adherence of the estimated derivatives of g to (H) over the whole data sample. But such pointwise characterizations are notoriously imprecise, converging to the true derivatives at very slow rates as sample size increases, especially when x has more than two or three components.[2]

Instead, this paper proposes a method for testing derivative constraints based on a regression analysis of the departures from (H). Suppose that the departure from (H) was observed for each observation, and one performed an ordinary least squares (OLS) regression analysis of the departures on the components of x and the squares and cross-products of the components of x. If

2

(H) is valid, then all such regression coefficients must be zero. This paper gives estimators of the OLS coefficients from such a regression, and test statistics of the hypothesis that the coefficients vanish.

There are several attractive features of the proposed tests. First, the tests are nonparametric; they do not require a specification of the functional form of g(x). Second, the tests are interpretable; when rejection occurs, the source of rejection will be indicated by the regression coefficients on which the tests are based. Third, the tests are computationally simple; they are based on kernel density estimators that are computed directly from the data. Fourth, the tests have precision properties comparable to tests based on parametric models; the regression coefficients converge at rate $\sqrt{N}$ (where N is sample size), or the same rate of convergence displayed by parametric estimators. Consequently, the tests have non-zero efficiency relative to those based on a (correctly specified) parametric model.

Section 2 begins with several examples of constraints of the form (H). Section 3 introduces the regression approach to testing, and discusses its statistical power. Section 4 presents the estimators, test statistics, their interpretation and immediate extensions, and Section 5 gives some further remarks. Section 6 lists and briefly discusses the formal assumptions; as such, Section 6 can be read concurrently with the results, separately, or skipped depending upon the reader's interest in the technical requirements.[3]

2. Examples of Additive Derivative Constraints

The basic framework we consider is where the data (y_i, x_i), i=1,...,N represents random drawings from an underlying (joint) distribution of y and x. The relevant economic structure of the model is captured in the conditional expectation $E(y|x) = g(x)$, so that the constraints of interest are restrictions on the derivatives of g.[4] The marginal density of x is denoted f(x), which is taken to vanish on the boundary of x values.

As a notational convention, the subscript $i=1,\ldots,N$ always denotes an observation, and the subscripts $j,k=1,\ldots,M$ denote components of the vector x. For instance, $\partial g/\partial x_j$ is the jth partial derivative of g, x_i is the ith observation on x, and x_{ji} is the jth component of x_i.

We begin the examples with two cases of derivative constraints familiar from economic theory, namely homogeneity (of some degree) and symmetry.

Example 1: Homogeneity Restrictions. For concreteness, suppose that g(x) represents the logarithm of production and x represents the vector of log-input values: input levels are $I=e^x$ and quantity produced is $P(I)=e^{g(x)}$. $P(I)$ is homogeneous of degree d_0 in I if $P(\kappa I)=\kappa^{d_0}P(I)$ for any positive scalar κ, which obtains if and only if the log-form Euler equation is valid:

$$\sum_j \frac{\partial g(x)}{\partial x_j} = d_0 \tag{2.1}$$

where $\partial g/\partial x_j$ is the jth output elasticity and $\sum_j \partial g/\partial x_j$ is the "scale" elasticity. Constant returns-to-scale occur when $d_0=1$. (2.1) is clearly in the form (H) where $D(x)=d_0$, $G_0(x)=0$ and $G_j(x)=1$, $H_{jk}(x)=0$ for all j,k.[5] Note also that zero degree homogeneity restrictions, such as those applicable to cost and demand functions, take the form (2.1) with $d_0=0$.

Example 2: Symmetry Restrictions of Cost Minimization. Now suppose $g^j(x)$, $j=1,\ldots,M-1$, represent the demands for M-1 inputs, where x_j, $j=1,\ldots,M-1$, are input prices and x_M is the output of the firm. Cost minimization implies that

$$\frac{\partial g^j(x)}{\partial x_k} + \frac{\partial g^k(x)}{\partial x_j} = 0, \quad j,k = 1,\ldots,M-1 \tag{2.2}$$

This set of restrictions involves several behavioral equations, as discussed in Section 4.5.[6]

The following examples treat simplifying restrictions on the structure of g(x): exclusion restrictions and additive and linear functional forms.

Example 3: "$\underline{x_j}$ $\underline{has}$ $\underline{no}$ $\underline{effect}$ $\underline{on}$ $\underline{y}$". x_j can be omitted from g(x) if and only if

$$\frac{\partial g(x)}{\partial x_j} = 0 \qquad\qquad (2.3)$$

Example 4: Additivity and Linearity. g(x) is additive if g(x) = $\sum_j g_j(x_j)$, which is equivalent to

$$\frac{\partial^2 g(x)}{\partial x_j \partial x_k} = 0, \qquad j \neq k, \quad j,k=1,\ldots,M \qquad\qquad (2.4)$$

Moreover, g(x) is linear if g(x) = $\eta_0 + \sum_j \eta_j x_j$, which requires (2.4) for all j,k=1,...,M. Each restriction of (2.4) is in the form (H); joint tests of all the restrictions are discussed in Section 4.5.

Many other examples of additive derivative constraints can be derived. As with (2.1-4), specific constraints will typically involve zero restrictions on many of the functions G_0, G_j, H_{jk} and D of (H), which impart analogous simplications to the test statistics presented below.

3. The Approach to Testing Derivative Constraints

3.1 Regression Analysis of Departures

Our approach for testing (H) is based on the departure function

$$\Delta(x) = G_0(x)g(x) + \sum_j G_j(x) \frac{\partial g(x)}{\partial x_j} + \sum_{j \leq k} H_{jk}(x) \frac{\partial^2 g(x)}{\partial x_j \partial x_k} - D(x), \qquad (3.1)$$

so that (H) is summarized as $\Delta(x)=0$ for all x.

Suppose for the moment, that we observed the departure value at each data point; $\Delta(x_i)$, i=1,...,N; say up to random error. A natural method of assessing whether $\Delta(x)=0$ would be to carry out an ordinary least squares (OLS) regression analysis of $\{\Delta(x_i)\}$, to check whether $\Delta(x)$ has a nonzero mean or varies linearly or nonlinearly with x. In particular, we could estimate the coefficients of the quadratic equation

5

$$\Delta(x_i) = \gamma_c + \sum_j \gamma_{1j} x_{ji} + \sum_{j \le k} \gamma_{sjk} x_{ji} x_{ki} + u_i \qquad (3.2)$$

$$= \gamma_c + x_i' \gamma_1 + s_i' \gamma + u_i \qquad i=1,\ldots,N$$

where $s_i = (x_{1i}^2, x_{1i} x_{2i}, \ldots, x_{Mi}^2)'$ denotes the $M(M+1)/2$ vector of squared and cross product terms: γ_c, γ_1 and γ denote large sample (limits of) OLS coefficient values, and $E(u)=0$, $Cov(x,u)=0$ and $Cov(s,u)=0$ by the definition of least squares. If any of the estimates of the regression parameters of (3.2) are significantly different from zero, then there is sufficient evidence to reject the constraint (H) that $\Delta(x)=0$ for all x.

The proposal of this paper is to impliment this procedure using nonparametric estimators of the coefficients γ_c, γ_1 and γ. Using additivity of the constraint (H), Section 4 derives the estimators and a consistent estimator of their covariance matrix. The test statistic proposed is the Wald statistic of the joint hypothesis that $\gamma_c=0$, $\gamma_1=0$ and $\gamma=0$.

For later reference, the nonparametric estimators can be used to estimate the coefficients of the lower order regression equations

$$\Delta(x_i) = \alpha + u_{1i} \qquad (3.3)$$
$$i=1,\ldots,N$$

$$\Delta(x_i) = \beta_c + x_i' \beta + u_{2i} \qquad (3.4)$$

where $\alpha = E[\Delta(x)]$ and $\beta = (\Sigma_{xx})^{-1} \Sigma_{x\Delta}$, $\beta_c = E[\Delta(x)] - E(x)'\beta$ denote large sample OLS values.[7] Tests of (H) based on $\alpha=0$, $\beta_c=0$ and $\beta=0$ can be formulated, but these tests would be redundant, in view of the regression identities

$$\alpha = \gamma_c + E(x)' \gamma_1 + E(s)' \gamma$$

$$\beta_c = \gamma_c - E(x)' \Sigma_{xx}^{-1} \Sigma_{xs} \gamma + E(s)' \gamma \qquad (3.5)$$

$$\beta = \gamma_1 + \Sigma_{xx}^{-1} \Sigma_{xs} \gamma$$

But while redundant, such tests may be useful when x has many components, where estimating all of the coefficients of (3.2) is impractical.[8]

3.2 Interpretation and Power

The motivation of the procedure is that it is well founded and interpretable. The large sample coefficients γ_c, γ_1 and γ exist under weak regularity conditions, so that the procedure is not based on a maintained functional form. Moreover, when a constraint is rejected, the estimated coefficient values give an empirical depiction of how the constraint is violated in the data. Instead of just a "yes" or "no" answer, the regression test can provide useful information for revising the modeling approach or reconsidering the theory at issue.[9]

We could also compute and graph the "fitted values" of (3.2), say $\hat{\Delta}(x_i)$, $i=1,\ldots,N$, using the estimated coefficients. However, it should be noted that $\hat{\Delta}(x_i)$ is not a consistent nonparametric estimator of $\Delta(x_i)$, but rather a quadratic least squares approximation in the sense of White(1980). Testing $\hat{\Delta}(x_i)=0$ amounts to testing $\gamma_c=0$, $\gamma_1=0$ and $\gamma=0$ as above, which is not a complete (pointwise) test of $\Delta(x)=0$. Consequently, it is of interest to establish a precise characterization as to how departures of $\Delta(x)$ from 0 are reflected in the large sample regression parameters.

One such characterization arises from considering the mean departure when the data set is reweighted. Suppose that the marginal density of x were altered from $f(x)$ to $f(x|\mu)$ by (exponential family) reweighting as

$$f(x|\mu) \;=\; f(x)\, C[\pi(\mu)] e^{\pi(\mu)'x} \tag{3.6}$$

where μ refers to the reweighted mean of x. This reweighting is locally unique: $C[\pi(\mu)]=\{\int f(x)e^{\pi(\mu)'x}\,dx\}^{-1}$ is the normalizing constant determined by $\pi(\mu)$, and $\pi(\mu)$ is determined uniquely in a neighborhood of $\mu=E(x)$ by the equation $\mu = E(x|\mu) = \int x f(x) C[\pi(\mu)] e^{\pi(\mu)'x} dx.$[10] By construction, $\mu=E(x)$ corresponds to no reweighting; $\pi[E(x)]=0$, $C(0)=1$ and $f[x|E(x)] = f(x)$. The mean departure under the reweighted sample is

7

$$\phi(\mu) \;=\; E[\Delta(x)|\mu] \;=\; \int \Delta(x)\ f(x|\mu)\ dx \tag{3.7}$$

The function $\phi(\mu)$ obeys $\phi[E(x)]=E[\Delta(x)]$ and is analytic in μ in an open neighborhood of $\mu=E(x)$ (c.f. Lehmann(1959)).

The structure of $\Delta(x)$ bears an intimate relationship to $\phi(\mu)$ and its derivatives, as follows. If $\Delta(x)=0$ a.s., then obviously $\phi(\mu)=0$ for all μ. The converse is implied by completeness of the exponential family (Lehmann and Scheffe(1950-1955)), namely if $\phi(\mu)=0$ for μ in a neighborhood of $\mu=E(x)$, then $\Delta(x)=0$ a.s. Since ϕ is analytic, $\phi(\mu)=0$ in a neighborhood of $\mu=E(x)$ if and only if the derivatives of ϕ (of all orders) vanish at $\mu=E(x)$; hence the derivatives of ϕ vanish if and only if $\Delta(x)=0$ a.s.

It is the low order derivatives of ϕ at $\mu=E(x)$ that the regression coefficients of (3.2-4) measure. Clearly $\alpha=E[\Delta(x)]=\phi[E(x)]$, so that α coincides with (the zero order derivative of) ϕ at $\mu=E(x)$. From Stoker(1982,1986a), the linear coefficients β equal the first derivative $\partial\phi/\partial\mu$ evaluated at $\mu=E(x)$. Finally, the quadratic coefficients γ are uniquely connected to the matrix of second derivatives $\partial^2\phi/\partial\mu\partial\mu'$ as

<u>Theorem 3.1</u>: Under Assumptions A and B, γ is a linear, homogeneous, invertible function of $\partial^2\phi/\partial\mu\partial\mu'$ evaluated at $\mu=E(x)$, and so $\gamma=0$ if and only if $\partial^2\phi/\partial\mu\partial\mu'=0$ at $\mu=E(x)$.

Thus $\alpha=0$, $\beta=0$, $\gamma=0$ coincides uniquely with $\phi[E(x)]=0$, $\partial\phi/\partial\mu=0$ and $\partial^2\phi/\partial\mu\partial\mu'=0$, which from (3.5), coincides uniquely with $\gamma_c=0$, $\gamma_1=0$, $\gamma=0$. Consequently, the failure to reject $\gamma_c=0$, $\gamma_1=0$, $\gamma=0$ of the quadratic regression (3.2) implies that departures of $\Delta(x)$ from 0 induce at most third-order changes in the mean departure under reweighting. This is the precise characterization we were after, indicating what the coefficients measure and how they might fail to detect departures from the constraint (H).

4. Estimation of the Regression Coefficients

We now turn to the procedure for estimating the regression coefficients and testing that they vanish. The procedure is constructive but somewhat complicated and so we take it up in steps: reduce the problem to that of estimating weighted first and second average derivatives of $g(x)$ (Section 4.1), reformulate the average derivatives for estimation (Section 4.2) and propose nonparametric estimators (Section 4.3).

4.1 Reduction to Average Derivatives

Write the quadratic regression (3.2) compactly as

$$\Delta(x_i) = X_i'\Gamma + u_i \qquad i=1,\ldots,N \qquad (4.1)$$

where $X_i \equiv (1,x_i',s_i')'$ and $\Gamma = (\gamma_c, \gamma_1', \gamma')'$. Γ is written explicitly as

$$\Gamma \equiv \Pi_{XX}^{-1} C \qquad (4.2)$$

where $\Pi_{XX} = E(XX')$ and $C = E[X\Delta(x)]$.

Suppose that we had an estimator $\hat{C}$ such that $\sqrt{N}(\hat{C} - C)$ had a limiting normal distribution with mean 0 and covariance matrix V_C, and also a consistent estimator $\hat{V}_C$ of V_C. Then Γ is estimated consistently by

$$\hat{\Gamma} \equiv P_{XX}^{-1} \hat{C} \qquad (4.3)$$

where $P_{XX} = \sum_i X_i X_i'/N$, since P_{XX} is a consistent estimator of Π_{XX}. Moreover, it is a standard exercise to verify that $\sqrt{N}(\hat{\Gamma} - \Gamma)$ has a limiting normal distribution with mean 0 and covariance matrix $\Pi_{XX}^{-1}V_C\Pi_{XX}^{-1}$. For testing, under the null hypothesis that $\Gamma=0$, the Wald statistic

$$W \equiv N\hat{\Gamma}'P_{XX}\hat{V}_C^{-1}P_{XX}\hat{\Gamma} \qquad (4.4)$$

has a limiting χ^2 distribution with $1+M+[M(M+1)/2]$ degrees of freedom.

Therefore, the difficulty lies in the estimation of $C=E[X\Delta(x)]$. It is easy to see that C is the sum of terms that can be estimated with sample averages and weighted average derivative estimators, although some awkward notation is required to express this formally. In particular, partition C as

9

$C=(C_0,\{C_j\},\{C_{jk}\})'$, and write the components as

$$C_0 \quad = \quad E[\Delta(x)] \quad = \quad c_0^0 + \sum_{j'} c_0^{j'} + \sum_{j'\leq k'} c_0^{j'k'}$$

$$C_j \quad = \quad E[x_j\Delta(x)] \quad = \quad c_j^0 + \sum_{j'} c_j^{j'} + \sum_{j'\leq k'} c_j^{j'k'} \qquad j=1,\ldots,M \qquad (4.5)$$

$$C_{jk} \quad = \quad E[x_jx_k\Delta(x)] \quad = \quad c_{jk}^0 + \sum_{j'} c_{jk}^{j'} + \sum_{j'\leq k'} c_{jk}^{j'k'} \qquad j\leq k=1,\ldots,M$$

where the c's are individual components defined as follows. The zero order terms are

$$c_0^0 \quad = \quad E[G_0(x)g(x)-D(x)]$$

$$c_j^0 \quad = \quad E[x_j\ (G_0(x)g(x)-D(x))] \qquad\qquad j=1,\ldots,M \qquad (4.6)$$

$$c_{jk}^0 \quad = \quad E[x_jx_k\ (G_0(x)g(x)-D(x))] \qquad\qquad j\leq k=1,\ldots,M,$$

the first order terms are

$$c_0^{j'} \quad = \quad E\left[G_{j'}(x)\ \frac{\partial g(x)}{\partial x_{j'}}\right] \qquad\qquad j'=1,\ldots,M$$

$$c_j^{j'} \quad = \quad E\left[x_j\ \left(G_{j'}(x)\ \frac{\partial g(x)}{\partial x_{j'}}\right)\right] \qquad\qquad j,j'=1,\ldots,M \qquad (4.7)$$

$$c_{jk}^{j'} \quad = \quad E\left[x_jx_k\ \left(G_{j'}(x)\ \frac{\partial g(x)}{\partial x_{j'}}\right)\right] \qquad\qquad j\leq k,\ j'=1,\ldots,M$$

and the second order terms are

$$c_0^{j'k'} \quad = \quad E\left[H_{j'k'}(x)\ \frac{\partial^2 g(x)}{\partial x_{j'}\partial x_{k'}}\right] \qquad\qquad j'\leq k'=1,\ldots,M$$

$$c_j^{j'k'} \quad = \quad E\left[x_j\ \left(H_{j'k'}(x)\ \frac{\partial^2 g(x)}{\partial x_{j'}\partial x_{k'}}\right)\right] \qquad\qquad j,j'\leq k'=1,\ldots,M \quad (4.8)$$

$$c_{jk}^{j'k'} \quad = \quad E\left[x_jx_k\ \left(H_{j'k'}(x)\ \frac{\partial^2 g(x)}{\partial x_{j'}\partial x_{k'}}\right)\right] \qquad\qquad j\leq k,\ j'\leq k'=1,\ldots,M$$

The zero order terms (4.6) can be estimated with sample averages and the first order (4.7) and second order terms (4.8) can be estimated with weighted average first and second derivative estimators.

The next sections derive estimators for the (4.7) and (4.8) terms, as

10

well as their (joint) covariance matrix. The procedure is to construct $\hat{C}$ and $\hat{V}_C$ with these estimators, and then construct $\hat{\Gamma}$ and W via (4.3,4).

4.2 The Reformulation of Average Derivatives

We now turn to how weighted average derivative expressions can be reformulated to depend on density derivatives. This formulation motivates estimators whose properties follow from minor alterations of the statistical theory given in Powell, Stock and Stoker(1987) and Härdle and Stoker(1988). While other procedures can be proposed for estimation,[11] it is useful to note that recent work by Robinson(1987) has established the properties of such "density based" average estimators when the individual data points follow a general dependent stochastic process, an important practical situation not addressed here.

The reformulation is accomplished by integrating the weighted average derivative by parts, as

Theorem 4.1: Given Assumptions A, B and C, suppose that G(x) is a differentiable function, then

$$E\left[G(x)\,\frac{\partial g}{\partial x_j}\right] \;=\; E\left[-\left(\frac{\partial G}{\partial x_j} + \frac{G(x)}{f(x)}\,\frac{\partial f}{\partial x_j}\right) y\right] \qquad (4.9)$$

and if G(x) is twice differentiable, then

$$E\left[G(x)\frac{\partial^2 g}{\partial x_j\partial x_k}\right]=E\left[\left(\frac{\partial^2 G}{\partial x_j\partial x_k} + \frac{\partial G}{\partial x_j}\,\frac{1}{f(x)}\,\frac{\partial f}{\partial x_k} + \frac{\partial G}{\partial x_k}\,\frac{1}{f(x)}\,\frac{\partial f}{\partial x_j} + \frac{G(x)}{f(x)}\,\frac{\partial^2 f}{\partial x_j\partial x_k}\right) y\right] \quad (4.10)$$

Moreover, suppose that G(x) is an nth order differentiable function, then an nth order weighted average derivative can be expressed as

$$E\left[G(x)\,\frac{\partial^n g}{\partial x_{j_1}\cdots\,\partial x_{j_n}}\right] \;=\; E[\Psi(x)\,y] \qquad (4.11)$$

where $\Psi(x)$ is determined by G(x) and the density f(x).

Each of the terms of (4.7), (4.8) can be written in the forms (4.9), (4.10) respectively. Moreover, estimators of these terms can be constructed

11

from general estimators of the two functionals

$$\delta_{1j} = E\left[\left(\frac{G(x)}{f(x)} \frac{\partial f}{\partial x_j}\right) y\right] \tag{4.12}$$

and

$$\delta_{2jk} = E\left[\left(\frac{G(x)}{f(x)} \frac{\partial^2 f}{\partial x_j \partial x_k}\right) y\right] \tag{4.13}$$

(where G is a known function), to which we now turn.

4.3 Kernel Estimation of Average Derivatives

The estimators of (4.12) and (4.13) are sample analogues, where the density $f(x)$ and its derivatives are replaced by a nonparametric estimators. In particular, define the (Rosenblatt-Parzen) kernel density estimator[12] as

$$\hat{f}(x) = N^{-1}h^{-M} \sum_{i=1}^{N} K\left[\frac{x - x_i}{h}\right] . \tag{4.14}$$

$\hat{f}(x)$ is a local average estimator, where h is the bandwidth parameter controlling the area (window) over which averaging is performed, and $K(.)$ is a kernel function giving the weights for local averaging. $K(.)$ must be a "higher order" kernel, as discussed in Section 6. To facilitate nonparametric approximation, the asymptotic theory for $\hat{f}(x)$ requires smaller averaging windows for larger samples; $h \to 0$ as $N \to \infty$.

The derivatives of $f(x)$ are estimated by the corresponding derivatives of $\hat{f}(x)$ (see Appendix 1). In addition, because (4.12) and (4.13) each involve division by $f(x)$, we do not include terms with estimated density smaller than a bound b; a technical requirement that avoids inducing erratic behavior into the estimators.[13] For this define a trimming indicator $I_i = I[\hat{f}(x_i) > b]$, where $I[.]$ is the indicator function. For larger samples, the criterion for trimming is weakened; $b \to 0$ as $N \to \infty$.

With this setup, the functionals (4.12) and (4.13) are estimated by

$$\hat{\delta}_{1j} = N^{-1} \sum_{i=1}^{N} \left[\left(\frac{G(x_i)}{\hat{f}(x_i)} \frac{\partial \hat{f}(x_i)}{\partial x_j} \hat{I}_i \right) y_i \right] \tag{4.15}$$

and

$$\hat{\delta}_{2jk} = N^{-1} \sum_{i=1}^{N} \left[\left(\frac{G(x_i)}{\hat{f}(x_i)} \frac{\partial^2 \hat{f}(x_i)}{\partial x_j \partial x_k} \hat{I}_i \right) y_i \right] \tag{4.16}$$

The statistical properties of these estimators are given as

Theorem 4.2: Given Assumptions A, B, C and D, (where p is cited in D), if

(i) $N \to \infty$, $h \to 0$, $b \to 0$, $b^{-1} h \to 0$,

(ii) For some $\epsilon > 0$, $b^4 N^{1-\epsilon} h^{2M+4} \to \infty$,

(iii) $N h^{2p-4} \to 0$,

then a)

$$\sqrt{N}(\hat{\delta}_{1j} - \delta_{1j}) = \sqrt{N} \sum_{i=1}^{N} R_j(y_i, x_i) + o_p(1) \tag{4.17}$$

where $R_j(y,x)$ is a function with mean 0 given in Appendix 1. As a consequence, $\sqrt{N}(\hat{\delta}_{1j} - \delta_{1j})$ has a limiting normal distribution with mean 0 and variance $\sigma_j = \text{Var}(R_j)$. σ_j is consistently estimated by the sample variance of $\hat{R}_j(G, y_i, x_i)$ given in Appendix 1,

and b)

$$\sqrt{N}(\hat{\delta}_{2jk} - \delta_{2jk}) = \sqrt{N} \sum_{i=1}^{N} R_{jk}(y_i, x_i) + o_p(1) \tag{4.18}$$

where $R_{jk}(y,x)$ is a function with mean 0 given in Appendix 1. As a consequence, $\sqrt{N}(\hat{\delta}_{2jk} - \delta_{2jk})$ has a limiting normal distribution with mean 0 and variance $\sigma_{jk} = \text{Var}(R_{jk})$. σ_{jk} is consistently estimated by the sample variance of $\hat{R}_{jk}(G, y_i, x_i)$ given in Appendix 1.

The proof follows Härdle and Stoker(1988).[14]

This completes all of the ingredients of the procedure for estimating

C=E[XΔ(x)], which is summarized as follows. First estimate the density $f(x)$ and its derivatives at each point x_i, i=1,...,N. For each of the c's in (4.6), form the sample analogue estimator. For each of the c's in (4.7) and (4.8), reformulate the average derivative expressions as in (4.9) and (4.10), and form the sample analogue estimator using the estimated density. Sum the averages according to (4.5) to form $\hat{C}$. For the covariance matrix of $\hat{C}$, sum the individual components ($\hat{R}$'s) to form a grand variance component $\hat{S}_i$ for each i, and denote the sample covariance matrix of $\hat{S}_i$ as $\hat{V}_C$. The formulae corresponding to these instructions are presented in Appendix 1. We then have

Corollary 4.1: Under the conditions of Theorem 4.2, $\sqrt{N}(\hat{C} - C)$ has a limiting normal distribution with mean 0 and covariance matrix V_C. V_C is consistently estimated by $\hat{V}_C$.

The regression coefficients are then estimated by $\hat{\Gamma}$ of (4.3), with their asymptotic covariance matrix consistently estimated by $P_{XX}^{-1}\hat{V}_C P_{XX}^{-1}$. To test $\Gamma=0$, the value of the Wald statistic W of (4.4) is compared to the critical values of a χ^2 distribution with 1+M+[M(M+1)/2] degrees of freedom. This is the regression test of the derivative constraint (H).

4.4 Interpretation via Constant Returns-to-Scale

With the conceptual framework of the technique in hand, it is useful to consider some of its features relative to a concrete example. For this, we return to Example 1 specialized to testing constant returns-to-scale. The departure (3.1) takes the form

$$\Delta(x) = \sum_j \frac{\partial g(x)}{\partial x_j} - 1 \qquad (4.19)$$

so that the regression (3.2) is a quadratic approximation to the "scale" elasticity less 1.[15] The mean departure $E[\Delta(x)]$ is

14

$$E[\Delta(x)] = \sum_j E\left[\frac{\partial g(x)}{\partial x_j}\right] - 1 = \sum_j E\left[-\left(\frac{1}{f(x)}\frac{\partial f}{\partial x_j}\right)y\right] - 1 \qquad (4.20)$$

where the latter expression is the reformulation according to Theorem 4.1.

For testing constant returns-to-scale, one considers the experiment of increasing all input levels by a factor $d\theta$ (adding $d\theta$ to each x_j), and comparing the log-output response $[\sum_j \partial g/\partial x_j]d\theta$ to $d\theta$. Examining the mean departure corresponds to the experiment of increasing the input levels of all firms by $d\theta$, and comparing the mean response $\sum_j [E(\partial g/\partial x_j)]d\theta$ to $d\theta$.

The reformulation arises by recasting the last experiment as a reconfiguration of the population. After expansion of all input levels, firms at initial log-input level x will move to $x+d\theta$, so that the density of firms at $x+d\theta$ is $f(x)$. Equivalently, this experiment can be regarded as an adjustment of the density of firms at log-input level x by $[-\sum_j \partial f/\partial x_j]d\theta$. The mean log-output response from the readjustment is $\{\int [-\sum_j \partial f/\partial x_j]g(x)dx\}d\theta$ $= \sum_j E\{-[1/f(x)][\partial f/\partial x_j]g(x)\}d\theta$, or the reformulation above.[16]

In addition, this example illustrates how the form of the estimators will typically be simpler than they would appear by the formulae of Sections 4.2 and 4.3. From the estimated density $\hat{f}$, define $\hat{L}_i = [\sum_j \partial \hat{f}/\partial x_j]\hat{I}_i/\hat{f}(x_i)$. The estimator $\hat{C} = (\hat{C}_0,\{\hat{C}_j\},\{\hat{C}_{jk}\})'$ of $C = (E[\Delta(x)],\{E[x_j\Delta(x)]\},\{E[x_jx_k\Delta(x)]\})'$ is

$$\hat{C}_0 = N^{-1}\sum_{i=1}^{N}(-\hat{L}_i y_i) - 1$$

$$\hat{C}_j = N^{-1}\sum_{i=1}^{N}[-(1 + x_{ji}\hat{L}_i)y_i - x_{ji}] \qquad j \leq k = 1,\dots,M \qquad (4.21)$$

$$\hat{C}_{jk} = N^{-1}\sum_{i=1}^{N}[-(x_{ji} + x_{ki} + x_{ji}x_{ki}\hat{L}_i)y_i - x_{ji}x_{ki}]$$

which is easily verified.

Theorem 4.2 and Corollary 4.2 point out that the nonparametric estimators $\hat{C}$ and $\hat{\Gamma}$ have comparable efficiency properties to sample averages, and hence to

parameter estimators from a correctly specified model. In particular, if the true production function were in Cobb-Douglas form $g(x)=\eta_0+\sum_j\eta_j x_j$, then for larger sample sizes the precision of $\hat{\Gamma}$ improves at the same rate ($\sqrt{N}$) as that of the estimated parametric scale elasticity $\sum_j\hat{\eta}_j$. In this sense the cost of our nonparametric approach to testing is not infinite, whereas if the Cobb-Douglas specification were incorrect, testing based on $\sum_j\hat{\eta}_j$ yields uninterpretable results.

Alternatively, if $\Delta(x)$ were estimated nonparametrically using pointwise estimators of the output elasticities $(\partial g/\partial x_j)$, the precision of the pointwise estimates necessarily improve at a lower rate than that of $\hat{\Gamma}$ (c.f. Stone(1982)). This problem is exascerbated when the number of inputs M is larger than two or three, owing to the (practical and theoretical) problem of obtaining enough data points for adequate local approximation in higher dimensions; for every observed level of inputs, one must find firms whose input levels are close enough to reasonably measure all the output elasticities. Theorem 4.2 states that these problems are avoided by averaging the nonparametric components as in $\hat{C}$ and $\hat{\Gamma}$.

4.5 Extensions

The regression formulation facilitates natural methods of application for various extensions of the framework. For instance, if there are important differences in the observations (typically modeled with discrete predictor variables) then the regression estimators can be constructed for each homogeneous segment, and pooled for the constraint test. Formally, suppose that u=1,...,U denotes segments, where segment u has N_u observations and $\hat{C}_u$ denotes the estimator of C. If we define $\hat{C} = \sum_u(N_u/N)\hat{C}_u$, then (4.3) gives the pooled estimator, and the variance is estimated by an analogous combination.

Tests of constraints involving several regression functions (such as in Example 2) are obviously possible, by estimating average derivatives using the

16

corresponding dependent variable. For instance, in (2.2), estimators of average derivatives of $g^j(x)$ use observed input y^j, and those of $g^k(x)$ use y^k.

Simultaneous tests of several constraints (as in Examples 2, 3 and 4) are formulated by stacking the regressions. Let $(H^{(q)})$, $q=1,\ldots,Q$ denote (linearly independent) constraints with departure functions $\{\Delta^{(q)}(x)\}$, and let $\hat{c}^{(q)}$ denote the associated estimator of $c^{(q)}=E[X\Delta^{(q)}(x)]$. Testing $\{(H^{(q)})\}$ amounts to testing $\Gamma^Q=0$, where $\Gamma^Q = (\Gamma^{(1)'},\ldots,\Gamma^{(Q)'})'$ is the stacked vector of regression coefficients $\Gamma^{(q)} = \Pi_{XX}^{-1}c^{(q)}$. If $\hat{c}^Q = (\hat{c}^{(1)'},\ldots,\hat{c}^{(Q)'})'$, then Γ^Q is estimated as

$$\hat{\Gamma}^Q = (I_Q \otimes P_{XX}^{-1})\hat{c}^Q \qquad (4.22)$$

where I_Q is the identity matrix. The (joint) covariance matrix is estimated as before, using the sample covariance of the (stacked) variance component of $\hat{c}^Q$. The resulting Wald statistic for testing $\Gamma^Q=0$ has $Q\{1+M+[M(M+1)/2]\}$ degrees of freedom.

5. Further Remarks

We have established a nonparametric regression test of additive derivative constraints. The nonparametric statistics have precision properties comparable to those based on (correctly specified) parametric models, as their relative efficiency is finite. Practical advantages arise from the familiar regression format; easy interpretability and natural extensions to situations such as simultaneous tests of many constraints.

Several open research questions surround the practical application of the estimators. Are there automatic (sample based) rules for choosing bandwidths and/or kernel functions, that assure good small sample performance of the tests? Are there estimation methods applicable to circumstances where our assumptions fail, for instance when a significant number of observations lie on the boundary of x values? Can similar nonparametric techniques be proposed

17

for testing nonlinear derivative constraints, such as the Slutsky equation of demand analysis.

This work is part of the rapidly growing literature on semiparametric and nonparametric estimation in statistics and econometrics. It is especially "econometric" in the sense that derivative constraints have played an essential role in modeling economic behavior, both historically and practically. As such, it is the hope of the author that this work becomes regarded as an early argument for flexible approaches to questions directly motivated by economic concerns.

6. Assumptions and Technical Discusssion

The assumptions are grouped into categories as they are used in the exposition. We begin with

A. *Basic Regularity Conditions*: (y,x) is distributed with distribution T that is absolutely continuous with respect to a σ-finite measure ν, with density $F(y,x)=q(y|x)f(x)$. The support Ω of $f(x)$ is a convex subset of R^M with nonempty interior. The measure ν can be written as $\nu=\nu_y \times \nu_x$, where ν_x is Lebesgue measure. The third moments of (y,x) exist and the fourth moments of x exist. $g(x)=E(y|x)$ is twice continuously differentiable in the components of x. $E[X\Delta(x)]$ exists, and Π_{XX} is nonsingular.

These assumptions comprise the basic set-up for the constraint (H) and the quadratic regression (3.2). The vector x is continuously distributed in accordance with taking derivatives, although the side presence of discrete variables can be accomodated as discussed in Section 4.5.

B. *Reweighted Mean Departure*: The expectation $E[\Delta(x)|\mu]$ of (3.7) exists for all μ in a neighborhood of $\mu=E(x)$.

While apparently quite minimal, this assumption underlies all the

properties discussed in Section 3.2.

C. *Reformulation of Average Derivatives (Theorem 4.1):* $f(x)$, $G_j(x)f(x)$, $H_{jk}(x)f(x)$, $j \leq k=1,\ldots,M$, (and for (4.11), $G(x)f(x)$) vanish on the boundaries of their supports. $G_j(x)$, $j=1,\ldots,M$ are continously differentiable, and $f(x)$, $H_{jk}(x)$, $j \leq k=1,\ldots,M$ are twice continuously differentiable. As applied to the terms of (4.7,8), the RHS expectations of (4.9,10) exist, as do the expectations comprising (***) in the inductive step of Theorem 4.1.

These conditions eliminate boundary terms in the applications of integration by parts. The existence of expectations is assumed directly, because a more primitive condition assuring their existence is not used.

D. *Kernel Estimation (Theorem 4.2):* Let p be an integer, $p > M+4$. These assumptions are used to establish the statistical properties of $\hat{\delta}_{1j}$ and $\hat{\delta}_{2jk}$, for given j,k, and are taken to hold for each functional involved in the estimation of $C=[X\Delta(x)]$. For simplicity, let (') denote differentiation with respect to x_j, and (") denote differentiation with respect to x_j and x_k: for instance $f' = \partial f / \partial x_j$, $f'' = \partial^2 f / \partial x_j \partial x_k$. We assume

D1. All derivatives of $f(x)$ of order p exist. If $f^{(p)}$ denotes any p^{th} order derivative of f, $f^{(p)}$ is Hölder continuous: there exists r and c such that $|f^{(p)}(x+\nu)-f^{(p)}(x)| \leq c|\nu|^r$.

D2. The kernel function has support $\{u|\ |u|\leq 1\}$, is symmetric, has $p+r$ moments and $K(u)=0$, $\partial K/\partial u=0$ for all $u \in \{u|\ |u|=1\}$. $K(u)$ is of order p:

$$\int K(u)du=1$$

$$\int u^{l_1}u^{l_2}\ldots u^{l_\rho}K(u)du = 0, \qquad l_1+l_2+\ldots+l_\rho < p$$

$$\int u^{l_1}u^{l_2}\ldots u^{l_\rho}K(u)du \neq 0, \qquad l_1+l_2+\ldots+l_\rho = p$$

D3. $E[(Gg)'^2]$, $E[(Gg)''^2]$, $E[(f'Gy/f)^2]$ and $E[(f''Gy/f)^2]$ exist, and $E[(Gy)^2|x]$ is continuous in x.

D4. The following local Lipschitz conditions obtain

$$|[Gg]'(x+\nu)-[Gg]'(x)| < \omega_{g1}(x)|\nu| \qquad |[Gg]''(x+\nu)-[Gg]''(x)| < \omega_{g2}(x)|\nu|$$

$$|[f'g/f](x+\nu)-[f'g/f](x)| < \omega_{11}(x)|\nu| \qquad |[f''g/f](x+\nu)-[f''g/f](x)| < \omega_{12}(x)|\nu|$$

$$|f(x+\nu)-f(x)| < \omega_{f1}(x)|\nu| \qquad |f'(x+\nu)-f'(x)| < \omega_{f2}(x)|\nu|$$

$$|f''(x+\nu)-f''(x)| < \omega_{f3}(x)|\nu|$$

where $E[\omega_{g1}^2]$, $E[\omega_{g2}^2]$, $E[\omega_{11}^2]$, $E[\omega_{12}^2]$, $E[(Gyf'\omega_{f1}/f)^2]$, $E[(Gyf''\omega_{f1}/f)^2]$, $E[(Gy\omega_{f2})^2]$ and $E[(Gy\omega_{f3})^2]$ exist.

D5. Let $A_N=\{x|f(x)>b\}$ and $B_N\{x|f(x)\le b\}$. The following conditions hold

$$\int_{B_N} G\, g\, f'\, dx = o(N^{-1/2}) \qquad \int_{B_N} G\, g\, f''\, dx = o(N^{-1/2})$$

and for $f^{(p)}$ any p^{th} order partial derivative of f, and as $N\to\infty$ the following integrals are bounded

$$\int_{A_N} G\, g\, f^{(p)}\, dx; \qquad h^r\int_{A_N} G\, g\, dx; \qquad h^2\int_{A_N} G\, g\, (f'/f)\, f^{(p)}\, dx;$$

$$h^{r+2}\int_{A_N} G\, g\, (f'/f)\, dx; \qquad h^2\int_{A_N} G\, g\, (f''/f)\, f^{(p)}\, dx; \qquad h^{r+2}\int_{A_N} G\, g\, (f'/f)\, dx$$

These assumptions are sufficient for Theorem 4.2 as follows. D1 and D4 are smoothness conditions facilitating nonparametric approximation of the density f(x) and D3 assures that the limiting variance of the estimators exists. Assumption D2 implies that positive and negative local weights are used in averaging, which is a sufficient condition used to demonstrate that the estimators have no asymptotic bias (as, for instance, in Robinson(1988)). The value p is set so that conditions (ii) and (iii) of Theorem 4.2 hold simultaneously; if (ii) were relaxed (say by an alternative method of proof), then p could be set lower.[17] Also used in the analysis of asymptotic bias is Assumption D5, which governs the structure of the data in the tails, for

instance ruling out explosive behavior. While more primitive assumptions guaranteeing these conditions are of interest,[18] as a practical matter D5 only structures the area of data that is not used (trimmed out) in the estimators. Thus, if the low density areas of the data fail these conditions, asymptotic bias can arise in the limiting theory of the estimators.

Appendix 1: Estimator Formulae

With the kernel density estimator as in (4.14), denote the partial derivatives of the kernel function K(u) as $K^{(j)} \equiv \partial K/\partial u_j$, $K^{(jk)} \equiv \partial^2 K/\partial u_j \partial u_k$. The partial derivatives of f are then estimated as

$$\frac{\partial \hat{f}(x)}{\partial x_j} = N^{-1} h^{-M-1} \sum_{i=1}^{N} K^{(j)} \left[\frac{x - x_i}{h} \right]$$

$$\frac{\partial^2 \hat{f}(x)}{\partial x_k \partial x_j} = N^{-1} h^{-M-2} \sum_{i=1}^{N} K^{(jk)} \left[\frac{x - x_i}{h} \right]$$

where h is the bandwidth value as above.

By defining components, the estimators for δ_{1j}, δ_{2jk} and all the ingredients of $\hat{C}$ and $\hat{V}_C$ are easily stated. Begin by defining

$$T_j(G, y_i, x_i) = \left[\frac{G(x_i)}{\hat{f}(x_i)} \frac{\partial \hat{f}(x_i)}{\partial x_j} \hat{I}_i \right] y_i$$

$$T_{jk}(G, y_i, x_i) = \left[\frac{G(x_i)}{\hat{f}(x_i)} \frac{\partial^2 \hat{f}(x_i)}{\partial x_j \partial x_k} \hat{I}_i \right] y_i$$

so that

$$\hat{\delta}_{1j} = N^{-1} \sum_i T_j(G, y_i, x_i)$$

$$\hat{\delta}_{2jk} = N^{-1} \sum_i T_{jk}(G, y_i, x_i)$$

The variance components R_j and R_{jk} cited in Theorem 4.2 are

$$R_j(y, x) = \frac{G(x)}{f(x)} \frac{\partial f}{\partial x_j} [y - g(x)] + \frac{\partial [Gg]}{\partial x_j} - E\left[\frac{\partial [Gg]}{\partial x_j} \right]$$

$$R_{jk}(y, x) = \frac{G(x)}{f(x)} \frac{\partial^2 f}{\partial x_j \partial x_k} [y - g(x)] + \frac{\partial^2 [Gg]}{\partial x_j \partial x_k} - E\left[\frac{\partial^2 [Gg]}{\partial x_j \partial x_k} \right]$$

The "estimated variance components" are

$$\hat{R}_j(G,y_i,x_i) = T_j(G,y_i,x_i) - \hat{\delta}_{1j}$$

$$+ N^{-1} \sum_{i'=1}^{N} \left[h^{-M-1}K^{(j)}\left[\frac{x_i-x_{i'}}{h}\right] + h^{-M}K\left[\frac{x_i-x_{i'}}{h}\right]\left(\frac{\partial \hat{f}(x_{i'})}{\partial x_j}\right)\left(\frac{1}{\hat{f}(x_{i'})}\right)\right]\frac{G(x_{i'})y_i,\hat{I}_{i'}}{\hat{f}(x_{i'})}$$

$$\hat{R}_{jk}(G,y_i,x_i) = T_{jk}(G,y_i,x_i) - \hat{\delta}_{2jk}$$

$$+ N^{-1} \sum_{i'=1}^{N} \left[h^{-M-2}K^{(jk)}\left[\frac{x_i-x_{i'}}{h}\right] + h^{-M}K\left[\frac{x_i-x_{i'}}{h}\right]\left(\frac{\partial^2 \hat{f}(x_{i'})}{\partial x_j \partial x_k}\right)\left(\frac{1}{\hat{f}(x_{i'})}\right)\right]\frac{G(x_{i'})y_i,\hat{I}_{i'}}{\hat{f}(x_{i'})}$$

so that the sample variance of $\hat{R}_j(G,y_i,x_i)$ is a consistent estimator of $\sigma_j = \mathrm{Var}[R_j(y,x)]$, and that the sample variance of $\hat{R}_{jk}(G,y_i,x_i)$ is a consistent estimator of $\mathrm{Var}[R_{jk}(y,x)]$.

We can now state the estimators of the c's of (4.6-8), together with their estimated variance components, denoted as $\hat{s}$'s (the true variance components are omitted for simplicity). The estimators of the zero order terms of (4.6) take the form

$$\hat{c}^0[t(y,x)) = N^{-1} \sum_i t(y_i,x_i); \qquad \hat{s}^0_i = t(y_i,x_i).$$

In particular, $\hat{c}^0_0$, $\hat{s}^0_{0,i}$ set $t(y,x) = G_0(x)y-D(x)$, $\hat{c}^0_j$, $\hat{s}^0_{j,i}$ set $t(y,x) = x_j[G_0(x)y-D(x)]$, and $\hat{c}^0_{jk}$, $\hat{s}^0_{jk,i}$ set $t(y,x) = x_j x_k[G_0(x)y-D(x)]$.

The estimators of the first order terms of (4.7) take the form

$$\hat{c}^{j'}(G) = N^{-1} \sum_i \left[- \frac{\partial G(x_i)}{\partial x_{j'}} y_i - T_{j'}(G,y_i,x_i)\right]$$

$$\hat{s}^{j'}_i(G) = - \frac{\partial G(x_i)}{\partial x_{j'}} y_i - \hat{R}_{j'}(G,y_i,x_i)$$

In particular, $\hat{c}^{j'}_0$, $\hat{s}^{j'}_{0,i}$ set $G = G_{j'}$, $\hat{c}^{j'}_j$, $\hat{s}^{j'}_{j,i}$ set $G = x_j G_{j'}$, and $\hat{c}^{j'}_{jk}$, $\hat{s}^{j'}_{jk,i}$ set $G = x_j x_k G_{j'}$.

The estimators of the second order terms of (4.8) take the form

$$\hat{c}^{j'k'}(G) = N^{-1}\sum_i\left[\frac{\partial^2 G(x_i)}{\partial x_j,\partial x_{k'}}\,y_i +T_{j'}\left(\frac{\partial G}{\partial x_j},,y_i,x_i\right) +T_{k'}\left(\frac{\partial G}{\partial x_{k'}},,y_i,x_i\right) +T_{j'k'}(G,y_i,x_i)\right]$$

$$\hat{s}_0^{j'k'}(G) = \frac{\partial^2 G(x_i)}{\partial x_j,\partial x_{k'}}\,y_i + \hat{R}_{j'}\left(\frac{\partial G}{\partial x_j},,y_i,x_i\right) + \hat{R}_{k'}\left(\frac{\partial G}{\partial x_{k'}},,y_i,x_i\right) + \hat{R}_{j'k'}(G,y_i,x_i)$$

In particular, $\hat{c}_0^{j'k'}$, $s_{0,i}^{j'k'}$ set $G = H_{j'k'}$, $\hat{c}_j^{j'k'}$, $\hat{s}_{j,i}^{j'k'}$ set $G = x_j H_{j'k'}$ and $\hat{c}_{jk}^{j'k'}$, $\hat{s}_{jk,i}^{j'k'}$ set $G = x_j x_k H_{j'k'}$.

The vector $C = E[X\Delta(x)]$ is then estimated by $\hat{C} = (\hat{C}_0,\{\hat{C}_j\},\{\hat{C}_{jk}\})'$ where

$$\hat{C}_0 = \hat{c}_0^0 + \sum_{j'}\hat{c}_0^{j'} + \sum_{j'\le k'}\hat{c}_0^{j'k'}$$

$$\hat{C}_j = \hat{c}_j^0 + \sum_{j'}\hat{c}_j^{j'} + \sum_{j'\le k'}\hat{c}_j^{j'k'} \qquad j=1,\dots,M$$

$$\hat{C}_{jk} = \hat{c}_{jk}^0 + \sum_{j'}\hat{c}_j^{j} + \sum_{j'\le k'}\hat{c}_{jk}^{j'k'} \qquad j\le k=1,\dots,M$$

The overall estimated variance component $\hat{S}_i = (\hat{S}_{0,i},\{\hat{S}_{j,i}\},\{\hat{S}_{jk,i}\})'$ is found by analogous summation of individual components

$$\hat{S}_{0,i} = \hat{s}_{0,i}^0 + \sum_{j'}\hat{s}_{0,i}^{j'} + \sum_{j'\le k'}\hat{s}_{0,i}^{j'k'}$$

$$\hat{S}_{j,i} = \hat{s}_{j,i}^0 + \sum_{j'}\hat{s}_{j,i}^{j'} + \sum_{j'\le k'}\hat{s}_{j,i}^{j'k'} \qquad j=1,\dots,M$$

$$\hat{S}_{jk,i} = \hat{s}_{jk,i}^0 + \sum_{j'}\hat{s}_{jk,i}^{j'} + \sum_{j'\le k'}\hat{s}_{jk,i}^{j'k'} \qquad j\le k=1,\dots,M$$

and $\hat{V}_C$ is the sample covariance matrix of S_i.

Appendix 2: Proofs of Theorems

Proof of Theorem 3.1: Define the large sample residual functions

$$\xi = \Delta(x) - \beta_c - x'\beta \quad \text{and} \quad \zeta = s - E(s) - \Sigma_{sx}\Sigma_{xx}^{-1}[x - E(x)]$$

so that ξ, ζ are the large sample residuals of $\Delta(x)$, s regressed on x and a constant. By the definition of OLS coefficients, the quadratic coefficients γ arise from regressing ξ on ζ; namely $\gamma = \Sigma_{\zeta\zeta}^{-1}\Sigma_{\zeta\xi}$. But since $\Sigma_{\zeta\xi}=\Sigma_{s\xi}$ by construction, we have

$$\Sigma_{s\zeta} = \Sigma_{\zeta\zeta}^{-1}\gamma \qquad\qquad (*)$$

From Theorem 7 of Stoker(1982), the matrix of second derivatives of $E[\Delta(x)|\mu] = \phi(\mu)$ can be written as

$$\frac{\partial^2\delta}{\partial\mu\partial\mu'} = \Sigma_{xx}^{-1}\Sigma_{xx\Delta}\Sigma_{xx}^{-1} - \Sigma_{xx}^{-1}\Sigma_{xxx}[\Sigma_{xx}^{-1}\Sigma_{x\Delta}\otimes\Sigma_{xx}^{-1}]$$

where $\Sigma_{xx\Delta}$ is the M x M matrix with j,k element $E[(x_j-E(x_j))(x_k-E(x_k))(\Delta-E(\Delta))]$ and Σ_{xxx} is the M x M^2 matrix $\Sigma_{xxx}=[\Sigma_{1xx}|...|\Sigma_{Mxx}]$, with Σ_{1xx} the M x M matrix with j,k element $E[(x_1-E(x_1))(x_j-E(x_j))(x_k-E(x_k))]$. Simplifying,, we have

$$\frac{\partial^2\phi}{\partial\mu\partial\mu'} = \Sigma_{xx}^{-1}\Sigma_{xx\Delta}\Sigma_{xx}^{-1} - \Sigma_{xx}^{-1}[\Sigma_{xxx}(\beta\otimes I_M)]\Sigma_{xx}^{-1}$$

$$= \Sigma_{xx}^{-1}[\Sigma_{xx\Delta} - \Sigma_{xxx}(\beta\otimes I_M)]\Sigma_{xx}^{-1}$$

$$= \Sigma_{xx}^{-1}[\Sigma_{xx\xi}]\Sigma_{xx}^{-1}$$

where $\Sigma_{xx\xi}$ is the M x M matrix with j,k element $E[(x_j-E(x_j))(x_k-E(x_k))\xi] = \text{Cov}(x_j x_k,\xi)$, the latter equality from $E(x_j\xi)=0$ for all j. Thus

$$\Sigma_{xx}\frac{\partial^2\phi}{\partial\mu\partial\mu'}\Sigma_{xx} = \Sigma_{xx\xi} \qquad\qquad (**)$$

The result follows from the correspondence between (*) and (**). By (*), γ is a linear, homogeneous, nonsingular transformation of $\Sigma_{s\xi}$. $\Sigma_{xx\xi}$ is the M x M matrix uniquely constructed from the elements of the [M(M+1)/2] x 1 matrix $\Sigma_{s\xi}$. Finally, $\partial^2\phi/\partial\mu\partial\mu'$ is a linear, homogeneous, nonsingular transformation of $\Sigma_{xx\xi}$ by (**). Obviously $\partial^2\phi/\partial\mu\partial\mu'=0$ if and only if $\gamma=0$. QED

<u>Proof of Theorem 4.1</u>: (4.9) follows from integration by parts as in Stoker(1986b), using the boundary conditions of C. (4.10) follows in the same fashion from integration by parts applied twice. For (4.11), assume the result for n-1, and note that by integration by parts implies

$$E\left[G(x)\,\frac{\partial^n g(x)}{\partial x_{j_1}\ldots\partial x_{j_n}}\right] \;=\; E\left[\left(-\frac{\partial G}{\partial x_{j_n}} - G(x)\frac{\partial f}{\partial x_{j_n}}\,\frac{1}{f(x)}\right)\frac{\partial^{n-1} g(x)}{\partial x_{j_1}\ldots\partial x_{j_{n-1}}}\right] \qquad (***)$$

The result follows by induction, by applying the assumed result for n-1 to the latter term. QED

<u>Notes on the Proof of Theorem 4.2</u>: The proof structure mirrors that of Theorem 3.1 of Härdle and Stoker(1987), and so is not repeated. The key feature for $\hat{\delta}_{1j}$ is that it can be approximated by the "statistic"

$$N^{-1}\sum_{i=1}^{N}\left[\frac{G(x_i)}{f(x_i)}\,\frac{\partial f(x_i)}{\partial x_j}\right]I_i y_i \;+\; \binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{i'=i+1}^{N} P_{1j}[(x_i,y_i),(x_{i'},y_{i'})]$$

where

$$2P_{1j}[(x_i,y_i),(x_{i'},y_{i'})] = -h^{-M-1}K^{(j)}\left(\frac{x_i-x_{i'}}{h}\right)\left[\frac{G(x_i)y_i I_i}{f(x_i)} + \frac{G(x_{i'})y_{i'}I_{i'}}{f(x_{i'})}\right]$$

$$+\; h^{-M}K\left(\frac{x_i-x_{i'}}{h}\right)\left[\frac{f^{(j)}(x_i)G(x_i)y_i I_i}{f(x_i)^2} + \frac{f^{(j)}(x_{i'})G(x_{i'})y_{i'}I_{i'}}{f(x_{i'})^2}\right]$$

where $I_i=I[f(x_i)>b]$. Asymptotic normality of $\hat{\delta}_{1j}-E(\hat{\delta}_{1j})$ follows from analysis of this statistic, and $E(\hat{\delta}_{1j})-\delta_{1j}=o(N^{-1/2})$ is verified directly. Similarly, $\hat{\delta}_{2jk}-E(\hat{\delta}_{2jk})$ can be approximated by the statistic

$$N^{-1}\sum_{i=1}^{N}\left[\frac{G(x_i)}{f(x_i)}\,\frac{\partial^2 f(x_i)}{\partial x_j\partial x_k}\right]I_i y_i \;+\; \binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{i'=i+1}^{N} P_{2jk}[(x_i,y_i),(x_{i'},y_{i'})]$$

where

$$2p_{2jk}[(x_i,y_i),(x_{i'},y_{i'})] = h^{-M-2}K^{(jk)}\left[\frac{x_i-x_{i'}}{h}\right]\left[\frac{G(x_i)y_iI_i}{f(x_i)} + \frac{G(x_{i'})y_{i'}I_{i'}}{f(x_{i'})}\right]$$

$$+ h^{-M}K\left[\frac{x_i-x_{i'}}{h}\right]\left[\frac{f^{(jk)}(x_i)G(x_i)y_iI_i}{f(x_i)^2} + \frac{f^{(jk)}(x_{i'})G(x_{i'})y_{i'}I_{i'}}{f(x_{i'})^2}\right]$$

The variance components $\hat{R}_j$ and $\hat{R}_{jk}$ arise from averaging p_{1j} and p_{2jk} over

i', for each i.

[1] One reaction to this problem has been the development of "flexible" functional forms, as pioneered by Diewert(1971,1973a), Christensen, Jorgenson and Lau(1971,1973) and Sargan(1971); for more recent work, see the citations in Barnett and Lee(1985). Another reaction has been the development of nonlinear programming techniques to verify the inequality contraints implied by consistency of choice; see Afriat(1967,1972a,1972b,1973), Diewert(1973b) and Varian(1982,1983,1984b), among others.

[2] This "curse of dimensionality" of pointwise nonparametric estimators is well studied in the statistics literature (see Stone(1982) for instance), and is discussed vis-a-vis econometric applications by McFadden(1985).

[3] Proofs of results are given in Appendix 2.

[4] (H) is often implied by an analogous constraint on the underlying stochastic model. If $y=\tilde{g}(x,\epsilon)$ is the stochastic model, then if ϵ is additive or independent of x, it can be shown that $E[\partial\tilde{g}/\partial x_j|x]=\partial g/\partial x_j$ and $E[\partial^2\tilde{g}/\partial x_j\partial x_k|x]=\partial^2 g/\partial x_j\partial x_k$. Consequently, (H) is implied by the same constraint on the derivatives of $\tilde{g}$.

[5] Homogeneity restrictions can alternatively be studied in level form: namely $P(I)$ is homogeneous of degree d_0 if and only if $\sum_j I_j\partial P/\partial I_j = d_0 P(I)$, which is of the form (H) (here with g set to P and x to I).

[6] Symmetry restrictions cannot always be written in the additive form (H); for instance the traditional Slutsky conditions for demand functions have products of demand functions and derivatives of demands with respect to income.

[7] As notation, Σ_{wz} denotes the covariance matrix $E[(w-E(w))(z-E(z))']$.

[8] For instance, if M=3, (3.4) has 4 coefficients and (3.2) has 10, but if M=10, then (3.4) has 11 coefficients and (3.2) has 66.

[9] This can extend to the appropriateness of specific functional forms: see note 15 of Section 4.4.

[10] This follows directly from standard properties of the exponential family (c.f. Stoker(1982) for references). In particular, $d\mu = [Var(x|\mu)] d\pi$ and $Var(x|\mu)$ is nonsingular in a neighborhood of $\mu=E(x)$, so that the inverse function theorem gives the result.

[11]For instance, one could form a nonparametric estimator $\hat{g}(x)$ of $g(x)$ and then form the sample analogues of (4.6) and (4.7) using $\hat{g}(x_i)$, $i=1,\ldots,N$. While there is no general established theory for this procedure, certain results suggest that there is nothing to gain relative to the methods proposed here. In particular, Stoker(1988) shows that for estimating unweighted average derivatives, a "regression based" estimator is asymptotically equivalent to the "density based" estimator discussed here.

[12]For a survey of nonparametric estimators, see Prakasa Rao(1983), among others.

[13]Trimming is used for analogous reasons in Bickel(1981), Manski(1984) and Robinson(1988).

[14]See this reference for related statistical discussion. The result that $\sqrt{N}$ convergence can be achieved when nonparametric estimators are combined has been shown for other semiparametric and nonparametric estimation problems; for instance, see the references for partially linear models and linear heteroscedastic models in Robinson(1988).

[15]As a further point on interpretating the regression coefficients, note how they may be useful for guiding choice of functional form. If $g(x)$ were Cobb-Douglas, $g(x)=\eta_0 + \sum_j \eta_j x_j$, then $\gamma_c = \sum \eta_j - 1$, $\gamma_I = 0$ and $\gamma=0$, and if $g(x)$ were translog, $g(x)=\eta_0 + \sum_j \eta_j x_j + (1/2)\sum_{j,k}\eta_{jk}x_j x_k$, then $\gamma_c = \sum \eta_j - 1$, $\gamma_{lk}=\sum_j \eta_{jk}$ and $\gamma=0$. While noting that testing $\gamma_c=0$, $\gamma_I=0$, $\gamma=0$ checks similar features as tests based on these functional forms (when the forms are correct), a nonzero estimate of γ suggests the use of at least a third-order polynomial to model $g(x)$.

[16]This argument could break down if there were a significant number of firms on the boundary of log-input values; this is why it is assumed that $f(x)$ vanishes on the boundary of its support.

[17]This would be valuable for practical applications because larger p requires more positive-negative oscillation in the kernel K. However, the simulation results of Powell, Stock and Stoker(1987) suggest that better small sample performance for the estimators may be obtained when K is a standard positive kernel (i.e. a density function).

[18]For certain estimation problems using trimmed estimators, such tail conditions are not necessary; for instance, see Bickel's(1981) analysis of adaptive estimators and Robinson's(1988) estimator of coefficients in partially linear models.

References

Afriat, S.(1967), "The Construction of a Utility Function from Expenditure Data," _International Economic Review_, 8, 67-77.

Afriat,S.(1972a), "The Theory of International Comparison of Real Income and Prices," in _International Comparisons of Prices and Output_, D.J. Daly ed., New York, National Bureau of Economic Research.

Afriat, S.(1972b), "Efficiency Estimates of Production Functions," _International Economic Review_, 13, 568-598.

Afriat, S.(1973), "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method," _International Economic Review_, 14, 460-472.

Barnett, W.A. and Y.W. Lee(1985), "The Global Properties of the Minflex Laurent, Generalized Leontief, and Translog Flexible Functional Forms," _Econometrica_, 53, 1421-1437.

Bickel, P.(1981), "On Adaptive Estimation," _Annals of Statistics_, 10, 647-671.

Billingsley, P.(1979), _Probability and Measure_, New York, Wiley.

Christensen, L.R., D.W. Jorgenson and L.J. Lau(1971), "Conjugate Duality and the Transcendental Logarithmic Production Function," _Econometrica_, 39, 255-266.

Christensen, L.R., D.W. Jorgenson and L.J. Lau(1973), "Transcendental Logarithmic Production Frontiers," _Review of Economics and Statistics_, 28-45.

Diewert, W.E.(1971), "An Application of the Shepard Duality Theorem: A Generalized Leontief Production Function," _Journal of Political Economy_, 79,481-507.

Diewert, W.E.(1973a), "Functional Forms for Profit and Transformation Functions," _Journal of Economic Theory_, 6, 284-316.

Diewert, W.E.(1973b), "Afriat and Revealed Preference Theory," _Review of Economics Studies_, 40, 419-426.

Härdle, W. and T.M. Stoker(1988), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," MIT School of Management Working Paper, No. 2004-88, revised April 1988.

Lehmann, E.L.(1959), _Testing Statistical Hypotheses_, Wiley, New York.

Lehmann, E.L. and H. Scheffe(1950), "Completeness, Similar Regions and Unbiased Estimation, Part I," _Sankya_, 10, 305-340.

Lehmann, E.L. and H. Scheffe(1955), "Completeness, Similar Regions and Unbiased Estimation, Part II," _Sankya_, 15, 219-236.

McFadden, D.(1985), "Specification of Econometric Models," Presidential Address to the Fifth World Congress of the Econometric Society, Cambridge, Massachusetts, August.

Manski, C.F.(1984), Adaptive Estimation of Non-Linear Regression Models," *Econometric Reviews*, 3, 145-194.

Prakasa Rao, B.L.S.(1983), *Nonparametric Functional Estimation*, Academic Press, New York.

Powell, J.L., J.H. Stock and T.M. Stoker(1987), "Semiparametric Estimation of Index Coefficients," MIT School of Management Working Paper No. #1793-86, revised October 1987.

Robinson, P.M.(1987), "Hypothesis Testing in Semiparametric and Nonparametric Models for Economic Time Series," Discussion Paper, London School of Economics.

Robinson, P.M.(1988), "Root-N Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

Sargan, J.D.(1971), "Production Functions," in *Qualified Manpower and Economic Performance*, P.R.G. Layard, J.D. Sargan, M.E. Ager and D.J. Jones, eds., 145-204, London, Penguin Press.

Stoker, T.M.(1982), "The Use of Cross Section Data to Characterize Macro Functions," *Journal of the American Statistical Association*, 77, 369-380.

Stoker, T.M.(1986a), "Aggregation, Efficiency and Cross Section Regression," *Econometrica*, 54, 171-188.

Stoker, T.M.(1986b), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.

Stoker, T.M.(1988), "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," MIT School of Management Working Paper No. 1961-87, revised May 1988.

Stone, C.J.(1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040-1053.

Varian, H.(1982), "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945-973.

Varian, H.(1983), "Nonparametric Tests of Consumer Behavior," *Review of Economic Studies*, 50, 99-110.

Varian, H.(1984b), "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579-598.

White, H.(1980), "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 149-170.