TESTS OF DERIVATIVE CONSTRAINTS

by

Thomas M. Stoker
Massachusetts Institute of Technology

ABSTRACT

This paper proposes nonparametric regression tests of constraints involving first and second derivatives of any model $E(y|X)=F(X)$, where the true function F is unknown. The tests are based on a statistical characterization of the departures from the constraint. The test statistics are averages computed using data on y and X and knowledge of the marginal distribution of X, and their asymptotic distribution is derived. The applicability of the results is illustrated using the economic restrictions of homogeneity and symmetry, and the statistical restrictions of additivity and linearity of F in X. Extensions as well as the use of estimates for the marginal density of X are discussed.

1. Introduction

Derivative constraints play an important role in the application of econometric methods. The basic modeling restrictions implied by economic theory can often be written in the form of derivative constraints, as well as standard restrictions used to simplify econometric models. For instance, standard economic theory implies that costs are homogeneous in input prices and that demand functions are zero-degree homogeneous in prices and income, which are restrictions that can be written as constraints on the derivatives of cost and demand functions respectively. The symmetry restrictions inherent to optimization provide other examples - for instance, cost minimization implies equality constraints on the derivatives of input quantities with respect to input prices. Examples of derivative constraints not implied by basic economic theory but frequently used to simplify econometric models include constant returns-to-scale restrictions on production functions and exclusion restrictions on large demand or production systems. Such restrictions are valuable for increasing precision in estimation or facilitating applications of econometric models.[1]

Given the importance of derivative constraints, tests used to judge their statistical validity are of great interest in assessing model specification. Rejection of a constraint representing a basic implication of economic theory suggests either a revision of model specification, or reconsideration of the applicability of the theory to the specific empirical problem. The use of restrictions to simplify empirical models is only justified when the restrictions are not in conflict with the data evidence.

The major approach for testing derivative constraints in current practice is the parametric approach, whereby 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 coincide with standard hypothesis tests of the restrictions on the true parameter values. The limits of this approach concern the initially chosen parametric form, which must be held as a maintained assumption which the restrictions are tested against. The reaction to this problem has been the development of very general "flexible" functional forms, as pioneered by Diewert(1971,1973a), Christensen, Jorgenson and Lau(1971,1973) and Sargan(1971) and developed by many others, as well as sophisticated statistical techniques for implementing them in applications. Recent proposals by Gallant(1981,1982), Barnett and Jonas(1983), Barnett(1984) and Diewert and Wales(1984) display such flexible approximating properties that they often can be considered as nonparametric solutions.

Also related to tests of derivative constraints is the nonparametric approach to verifying the restrictions of optimizing behavior of Afriat(1967,1972a,1972b,1973), Diewert(1973b) and Varian(1982,1983), among others, which is based on direct verification of the inequality constraints implied by consistency of choice. This approach involves nonlinear programming techniques to check whether any consistent behavioral model could be found in accordance with observed data. When the data is in conflict with the basic inequality constraints, statistical variants of this technique can be used to produce measures of the severity of violation of the basic inequalities, as in Varian(1984b). A related approach to testing based on residual variance comparison is proposed by Epstein and Yatchew(1984), who also give a good survey of this literature.

The purpose of this paper is to propose a new nonparametric approach to testing derivative constraints, which utilizes information on the distribution of the independent variables in a behavioral equation. More formally, suppose

2

that a behavioral model explaining a dependent variable y in terms of a vector of continuous independent variables X implies that $E(y|X)=F(X)$, where the form of F is unknown. We propose tests of constraints of the form

$$(H) \qquad G_0(X)F(X) + \sum_i G_i(X) \frac{\partial F(X)}{\partial X_i} + \sum_{i,j} H_{ij}(X) \frac{\partial F^2(X)}{\partial X_i \partial X_j} = C(X)$$

where $G_0(X)$, $G_i(X)$ and $H_{ij}(X)$, $i,j=1,\ldots,M$ and $C(X)$ are known, prespecified functions of X. The tests utilize data on y and X, and require knowledge (or empirical estimates) of the density $p(X)$ of the independent variables.

The are several attractive features of the proposed tests. First, the tests are based on a statistical characterization of the departures from the derivative constraint exhibited in the data. Consequently, when a constraint is rejected, the source of rejection may be indicated by the procedure. Second, after the density $p(X)$ is characterized, the test statistics are based solely on sample averages and covariances, and therefore may be very simple to implement computationally. Third, for certain specific forms of the density $p(X)$, in particular multivariate normal, the test statistics are based on standard statistics such OLS coefficients of y regressed on X, which leads to alternative (nonparametric) interpretations of the standard statistics.

We begin by presenting the notation and basic assumptions in Section 2, together with several examples of derivative constraints of the form (H). Section 3 introduces the testing technique for the special case of a linear constraint on first derivatives, and explains the conceptual intuition of the procedure. Tests of constraints of the form (H) are presented in Section 4. Extensions of the procedure to more general constraints are discussed in Section 5. Issues and results on using statistical estimates of $p(X)$ are discussed in Section 6, and some concluding remarks are given in Section 7.

2. Notation, Examples and Basic Assumptions

We consider the situation where data is observed on a dependent variable y_k and an M-vector of independent variables $X_k = (X_{1k}, \ldots, X_{Mk})'$, for $k=1, \ldots, K$. (y_k, X_k), $k=1, \ldots, K$ represent random drawings from a distribution T which is absolutely continuous with respect to a σ-finite measure ν, with Radon-Nikodym density $P(y,X) = \partial T / \partial \nu$. $P(y,X)$ factors as $P(y,X) = q(y|X)p(X)$, where $p(X)$ is the density of the marginal distribution of X. The conditional density $q(y|X)$ represents the true behavioral econometric model, for which we assume the conditional expectation

$$(2.1) \qquad E(y|X) \equiv F(X)$$

exists for all X.

As indicated above, we propose tests of constraints of the form (H), which are nonparametric to the extent that the functional form of $F(X)$ is not prespecified or known. The characterizing feature of (H) is that it is "intrinsically" linear in $F(X)$ and its derivatives, as the coefficient functions are known. The principles upon which the tests are based are relatively straightforward, so for expositional clarity we introduce the basic technique and conceptual intuition of the tests for the special case of a linear derivative constraint of the form

$$(H^*) \qquad \sum_i c_i \frac{\partial F(X)}{\partial X_i} = c_0$$

where c_0 and c_i, $i=1, \ldots, M$ are known constants. Tests of (H^*) are covered in Section 3, and tests of (H) are covered in Section 4.

Before proceeding to specific examples, we first consider the interpretation of the derivatives $\partial F / \partial X_i$ relative to the derivatives of a more

primitive econometric model. In particular, suppose that the conditional density q(y|X) arises from a behavioral equation of the form

(2.2) $y = f(X,\varepsilon)$

where f is differentiable in X, and ε is assumed to stochastically represent individual heterogeneity not accounted for by X, with ε distributed with density $\bar{q}(\varepsilon|X)$. It is easy to see that if ε is an additive disturbance (with mean 0) in (2.2), or if ε is distributed independently of X, then $\partial F(X)/\partial X_i$ is the conditional mean of the behavioral derivatives $\partial f(X,\varepsilon)/\partial X_i$, given the value of X. Clearly, if ε represents an additive disturbance, as in $y=f(X)+\varepsilon$, then $f(X)=F(X)$ and $\partial F(X)/\partial X_i=\partial f(X,\varepsilon)/\partial X_i$ for all X. More generally, if X and ε are variation free and derivatives can be passed under expectations, we have that

(2.3) $$\frac{\partial F(X)}{\partial X_i} = E\left[\frac{\partial f(X,\varepsilon)}{\partial X_i}\bigg|X\right] + Cov\left[y, \frac{\partial \ln \bar{q}(\varepsilon|X)}{\partial X_i}\bigg|X\right]$$

(2.3) implies that $\partial F(X)/\partial X_i$ is the conditional mean of the derivative $\partial f(X,\varepsilon)/\partial X_i$ if and only if the covariance term vanishes, which is assured if ε and X are independent (since $\partial \ln \bar{q}/\partial X_i=0$ in this case). Moreover, under either sufficient condition it is easy to verify that $\partial^2 F/\partial X_i\partial X_j$ is the conditional mean of $\partial^2 f/\partial X_i\partial X_j$, $i,j=1,\ldots,M$. Consequently, under such sufficient conditions, (H) is implied by the same constraint with f replacing F, and tests of (H) coincide with tests of the same constraint on the derivatives of the primitive behavioral model f.[2]

We begin by presenting two examples of derivative constraints associated with economic properties of the function F(X), namely homogeneity (of some degree) and symmetry. For instance, demand functions derived from utility maximization are homogeneous of degree zero in prices and income, and cost

5

functions are homogeneous of degree one in input prices. In the analysis of production, it is often of interest to test whether production exhibits constant returns-to-scale, or homogeneity of degree one of output quantity with respect to input levels. Symmetry restrictions exist for virtually any model derived from optimizing behavior, such as models of input demand derived from cost minimization. These examples are included in the framework as

Example 1 - Homogeneity Restrictions: For concreteness, suppose that $F(X)$ represents the logarithm of production and X represents the vector of log-input values; input levels are $x = e^X$ and quantity produced is $\Phi(x) = e^{F(X)}$. $\Phi(x)$ is homogeneous of degree c_0 in x if $\Phi(\kappa x) = \kappa^{c_0} \Phi(x)$ for any positive scalar κ, which is valid if and only if the log-form Euler equation is valid;

$$(2.4) \qquad \sum_i \frac{\partial F(X)}{\partial X_i} = c_0$$

Here $\partial F / \partial X_i$ is the ith output elasticity, and (2.4) requires the output elasticities to add to c_0. For constant returns-to-scale we have $c_0 = 1$. (2.4) is clearly in form (H^*) where $c_i = 1$, $i = 1, \ldots, M$, and we utilize (2.4) to illustrate the results of Section 3.1. An alternative form of homogeneity constraints can be obtained from the Euler equation in level form. Specifically, suppose that $F(X)$ represents the quantity produced and X represents the vector of variable input levels. $F(X)$ is homogeneous of degree c_0 if and only if the following Euler equation is valid

$$(2.5) \qquad \sum_i X_i \frac{\partial F(X)}{\partial X_i} = c_0 F(X)$$

It should be noted that (2.4) and (2.5) involve different definitions of y and X relative to the homogeneity restriction, and will imply different tests below.

6

Example 2 - Symmetry: Suppose for concreteness that $F^i(X)$, $i=1,\ldots,M-1$, represent the demands for $M-1$ inputs, where X_i, $i=1,\ldots,M-1$ are the prices of the inputs and X_M is the output of the firm. Then cost minimization implies that

$$(2.6) \qquad \frac{\partial F^i(X)}{\partial X_j} - \frac{\partial F^j(X)}{\partial X_i} = 0 \qquad i,j=1,\ldots,M-1$$

This set of restrictions involves several behavioral equations, which are addressed in Section 5.

It should be noted that (H) does not include all symmetry restrictions of interest; for example the traditional form of the Slutsky restriction on demand functions includes products of quantities and income derivatives of other quantities, which are nonlinear terms in unknown functions.

The following two examples illustrate derivative constraints associated with the specific functional form structure of F(X).

Example 3: "X_i has no effect on y": X_i does not appear as an argument of $F(X)$ if and only if

$$(2.7) \qquad \frac{\partial F(X)}{\partial X_i} = 0$$

We will utilize (2.7) to specifically illustrate the results of Section 3.2.

Example 4 - Additivity and Linearity: $F(X)$ is additive in X_i, $i=1,\ldots,M$, if $F(X)=\sum_i F_i(X_i)$, which is equivalent to

$$(2.8) \qquad \frac{\partial^2 F(X)}{\partial X_i \partial X_j} = 0 \qquad i \neq j; \ i,j=1,\ldots,M$$

Moreover $F(X)$ is linear; $F(X)= \delta_0 + X'\delta_1$; if and only if (2.8) is valid for all $i,j=1,\ldots,M$. Each of the equality constraints in (2.8) is in the form (H); we discuss how to test them simultaneously in Section 5.

The formal assumptions that we utilize are as follows. We assume that X is continuously distributed, having carrier set Ω of the following form

Assumption 1: Ω is a ν-measurable, closed, convex subset of R^M with nonempty interior.

We will discuss the incorporation of discrete variables into the basic model in Section 6. We make the following assumptions on the conditional expectation (2.1) and on the coefficient functions of (H).

Assumption 2: F(X) is twice continuously differentiable in the components of X for all $X \in \bar{\Omega}$, where $\bar{\Omega}$ differs from Ω by at most a set of ν-measure 0.

Assumption 3: $G_i(X)$ is continuously differentiable, and $H_{ij}(X)$ is twice continuously differentiable for all $X \in \Omega$, $i,j=1,\dots,M$.

We make the following assumption on the marginal density p(X) of X.

Assumption 4: p(X) is twice continuously differentiable in the components of X for all $X \in \Omega$.

Assumption 5: For $X \in d\Omega$, where $d\Omega$ is the boundary of Ω, we have p(X)=0.

As further notation, we set

$$(2.9a) \qquad \ell_i(X) = -\frac{\partial \ln\ p(X)}{\partial X_i} \qquad i=1,\dots,M$$

$$(2.9b) \qquad \ell_{ij}(X) = -\frac{\partial^2 \ln\ p(X)}{\partial X_i \partial X_j} \qquad i,j=1,\dots,M$$

and $\ell(X)=(\ell_1(X),\dots,\ell_M(X))'$, so that $\ell(X)$ is a particular type of score vector of p.[3] We will often illustrate the results for p(X) in multivariate normal

form, as in

Example 5 - Normal Distribution: Suppose that X is distributed as a multivariate normal variable with mean μ_X^0 and covariance matrix Σ_X. Then

$$(2.10) \qquad \ell(X) = \Sigma_X^{-1}(X - \mu_X^0)$$

and $e_{ij}(X)$ is the i,j element of Σ_X^{-1}.

The basic posture of the paper is that the functions $\ell_i(X)$ and $e_{ij}(X)$ are either known (by assumption) or can be estimated, so that they can be evaluated for each X_k, k=1,......,K. For the main development of Sections 3 and 4, we assume that the functions are known, and denote their values at each X_k data value as $\ell_{ik} = \ell_i(X_k)$ and $e_{ijk} = e_{ij}(X_k)$, for i,j=1,....,M and k=1,....,K. We discuss in Section 6 the implications of utilizing estimated ℓ_i and e_{ij} functions.

We also include several regularity assumptions in Appendix 1, which, for example, assure the existence of expectations of several functions of y and X, including F(X) and its first and second derivatives. Appendix 2 contains proofs of theorems that are not presented in the exposition.

3. Tests of Linear First Derivative Constraints

In this section we derive nonparametric tests of derivative constraints of the form (H*). We first define tests based on the average departure from the constraint (H*), and then define tests based on the coefficients of departures regressed on functions of X.

3.1 Tests Based on Average Departures

We first consider the implications of directly averaging the derivatives in the constraint (H^*). To begin, define the departure $\Delta^*(X)$ from the constraint (H^*) as

$$(3.1) \qquad \Delta^*(X) \equiv \sum_i c_i \frac{\partial F(X)}{\partial X_i} - c_0$$

and the mean departure α^* as

$$(3.2) \qquad \alpha^* \equiv E(\Delta^*(X)) = \sum_i c_i E\left[\frac{\partial F(X)}{\partial X_i}\right] - c_0 = \sum_i c_i \delta_{1i} - c_0$$

where δ_{1i}, $i=1,\ldots,M$ are the mean derivatives

$$(3.3) \qquad \delta_{1i} \equiv E\left[\frac{\partial F(X)}{\partial X_i}\right] \qquad\qquad i=1,\ldots,M$$

When (H^*) is true, we clearly must have $\alpha^* = 0$, and so a test of (H^*) can be derived from a nonparametric estimate of α^*. A natural estimator can be constructed from the estimates of δ_{1i}, $i=1,\ldots,M$, that are suggested by

Theorem 1: Under Assumptions 1-5 and A1-A2, we have that

$$(3.4) \qquad \delta_{1i} = E\left[\frac{\partial F(X)}{\partial X_i}\right] = E(F(X)\ell_i(X)) = Cov(F(X),\ell_i(X)) \qquad i=1,\ldots,M$$

Proof: We begin by utilizing Fubini's Theorem (c.f. Billingsley(1979), Fleming(1977), among others) to write the expectation of the derivative $\partial F/\partial X_1$ as

$$(3.5) \quad \int_{\Omega} \frac{\partial F(X)}{\partial X_1} p(X) \, d\nu = \int \left[\int_{\omega(X_0)} \frac{\partial F(X)}{\partial X_1} p(X) d\nu_1(X_1) \right] d\nu_0(X_0)$$

where X_1 represents the first component of X and X_0 represents the other components of X. The set $\omega(X_0)$ is either a finite interval [a,b] (where a, b depend on X_0), or an infinite interval of the form [a,∞), (-∞,b] or (-∞,∞). Supposing first that $\omega(X_0)$=[a,b], integrate the inside integral of (3.5) by parts (c.f. Billingsley(1979)) as in

$$(3.6) \quad \int_{\omega(X_0)} \frac{\partial F(X)}{\partial X_1} p(X) d\nu_1(X_1) = - \int_{\omega(X_0)} F(X) \frac{\partial p(X)}{\partial X_1} d\nu_1(X_1)$$

$$+ F(b,X_0)p(b,X_0) - F(a,X_0)p(a,X_0)$$

The latter two terms represent Fp evaluated at boundary points, so that they vanish by Assumption 5. Moreover, the same is true if $\omega(X_0)$ is an infinite interval by Assumption A2 applied to limits of the boundary terms. Consequently, in all cases the RHS of (3.6) simplifies as

$$(3.7) \quad - \int_{\omega(X_0)} F(X) \frac{\partial p(X)}{\partial X_1} d\nu_1(X_1) = \int_{\omega(X_0)} F(X) \left[- \frac{\partial \ln p(X)}{\partial X_1} \right] p(X) d\nu_1(X_1)$$

$$= E(F(X)\ell_1(X))$$

$$= \text{Cov}(F(X),\ell_1(X))$$

where the latter equality holds because the mean of $\ell_1(X)$ is 0. The proof is completed by inserting (3.7) into (3.5), and repeating the same development for derivatives of F with respect to X_2, ..., X_M. QED

If we define the function $d_{11}(y,X) = y\ell_1(X)$, then Theorem 1 implies that $E(d_{11}(y,X))=\delta_{11}$. A natural estimator of δ_{11} is the sample average of the function d_{11}; or

(3.8) $\quad \hat{d}_{1i} = \dfrac{\Sigma \; y_k \ell_{ik}}{K} = \dfrac{\Sigma d_{1ik}}{K}$

where $d_{1ik} = d_{1i}(y_k, X_k)$, $i=1,\ldots,M$, $k=1,\ldots,K$. Consequently, a natural estimator of α^* is the sample average of the function $a^*(y,X) = \Sigma c_i d_{1i}(y,X) - c_0$, or

(3.9) $\quad \hat{a}^* = \dfrac{\Sigma a_k^*}{K}$

where $a_k^* = a^*(y_k, X_k) = \Sigma_i c_i d_{1ik} - c_0$, $k=1,\ldots,K$.

To present the properties of these estimators, define the vectors $c = (c_1,\ldots,c_M)'$, $\delta_1 = (\delta_{11},\ldots,\delta_{1M})'$, $d_1(y,X) = (d_{11}(y,X),\ldots,d_{1M}(y,X))'$, and $d_{1k} = d_1(y_k,X_k)$, $k=1,\ldots,K$; denote the covariance matrix of $d_1(.)$ as Σ_d, and the sample covariance matrix of $\{d_{1k}\}$ as S_d. The properties of $\hat{a}^*$ and $\hat{d}_1$ are summarized in

Theorem 2: Given Assumptions 1-5, and A1-A3, we have that $\lim \hat{d}_1 = \delta_1$ a.s., and that the limiting distribution of $\sqrt{K}(\hat{d}_1 - \delta_1)$ is normal with mean 0 and variance-covariance matrix Σ_d. Σ_d is consistently estimated by S_d. Moreover, we have that $\lim \hat{a}^* = \alpha^*$ a.s., the limiting distribution of $\sqrt{K}(\hat{a}^* - \alpha^*)$ is normal with mean 0 and variance $\Sigma_a^* = c'\Sigma_d c$, and Σ_a^* is consistently estimated by $S_a^* = c'S_d c$.

Proof: Because the data $\{y_k, \ell_k\}$ is a random sample, the consistency of $\hat{d}_1$ follows from Theorem 1 and the Strong Law of Large Numbers (c.f. Rao(1973), Section 2c.3, SLLN 2) applied to each component of $\hat{d}_1$. The consistency of S_d follows similarly from Assumption A3. The asymptotic normality of $\hat{d}_1$ follows directly from the multivariate Central Limit Theorem (c.f. Rao(1973), Section 2c.5). Finally, the properties of $\hat{a}^*$ follow immediately from the properties of $\hat{d}_1$. QED

The large sample distribution of $\sqrt{K}(\hat{a}^* - \alpha)/\sqrt{S_a^*}$ is univariate normal with mean 0 and variance 1, so that tests of $\alpha^* = 0$ can be performed using standard

methods.[4]

The underlying mathematical logic of the above estimator is straightforward. The basic idea is to use the average of the derivatives to test the constraint (H^*), as one could do by estimating α^* from performing the regression

$$(3.10) \qquad \Delta^*(X_k) = \alpha^* + u_k$$

The difficulty with performing this regression is that the individual derivatives $\partial F(X_k)/\partial X_i$ within $\Delta^*(X_k)$ are not directly observed. This problem is solved here by applying integration-by-parts, by which $\partial F(X_k)/\partial X_i$ is replaced by $d_{1k} = y_k \ell_{ik}$ for the purpose of estimating the average derivative. $\hat{d}_1$ and $\hat{a}^*$ are just the appropriate sample estimators using d_{1k}.

There is also a fairly straightforward economic logic to the above estimators, which involves reinterpreting the behavioral response represented by (H^*) as a sample reconfiguration. To see this, consider (2.4) of Example 1, where F represents a log-production function, X represents log-inputs and (H^*) represents the restriction of constant returns-to-scale ($c=\iota$, $c_0=1$). To test (H^*), one usually considers the experiment of increasing all inputs proportionately by a factor $d\theta$, or by adding $\iota d\theta$ to X. For a firm at initial log-input level X, the output response is $[\Sigma_i (\partial F/\partial X_i)]d\theta$, which is (statistically) compared to $d\theta$.

Here we consider the experiment of increasing all firm log-input levels by $\iota d\theta$, and compare the average log-output response, namely $E[\Sigma_i (\partial F/\partial X_i)]d\theta$, to $d\theta$. The test statistic derived above arises from considering the reconfiguration of the population of firms from this experiment. Namely, after expansion of inputs, all firms at initial log-input level X now have log-input level $X+\iota d\theta$, or that the density of firms (after expansion) at level $X+\iota d\theta$ is $p(X)$. Consequently, the experiment can be equivalently thought of as an

13

adjustment of the density of firms at log-input level X by $[-\Sigma_i(\partial p/\partial X_i)]'d\theta$.

The overall average log-output response is given by $[\int F(X)[-\Sigma_i(\partial p/\partial X_i)]d\nu]d\theta = E[F(X)(\Sigma_i \ell_i(X))]d\theta$, which is compared to $d\theta$. $\hat{a}^*$ just estimates this expression of the overall output response. The equivalence between behavioral response and population reconfiguration formulations would break down if there were a significant number of firms on the boundary of log-input values; we eliminate this by Assumption 5, the boundary condition.

For illustration of the specific form of the estimator $\hat{a}^*$, consider

Example 6: Consider the test of (2.4) of Example 1, where X is multivariate normally distributed, as in Example 5. From (2.10), we can write $\hat{d}_1$ as

$$(3.11) \qquad \hat{d}_1 = \Sigma_X^{-1} \left[\frac{\Sigma_k(X_k-\mu_X^0)y_k}{K} \right]$$

so that $\hat{d}_1$ is asymptotically equivalent to the OLS slope coefficients of y_k regressed on X_k.[5] $\hat{a}^* = \iota'\hat{d}_1 - 1$ is asymptotically equivalent to the sum of the OLS coefficients less 1. Notice that when y represents log-output and X the vector of log-inputs, $\hat{d}_1$ is asymptotically equivalent to the OLS coefficients from a "Cobb-Douglas" regression of y on X, although no specific functional form assumption has been applied to F(X).

3.2 Tests Based on Departure Regressions

The test proposed above is based on the fairly weak implication of (H^*) that $E(\Delta^*(X))=0$, or that (H^*) must be valid on average. In this section, we derive additional statistics which test whether $\Delta^*(X)=0$, or that (H^*) is valid for all X values. In particular, we consider statistics based on generalizations of the regression (3.10) of the following form

$$(3.12) \qquad \Delta^*(X_k) = \alpha^{**} + D(X_k)'\beta^* + u_k$$

where $D(X)=(D_1(X),\ldots,D_Q(X))'$ is a general Q vector function of X, and α^{**} and

β^* refer to the large sample limits of the OLS constant term and regression coefficients, respectively. We rewrite (3.12) as

$$(3.13) \qquad \Delta^*(X_k) = \alpha^* + (D(X_k) - \mu_D^0)'\beta^* + u_k$$

where $\mu_D^0 \equiv E(D(X))$, so that the true intercept is α^* of (3.2). If (H^*) is valid, we expect that α^* and β^* will equal 0. We first indicate how β^* can be consistently estimated with the data on y_k, ℓ_k and $D(X_k)$. We then give a concrete underpinning to this regression, and indicate the advantages of particular choices of the regressors $D(X)$, namely $D(X)=X$.

The problem as before, is that the dependent variable of the regressions (3.12,13) is not directly observed, so that OLS estimates of the coefficients of those equations could not be computed directly. However, also as before, we can solve this problem by appealing to integration-by-parts, as in Theorem 3, which is shown in Appendix 2.

Theorem 3: Under Assumptions 1-5 and A1-A2, if $D(X)$ is a continuously differentiable function of X, we have

$$(3.14) \qquad Cov(\frac{\partial F}{\partial X_i}, D_q(X)) = E(F(X)\delta_{2iq}(y,X;\mu_D^0)) = Cov(F(X),\delta_{2iq}(y,X;\mu_D^0))$$

$$i=1,\ldots,M; \quad q=1,\ldots,Q$$

where

$$(3.15) \qquad \delta_{2iq}(y,X;\mu_D^0) = \ell_i(X)\left[D_q(X) - \mu_{Dq}^0\right] - \frac{\partial D_q}{\partial X_i}$$

We can now construct an estimator of β^* from via the natural estimators of the covariances between $\partial F/\partial X_i$ and $D(X)$, $i=1,\ldots,M$, that are suggested by Theorem 3. In particular, define $d_{2iq}(y,X;\mu_D^0) = y\delta_{2iq}(y,X;\mu_D^0)$, $q=1,\ldots,Q$ and $d_{2i}(y,X;\mu_D^0)=(d_{2i1},\ldots,d_{2iQ})'$, $i=1,\ldots,M$, and denote the covariance matrix of $D(X)$ as Σ_D. Assemble the d_{2i} component terms as

$$(3.16) \qquad b^*(y,X;\mu_D^0,\Sigma_D) = \Sigma_D^{-1} \left[\sum_i c_i d_{2i}(y,X;\mu_D^0) - c_0(D(X)-\mu_D^0) \right]$$

Clearly $E(b^*) = \beta^*$, and so we define an estimator of β^* to be the sample average of the function b^*: set $b_k^* = b(y_k, X_k; \bar{D}, S_D)$, $k=1,\ldots,K$, where $\bar{D} = \Sigma D(X_k)/K$ is the sample average of $\{D(X_k)\}$ and S_D is the sample variance-covariance matrix of $\{D(X_k)\}$, and define the estimator $\hat{b}^*$ of β^* as

$$(3.17) \qquad \hat{b}^* = \frac{\Sigma b_k^*}{K}$$

For the purpose of testing (H^*), both $\hat{a}^*$ and $\hat{b}^*$ may be utilized simultaneously. For this, define the covariance matrix of $(a^*(.), b^*(.)')'$ as Σ_{ab}^* and denote the sample covariance matrix of $(a_k^*, b_k^*{}')'$, $k=1,\ldots,K$ as S_{ab}^*. Further define the covariance matrix of $(a^*(.),(b^*(.) - \Sigma_D^{-1}\alpha^* D(X)')'$ as $\Sigma_{ab(D)}^*$ and denote the sample covariance matrix of $(a_k^*,(b_k^* - S_D^{-1}\hat{a}^* D(X_k)')')'$, $k=1,\ldots,K$ as $S_{ab(D)}^*$. The properties of $(\hat{a}^*, \hat{b}^*{}')'$ can now be stated as

__Theorem 4:__ Under Assumptions 1-5 and A1-A3, we have that $\lim (\hat{a}^*, \hat{b}^*{}')' = (\alpha^*, \beta^*{}')'$ a.s. The limiting distribution of $\sqrt{K}[(\hat{a}^*, \hat{b}^*{}')' - (\alpha^*, \beta^*{}')']$ is normal with mean 0 and variance-covariance matrix $\Sigma_{ab(D)}^*$. $\Sigma_{ab(D)}^*$ is consistently estimated by $S_{ab(D)}^*$.

Asymptotic tests of (H^*) using $(\hat{a}^*, \hat{b}^*{}')'$ are possible using standard methods. In particular, a natural test statistic for (H^*) is given via

<u>Corollary 4</u>: Under the Assumptions of Theorem 4, under the null hypothesis (H^*), the limiting distribution of the statistic

$$(3.18) \qquad H^* = K \; (\hat{a}^*, \hat{b}^{*\prime}) \; S_{ab(D)}^{*-1} \begin{bmatrix} \hat{a}^* \\ \hat{b}^* \end{bmatrix}$$

is $\chi^2(1+Q)$. Moreover, $S_{ab(D)}^*$ can be replaced by S_{ab}^*.

   The following example illustrates these results, in a setting where $\alpha^* = E(\Delta^*(X)) = 0$.

<u>Example 7</u>: Consider the test of (2.7) of Example 3, where X is a univariate normal variable with mean 0 and variance σ_X^2, and the true function is $F(X) = X^2$. We have $\Delta^*(X) = \partial F/\partial X = 2X$, and $\alpha^* = E(\Delta^*(X)) = E(\partial F/\partial X) = 0$, which in view of Example 6, coincides with the fact that the large sample OLS coefficient of y on X is 0. For $D(X) = X$, from (3.13) we have that $\beta^* = 2$, which is estimated by $\hat{b}^*$. To verify the validity of (3.14), note that $\ell(X) = \sigma_X^{-2}X$, $\delta_2 = \sigma_X^{-2}X^2 - 1$, and $F(X)\delta_2 = \sigma_X^{-2}X^4 - X^2$. From the properties of the normal distribution,

$$(3.19) \qquad E(F(X)\delta_2) = \sigma_X^{-2}E(X^4) - E(X^2) = 3\sigma_X^2 - \sigma_X^2 = 2\sigma_X^2$$

so that $b^* = \sigma_X^{-2}(y\delta_2)$, with $E(b^*) = 2$.

   Theorems 3 and 4 are presented for a general differentiable function $D(X)$ to facilitate the study of a wide range of regression equations of the form (3.12,13). The choice of a particular $D(X)$ depends on the types of departures from the hypothesis (H^*) that one wants to study, because β^* is interpreted as the regression coefficients of the departures $\Delta^*(X)$ on $D(X)$. For example, to study whether a production function obeys constant returns to scale, setting $D(X) = X$ allows one to study whether returns-to-scale vary with log-input levels. However, from this point of view, the restriction that $D(X)$ be differentiable is costly, as one might want to set components of $D(X)$ equal to indicator functions (dummy variables), to see how (H^*) is violated. For

example, one is not allowed to set D(X) equal to a dummy variable indicating large versus small firms, because in that case D(X) is not differentiable.

This difficulty with utilizing discrete variables restricts the practical applicability of the regression (3.12,13), but does not affect the ability of β^* to detect general departures from (H^*). As stated briefly at the beginning of this section, the statistics of Theorem 4 and Corollary 4 test whether (H^*) is valid for differing X values, or $\Delta^*(X)=0$ for all X. To make this notion precise, the relationship between the value of β^* and the structure of $\Delta^*(X)$ is characterized along the lines of Stoker(1982,1985), as follows.

The large sample values of regression coefficients such as β^* of (3.12,13) can be characterized in terms of the changes in the mean of the departures $\Delta^*(X)$ implied by a reconfiguration of the population using weights in the exponential family form, following Stoker(1982,1985). In particular, suppose that the population density is reconfigured as $p_D^*(X|\pi)$ by setting a nonzero value of the Q vector $\pi=(\pi_1,\ldots,\pi_Q)'$ in

(3.20) $p_D^*(X|\pi) = p(X)c_D(\pi)\exp[\pi'D(X)]$

where $c_D(\pi)=(\int p(X)\exp[\pi'D(X)]d\nu)^{-1}$ is a normalizing constant. Clearly we have that $p(X)=p_D^*(X|0)$, and we consider only π values in a neighborhood $\Pi \subset R^Q$ of $\pi=0$. We also note that $p_D^*(X|\pi)$ can be equivalently parameterized by the mean $\mu_D=E[D(X)|\pi]\equiv H_D(\pi)$ as $p_D(X|\mu_D)=p_D^*(X|H_D^{-1}(\mu_D))$, where $\mu_D^0=H_D(0)$. Now consider the mean departure from (H^*) under the above population reconfiguration

(3.21) $\phi_D^*(\pi) \equiv E(\Delta^*(X)|\pi) = \sum_i c_i \phi_{Di}(\pi) - c_0$

where $\phi_{Di}(\pi)=E(\partial F/\partial X_i|\pi)$, and define $\phi_D(\mu_D)\equiv\phi_D^*(H_D^{-1}(\mu_D))$ as the implied

18

relation between the mean departure and μ_D.

As indicated in Stoker(1982,1985), the derivatives of ϕ_{Di}, $i=1,\ldots,M,$, ϕ_D^*, H_D and ϕ_D can be expressed in terms of second-order moments of the $\partial F/\partial X$, $D(X)$ distribution. In particular, we have the following theorem, which characterizes the large sample regression coefficients β^*.

__Theorem 5 (Stoker(1982,1985))__: Under Assumptions 1-5, A1-A4, we have that $\partial\phi_{Di}(0)/\partial\pi=Cov(\partial F/\partial X_i,D(X))$, $i=1,\ldots,M$, $\partial\phi_D^*(0)/\partial\pi=Cov(\Delta^*(X),D(X))$, $\partial H_D(0)/\partial\pi=\Sigma_D$ and

$$(3.22) \qquad \beta^* = \left[\frac{\partial H_D(0)}{\partial\pi}\right]^{-1} \left[\frac{\partial\phi_D^*(0)}{\partial\pi}\right] = \frac{\partial\phi_D(\mu_D^0)}{\partial\mu_D}$$

If (H^*) is valid for all $X\in\Omega$, we must have $\phi(\mu_D)=0$ for all μ_D in a neighborhood of μ_n^0. Therefore, the validity of (H^*) implies that $\beta^*=0$. However, is there any sense in which $\beta^*=0$ implies that $\Delta^*(X)=0$ for all X? The answer is given by the Lehmann-Scheffe Theorem on the completeness of the exponential family;[6]

__Theorem 6 (Lehmann and Scheffe(1950,1955))__: Under Assumptions 1-5 and A1-A4, if $Q \geq M$ and $\partial D/\partial X$ is of full rank M for all $X\in\Omega$, then $\phi_D^*(\pi)=0$ for all $\pi\in\Pi$ implies that $\Delta^*(X)=0$ a.s. for $X\in\Omega$.

Given that the variance-covariance matrix of $D(X)$ is nonsingular for all $\pi\in\Pi$, $\phi_D(\mu_D)=0$ for all $\mu_D\in H(\Pi)$ also implies that $\Delta^*(X)=0$ a.s. The rank condition is obeyed if M components of $D(X)$ can be inverted in X for all $X\in\Omega$. In particular, the condition is guaranteed if $Q=M$ and $D(X)=X$. In this case $b^*(.)$ of (3.16) can be written as

$$(3.23) \qquad b^*(y,X;\mu_X^0,\Sigma_X) = \Sigma_X^{-1}\left[\ \sum_i c_i y\left[\ell_i(X)(X-\mu_X^0) - e_i\right] - c_0(X-\mu_X^0)\right]$$

where μ_X^0 and Σ_X are the mean and variance-covariance matrix of X, and e_i is the unit vector with i^{th} component 1.

In brief, Theorem 6 says that for certain choices of D(X), e.g. D(X)=X, the aggregate functions ϕ_D^* and ϕ_D equalling 0 imply that (H^*) is valid for all $\tilde{X}\in\tilde{\Omega}$, where $\tilde{\Omega}$ differs from Ω by at most a set of measure 0. Theorems 3, 4 and 5 indicate how the first derivatives of these functions can be consistently estimated with data on y, X and the score vector $\ell(X)$. Clearly, $\alpha^*=0$ and $\beta^*=0$ are only necessary for ϕ_D^* or ϕ_D to vanish, but if $\alpha^*=0$ and $\beta^*=0$, then departures from (H^*) (nonzero values of $\Delta^*(X)$) display only second-order aggregate effects.

The practical suggestion of Theorem 6 is that the regressor function D(X) should include as M components either X or an invertible function of X, because then β^* will represent any departures from (H^*) with first-order aggregate effects. Moreover, for situations where further testing is indicated, estimates of second-order aggregate derivatives can be obtained by applying integration-by-parts to the formulae of Theorem 7 of Stoker(1982).

4. Tests of the Derivative Constraint (H)

In this section tests are developed for the general derivative constraint (H). The conceptual and mathematical features of the general tests are formally identical to the tests presented in Section 3, so that derivations are just sketched, and all proofs are relegated to Appendix 2.

We begin, as before, by defining the departure from (H) as

$$(4.1) \quad \Delta(X) = G_0(X)F(X) + \sum_i G_i(X) \frac{\partial F(X)}{\partial X_i} + \sum_{i,j} H_{ij}(X) \frac{\partial F^2(X)}{\partial X_i \partial X_j} - C(X)$$

The test statistics are based on consistent, asymptotically normal estimators of the mean $\alpha = E(\Delta(X))$ and the large sample values of the slope coefficients β of the regression

$$(4.2) \quad \Delta(X_k) = \alpha + (D(X_k) - \mu_D^0)'\beta + u_k$$

where $D(X)$ is a twice continuously differentiable Q-vector function of X. As above, when (H) is valid we have $\alpha = 0$ and $\beta = 0$. Moreover, we assemble consistent estimators of α and β from consistent estimators of the means and covariances with $D(X)$ of each of the separate terms in (H). The means and covariances of the derivative terms are expressed via

Theorem 7: Under Assumptions 1-5 and A1-A2, we have for $i,j=1,\ldots,M$ and $q=1,\ldots,Q$ that

$$(4.3a) \quad E\left[G_i(X) \frac{\partial F(X)}{\partial X_i}\right] = E(F(X)\gamma_{1i}(X)) = Cov(F(X), \gamma_{1i}(X))$$

$$(4.3b) \quad Cov\left[G_i(X) \frac{\partial F(X)}{\partial X_i}, D_q(X)\right] = E(F(X)\gamma_{2iq}(X;\mu_D^0)) = Cov(F(X), \gamma_{2iq}(X;\mu_D^0))$$

$$(4.4a) \quad E\left[H_{ij} \frac{\partial^2 F}{\partial X_i \partial X_j}\right] = E(F(X)\eta_{1ij}(X)) = Cov(F(X), \eta_{1ij}(X))$$

$$(4.4b) \quad Cov\left[H_{ij} \frac{\partial^2 F}{\partial X_i \partial X_j}, D_q(X)\right] = E(F(X)\eta_{2ijq}(X;\mu_D^0)) = Cov(F(X), \eta_{2ijq}(X;\mu_D^0))$$

where γ_{1i}, γ_{2iq}, η_{1ij} and η_{2ijq} are defined as

(4.5a) $\quad Y_{1i}(X) = G_i \ell_i - \dfrac{\partial G_i}{\partial X_i}$

(4.5b) $\quad Y_{2iq}(X;\mu_D) = Y_{1i}(X)[D_q(X) - \mu_{Dq}] - G_i \dfrac{\partial D_q}{\partial X_i}$

(4.6a) $\quad \eta_{1ij}(X) = \dfrac{\partial^2 H_{ij}}{\partial X_i \partial X_j} - \dfrac{\partial H_{ij}}{\partial X_i}\ell_j - \dfrac{\partial H_{ij}}{\partial X_j}\ell_i + H_{ij}[-\ell_{ij} + \ell_i \ell_j]$

(4.6b) $\quad \eta_{2ijq}(X;\mu_D) = \eta_{1ij}(X)[D_q(X) - \mu_{Dq}] + \left[\dfrac{\partial H_{ij}}{\partial X_j}\right]\left[\dfrac{\partial D_q}{\partial X_i}\right]$

$$+ \left[\dfrac{\partial H_{ij}}{\partial X_i}\right]\left[\dfrac{\partial D_q}{\partial X_j}\right] + H_{ij}\left[\dfrac{\partial^2 D_q}{\partial X_i \partial X_j} - \dfrac{\partial D_q}{\partial X_i}\ell_j - \dfrac{\partial D_q}{\partial X_j}\ell_i\right]$$

The usefulness of Theorem 7 arises from the fact that the Y and η functions depend only on the known G and H functions, as well as the density $p(X)$. To derive consistent estimators of α and β, we first define the following components for $i,j=1,\ldots,M$ and $q=1,\ldots,Q$

(4.7a) $\quad\quad g_{10}(y,X) = y\, G_0(X)$

(4.7b) $\quad\quad g_{1i}(y,X) = y\, Y_{1i}(X)$

(4.7c) $\quad\quad h_{1ij}(y,X) = y\, \eta_{1ij}(X)$

(4.8a) $\quad\quad g_{20q}(y,X;\mu_D) = y\, G_0(X)[D_q(X) - \mu_{Dq}]$

(4.8b) $\quad\quad g_{2iq}(y,X;\mu_D) = y\, Y_{2iq}(X;\mu_D)$

(4.8c) $\quad\quad h_{2ijq}(y,X;\mu_D) = y\, \eta_{2ijq}(X;\mu_D)$

and define the Q-vectors $g_{2i}(y,X;\mu_D)=(g_{2i1},\ldots,g_{2iQ})'$, $i=0,\ldots,M$, and $h_{2ij}(y,X;\mu_D)=(h_{2ij1},\ldots,h_{2ijQ})'$, $i,j=1,\ldots,M$. Next assemble the component terms as

$$(4.9a) \quad a(y,X) = g_{10}(y,X) + \sum_{i=1}^{M} g_{1i}(y,X) + \sum_{i,j=1}^{M} h_{1ij}(y,X) - C(X)$$

$$(4.9b) \quad b(y,X;\mu_D^0,\Sigma_D) = \Sigma_D^{-1}\left[g_{20}(y,X;\mu_D^0) + \sum_{i=1}^{M} g_{2i}(y,X;\mu_D^0) \right.$$

$$\left. + \sum_{i,j=1}^{M} h_{2ij}(y,X;\mu_D^0) - C(X)[D(X)-\mu_D^0] \right]$$

so that $E(a)=\alpha$ and $E(b)=\beta$. Now set $a_k = a(y_k, X_k)$ and $b_k = b(y_k, X_k; \bar{D}, S_D)$,
$k=1,\ldots,K$, where $\bar{D}$ and S_D are the sample average and sample covariance matrix
of $\{D(X_k)\}$, and define estimators of α and β as

$$(4.10a) \quad \hat{a} = \frac{\Sigma a_k}{K}$$

$$(4.10b) \quad \hat{b} = \frac{\Sigma b_k}{K}$$

For the purpose of testing (H), $\hat{a}$ and $\hat{b}$ can be utilized simultaneously.
To characterize their limiting distribution, define the covariance matrix of
$(a(.),b(.)')'$ as Σ_{ab} and denote the sample covariance matrix of $\{(a_k,b_k')'\}$ as
S_{ab}. Further define the covariance matrix of $(a(.),(b(.) - \Sigma_D^{-1}\alpha D(X)')')'$ as
$\Sigma_{ab(D)}$ and denote the sample covariance matrix of $(a_k,(b_k - S_D^{-1}\hat{a}D(X_k)')')'$,
$k=1,\ldots,K$ as $S_{ab(D)}$. The asymptotic properties of $(\hat{a},\hat{b}')'$ are given via

Theorem 8: Under Assumptions 1-4 and A1-A3, we have that
$\lim (\hat{a},\hat{b}')' = (\alpha,\beta')'$ a.s. The limiting distribution of $\sqrt{K}[(\hat{a},\hat{b}')' - (\alpha,\beta')']$
is normal with mean 0 and variance-covariance matrix $\Sigma_{ab(D)}$. $\Sigma_{ab(D)}$ is
consistently estimated by $S_{ab(D)}$.

As above, a χ^2 statistic for testing (H) is directly available, as in

<u>Corollary 8</u>: Under the Assumptions of Theorem 8, Under the null hypothesis (H), the limiting distribution of the statistic

$$(4.11) \qquad H = K\,(\hat{a},\hat{b}')\;S_{ab(D)}^{-1}\;\begin{bmatrix}\hat{a}\\\hat{b}\end{bmatrix}$$

is $\chi^2(1+Q)$. Moreover, $S_{ab(D)}$ can be replaced by S_{ab}.

We illustrate theorems 7 and 8 via:

Example 8: Consider the test of the second derivative constraint (2.8) for a particular (i',j'). We have $H_{ij}(X)=1$ if $(i,j)=(i',j')$, $H_{ij}(X)=0$ otherwise, $G_i(X)=0$, $G_0(X)=0$ and $C(X)=0$. We also have that $\eta_{1i'j'}=\ell_{i'}\ell_{j'}-\ell_{i'j'}$, and that $\alpha=E(\Delta(X))=E(\partial^2 F/\partial X_i,\partial X_j,)=E(y\eta_{1i'j'})$, where α is estimated by $\hat{a}$. For $D(X)=X$, we have that $\eta_{2i'j'q}=\eta_{1i'j'}(X_q-\mu_{Xq}^o)-\kappa_{qi'}\ell_{i'}-\kappa_{qj'}\ell_{j'}$, where $\kappa_{ij}=1$ if $i=j$ and $\kappa_{ij}=0$ otherwise, and that $Cov(\Delta(X),X_q)=Cov(\partial^2 F/\partial X_i,\partial X_j,,X_q)=E(y\eta_{2i'j'q})$. $\hat{b}$ estimates $\beta=\Sigma_X^{-1}[Cov(\Delta(X),X_1),\ldots,Cov(\Delta(X),X_M)]$. Finally, notice that if $p(X)$ is multivariate normal, then $\eta_{1i'j'}$ depends only on the product of deviations from means of X_i' and X_j', and that $\eta_{2i'j'q}$ depends on the product of deviations from means of X_i', X_j' and X_q.

Regarding the choice of $D(X)$, the same conclusions exist for testing constraint (H) as for testing constraint (H*); namely that when $D(X)$ contains a subvector that is an invertible transformation of X, β will represent all first-order changes in the mean of the departure $\Delta(X)$ induced by the reconfiguring the population via the exponential family density (3.20). This is verified as above, by defining the mean of the departure from (H) under the exponential family population reconfiguration as $E(\Delta(X))=\phi^*(\pi)$, recalling the interpretation of β as aggregate distributional effects (as in Theorem 5) and utilizing completeness of the exponential family (as in Theorem 6).

This completes the main development of the paper. We now turn to extensions and topics of interest to empirical applications of the results.

5. Extensions of the Testing Technique

The development of test statistics in Sections 3 and 4 is based on two constructive steps. First, estimates of the means and covariances with D(X) of each of the separate component terms in (H) are found by using integration-by-parts, where the estimates depend only on data on y and X, as well as the form of the density p(X). Second, the component estimates are assembled into statistics describing the departures from the constraint (H), from which tests are possible. Here we indicate how this technique extends to constraints involving higher order derivatives, constraints involving several different dependent variables and multi-equation constraints. The numerous extensions provide a further justification for understanding the constructive treatment of the test statistics for (H).

The general constraint (H) is limited to second-order derivatives for simplicity, since the majority of applications only involve low order derivatives. In principle, however, tests of constraints involving derivatives of any orders can be derived using the technique, as for

$$(H^1) \qquad \sum_i H_i(X) \frac{\partial^{n(i)} F(X)}{\partial X_{j_1} \ldots \partial X_{j_{n(i)}}} = C(X)$$

where $n(i)$ is any integer order and $j_1, \ldots, j_{n(i)}$ are $n(i)$ integers defining the derivatives. A test statistic for (H^1) can be derived as above using estimates of the mean and the covariance with D(X) of each component term of (H^1). Such estimates are available along precisely the same lines as before, where the estimates for nth order derivative terms will require nth order log-density derivatives, as well as derivatives of the known coefficient

functions. The form of these terms are similar, each representing a product between y and a function of X, which can be estimated via a sample average. The last property is presented formally as

Theorem 9: Given Assumptions 1-5 and A1-A2, assume that $G(X)$ is any function determined by $p(X)$ and known functions $\{H_j(X)\}$. We have that

$$(5.1a) \qquad E\left[G(X) \frac{\partial^n F(X)}{\partial X_{j_1} \ldots \partial X_{j_n}}\right] = E(F(X) \Gamma_1(X))$$

$$(5.1b) \qquad Cov\left[G(X) \frac{\partial^n F(X)}{\partial X_{j_1} \ldots \partial X_{j_n}}, D(X)\right] = E(F(X) \Gamma_2(X;\mu_D^0))$$

$$\Gamma_2(X;\mu_D^0) = - \Gamma_1(X)\mu_D^0 + \Gamma_2^*(X)$$

where $\Gamma_1(X)$ is determined by the density $p(X)$ and the known functions $\{H_j(X)\}$, and $\Gamma_2^*(X)$ is determined by $D(X)$, the density $p(X)$ and the known functions $\{H_j(X)\}$.

Theorem 9 validates the repeated application of integration-by-parts to obtain estimators of the means and covariances with $D(X)$ of general derivative terms. As before, the asymptotic structure of the average estimators implied by (5.1a,b) is straightforward, with the structure of $\Gamma_2(X;\mu_D^0)$ of (5.1b) providing a simple asymptotic variance correction when $\bar{D}$ is used in place of μ_D^0, as in Theorems 4 and 8.

The second extension is to derivative constraints among several behavioral equations, or equations describing several dependent variables. As long as the constraints are intrinsically linear in derivatives, with known coefficient functions, the technique can be applied directly. In particular, suppose that we observe two dependent variables, y^1 and y^2, with conditional expectations $E(y^1|X) \equiv F^1(X)$ and $E^2(y^2|X) \equiv F^2(X)$, and our interest is in testing

a constraint of the form

$$(H^2) \quad \sum_i H_i^1(X) \frac{\partial^{n^1(i)} F^1(X)}{\partial X_{j_1} \cdots \partial X_{j_{n^1(i)}}} + \sum_i H_i^2(X) \frac{\partial^{n^2(i)} F^2(X)}{\partial X_{j_1} \cdots \partial X_{j_{n^2(i)}}} = C(X)$$

where the coefficient functions $H_i^1(X)$, $H_i^2(X)$ are known functions of X. Our results apply directly to estimating the mean and covariance with D(X) of each of the separate terms in (H^2), and hence permit estimation of the regression coefficients of departures from (H^2) on D(X). The method of assembly of mean departure and regression coefficient estimates follows through exactly, as does the (sample covariance) method of estimating the covariance matrix of the estimates.

Example 9: Consider the test of the restriction (2.6), where y_k^i denotes the observed value of the ith input quantity, $i=1,\ldots,M-1$. An estimate $\hat{a}$ of $\alpha=E(\omega_i\hat{x}_j)-\hat{\omega}(\upsilon F^i/\iota x_j - \omega F^i/\sigma X_i)$ is defined by applying Theorem 1 to each term, yielding

$$(5.2) \qquad \hat{a} = \frac{\sum_k y_k^i \ell_{jk}}{K} - \frac{\sum_k y_k^j \ell_{ik}}{K}$$

Similarly, an estimate of $\beta = \Sigma_X^{-1} \text{Cov}(\partial F^i/\partial X_j - \partial F^j/\partial X_i, D(X))$ can be constructed in a component-by-component manner.

The third extension is to testing economic hypotheses that take the form of several (simultaneous) constraints, each in the form (H), (H^*), (H^1) or (H^2). We have discussed how each constraint can be tested individually; the only issue that remains is how to test them simultaneously. As above, we can obtain consistent, asymptotically normal estimates of the mean and regression coefficients of departures from each constraint, with each estimator in the form of an average. The joint distribution of all mean departure and regression coefficient estimates for all constraints is asymptotically normal,

and the joint covariance matrix can be estimated using sample covariances (modified for $\bar{D}$) of all components of the average statistics. A single grand $\chi^2(s(1+Q))$ statistic can be constructed (where s is the number of independent constraints), which tests whether the mean and regression coefficients of the departures from each constraint vanish simultaneously. In other words, the construction of Sections 3 and 4 applies in estimating the multivariate mean departure of several constraints, as well as the multivariate matrix of regression coefficients on D(X).

For these extensions, all of the previous development carries through, including the advantages of utilizing X (or an invertible transformation of X) as a subvector of D(X). The specific formulae for these extensions are derived in a straightforward fashion, and therefore are left for applications.

6. Characterization of the Distribution of X

In this section we consider several topics relevant to the empirical implimentation of the above tests; the incorporation of additional (discrete) independent variables and the estimation of the density p(X).

Very often derivative constraints of interest in an application involve only the effects of a subset of the X variables. Expand the notation slightly by supposing that there are two sets of X variables, namely an M_1-vector X_1 and an M_2-vector X_2, where the behavioral model has $E(y|X_1,X_2)=F(X_1,X_2)$. For the production example where y represents log-output, suppose that X_1 represents log-input values and X_2 represents additional technological variables affecting production. The hypothesis of constant returns-to-scale obviously involves only the derivatives of F with respect to X_1, with X_2 appearing as additional variables.

The techniques of the paper apply directly to this situation where $X=(X_1',X_2')'$, and the functions $\ell_i(X)$ and $e_{ij}(X)$ of (2.9a,b) are defined as

the derivatives with respect to the components of X_1. This is valid because the use of integration-by-parts in Theorems 1, 3, 7 and 9 isolates on derivatives with respect to single components of X, and therefore can be applied to the derivatives of F with respect to the components of X_1 for each value of X_2. There are three useful observations in this regard. First, the validity of a constraint on the derivatives of F with respect to X_1 is tested for all values of X_2; in the production example, constant returns-to-scale is tested for each value of the technological variables X_2. Second, the functions $\ell_i(X)$ and $e_{ij}(X)$ will in general depend on both X_1 and X_2; so that the joint distribution of X_1 and X_2 must be characterized, not just the (marginal) distribution of X_1 (unless X_1 and X_2 are independent, in which case X_2 has an analogous role to ϵ of Section 2). Third, the requirements of Assumptions 1-5 need only apply to X_1, or in particular that X_2 may represent discrete variables, or variables in which p and/or F are not differentiable.

Returning to the original notation, the results of this paper have taken the functions $\ell_i(X)$ and $e_{ij}(X)$ to be known, so that their values could be computed for each observation value X_k, $k=1,\ldots,K$. In certain cases, such as X normally distributed, the above test statistics can be written in terms of OLS estimators and other standard statistics (as in Examples 6-8); in such cases the framework can be applied directly without empirical characterization of $p(X)$. Realistically, however, general applications of the technique will require that the density $p(X)$ be characterized with the data $\{X_k, k=1,\ldots,K\}$. We now discuss the econometrics of using estimated values of the functions $\ell_i(X)$ and $e_{ij}(X)$.

Statistical results that are analogous to Theorems 4 and 8 can be obtained when a parametric approach is adopted for estimating the density $p(X)$. In particular, suppose that $p(X)$ is assumed to be in the parametric form $p(X|\Lambda)$, where Λ represents a finite vector of parameters with true value Λ^0,

or that $p(X|\Lambda^0)=p(X)$. If a $\sqrt{K}$ consistent estimate $\hat{\Lambda}$ of Λ^0 is available, then estimates $\hat{\ell}_{ik}=\ell_i(X_k,\hat{\Lambda})$ and $\hat{e}_{ijk}=e_{ij}(X_k,\hat{\Lambda})$ can be constructed for all

$i,j=1,\ldots,M$ and $k=1,\ldots,K$, and utilized in place of ℓ_{ik} and e_{ijk} in the test

statistics. Under the additional regularity conditions of Appendix 1, tests

are possible using estimated values, where the only proviso is that the

asymptotic variances utilized must reflect the estimated $\hat{\Lambda}$.

More formally, with regard to Section 4, suppose that $a(.)$ and $b(.)$ of

(4.9a,b) are written to include the parameters Λ, denote $\tilde{a}_k=a(y_k,X_k;\hat{\Lambda})$,

$\tilde{b}_k=b(y_k,X_k;\bar{D},S_D,\hat{\Lambda})$ as the components evaluated at $\hat{\Lambda}$, and define $\tilde{a}$ and $\tilde{b}$ as the

averages of $\{\tilde{a}_k\}$ and $\{\tilde{b}_k\}$ respectively. Following Newey(1984), for the

variance matrices $\Sigma_{ab(D,\lambda)}$ and $S_{ab(D,\lambda)}$ defined in Appendix 1, we can show

Theorem 10: Under Assumptions 1-5 and A1-A3, A5-A6, we have that
$\lim (\tilde{a},\tilde{b}')' = (\alpha,\beta')'$ a.s., and that the limiting distribution of
$\sqrt{K}[(\tilde{a},\tilde{b}')' - (\alpha,\beta')']$ is normal with mean 0 and variance-covariance matrix
$\Sigma_{ab(D,\lambda)} \cdot \Sigma_{ab(D,\lambda)}$ is consistently estimated by $S_{ab(D,\lambda)}$.

Therefore, tests of constraints can be performed with density parameter

estimates, as long as the variance estimates reflect the variability of $\hat{\Lambda}$.

Of natural importance is the question of whether using nonparametric

estimates of the functions $\ell_i(X)$ and $e_{ij}(X)$ in the test statistics will yield

consistent procedures, with $\sqrt{K}$ convergence to normality. This question is

difficult, unanswered, and beyond the scope of this paper. Here we indicate

two lines of work which may provide solutions to this problem. The first

approach is to estimate $\ell_i(X)$ and $e_{ij}(X)$ by differentiating a kernel density

estimator of $p(X)$. The current state of the literature, as surveyed by Prakasa

Rao (1983, see especially chapter 4), does not cover the use of density

estimates in a multivariate context as required here, although this is a

promising avenue for future research.[7]

A second approach is presented in Gallant and Nychka(1985), whose propose a flexible functional form approach to estimating p(X) by postulating that

(6.1) $\hat{p}(X) = p(X;\theta_K)N(X)$

where $N(X)$ is the multivariate normal distribution[8] and $p(X;\theta_K)$ is a (Hermite) polynomial, with parameters θ_K, whose degree is increased with the sample size K. Gallant and Nychka indicate that when the parameters θ_K are estimated via maximum likelihood, $\hat{p}(X)$ provides a consistent nonparametric estimate of p(X). Moreover, they show that differentiating ln $\hat{p}(X)$ gives consistent estimates of $\ell_i(X)$, and that the sample covariances between the data values y_k and the estimated ℓ_{ik} consistently estimate the population covariances. Thus, Gallant and Nychka provide a theoretical solution to the problem of consistently estimating α^* of Section 3 with nonparametric density estimates, and it is a natural conjecture that α and β of Section 4 can likewise be consistently estimated. The remaining open question is whether $\sqrt{K}$ consistent nonparametric estimators are available.

7. Concluding Remarks

This paper has proposed a new nonparametric technique for testing derivative constraints. The technique utilizes information of the density of the independent variables, which is related mathematically to the unknown derivatives by integration-by-parts. Tests statistics are constructed for constraints in the general form (H), with extensions outlined for higher-order derivative constraints, multi-equation constraints and constraints involving derivatives of several unknown behavioral functions.

There are several advantages of the test statistics. First, the statistics are based on a statistical characterization of the departures from the constraint exhibited in the data. This means that if a constraint is

31

rejected, one can study the estimates of α and β to learn how the data departs from the constraint. In particular, a nonzero value of α indicates a nonzero average departure from the tested constraint, and nonzero values of the components of β indicate how departures from the constraint vary with the components of the chosen regressor vector $D(X)$. When $D(X)$ includes an M-subvector which is an invertible function of X, $\alpha=0$ and $\beta=0$ implies that departures display at most second-order aggregate effects.

The second advantage is the computational simplicity of the test statistics, after the density of independent variables has been characterized. Given (parametric) estimates of the log-density derivative functions $\ell_i(X)$ and $\ell_{ij}(X)$, all estimators are constructed from sample averages, involving no sophisticated nonlinear programming or other complicated maximization techniques. This feature computationally facilitates the study of many derivative constraints for a given data set. Moreover, certain density assumptions, such as multivariate normality, imply that the tests statistics are naturally related to familiar statistics, such as OLS regression coefficients.

The main limitation of the technique arises from the lack of results on the use of nonparametric estimates of the log-density derivatives. While consistency of such test statistics has been established, the question of whether $\sqrt{K}$ consistent tests can be performed with nonparametric estimates of $p(X)$ remains open. This question may be resolved in the near future, because of the currently very active pursuit of related questions of nonparametric density estimation.

The second limitation of the testing technique is due to the intrinsic linearity of all of the constraints considered. We have not established tests for constraints involving products of derivatives of unknown functions (e.g. terms of the form $(\partial F/\partial X_i)(\partial F/\partial X_j)$) or products of derivatives and levels of

32

unknown functions (e.g. terms of the form $F^1(X)(\partial F^2/\partial X_i)$). This eliminates some economic hypotheses of interest, such as the traditional form of the Slutsky equations in demand analysis (which contains products of quantities of goods and the income derivatives of other goods). Further research is warranted to see whether an analogous testing technique can be applied to nonlinear derivative constraints.

APPENDIX 1: FURTHER ASSUMPTIONS

We begin by assuming that the expectations of individual terms in the derivative constraints (H^*), (H), and (H^1). Since a primitive condition assuring existence is not of further use here, we assume the existence directly via

Assumption A1: The expectations in the following formulae exist and are finite

1. (3.4)		(Theorems 1, 2)
2. (3.14)		(Theorems 3, 4)
3. (4.3a,b), (4.4a,b), $yG_0(X)$, $C(X)$		(Theorems 7, 8)
4. (5.1a,b)		(Theorem 9)

The results listed in parentheses indicate where the condition is required.

Many of the results utilize integration-by-parts to write expectations of derivatives as the sum of a covariance term and boundary terms. For unbounded carrier sets, we require an assumption that implies that (limits of) the boundary terms vanish. A generic condition which is sufficient for this property is given as follows. Define a single component sequence $\{X^n \in R^M\}$ via $X^n = (X_1, \ldots, X_i^n, \ldots, X_M)$ for some component i; so that $\{X^n\}$ is a set of points that differ with respect to only a single component. A function $G(X)$ obeys condition A if

Condition A: If $\{X^n | X^n \in \Omega\}$ is any single component sequence such that $\|X^n\| \to \infty$ as $n \to \infty$, then $G(X^n)p(X^n) \to 0$.

We require that condition A is obeyed by several functions, as in

Assumption A2: The following functions obey condition A, for all $i,j = 1, \ldots, M$, $q = 1, \ldots, Q$.

1. $F(X)$		(Theorems 1, 2)
2. $F(X)D_q(X)$		(Theorems 3, 4)
3. $F(X)G_i(X)$		(Theorems 7, 8)
4. $[\partial F/\partial X_i]H_{ij}(X)$		(Theorems 7, 8)
5. $F(X)[\partial H_{ij}/\partial X_j - H_{ij}\ell_j]$		(Theorems 7, 8)
6. $F(X)G_i(X)D_q(X)$		(Theorems 7, 8)

7. $[\partial F/\partial X_i]H_{ij}(X)D_q(X)$ (Theorems 7, 8)

8. $F(X)\{D_q(X)[\partial H_{ij}/\partial X_j D-H_{ij}\ell_j]+H_{ij}\partial D_q/\partial X_j\}$

                                        (Theorems 7, 8)

9. $G(X)[\partial^{n-1}F/\partial X_{j1},\ldots,\partial X_{jn-1}]$ (n to n-1 Step of Theorem 9)

10. $D_q(X)G(X)[\partial^{n-1}F/\partial X_{j1},\ldots,\partial X_{jn-1}]]$ (n to n-1 Step of Theorem 9)

All of the estimators discussed in Sections 3 and 4 are sample averages, and the asymptotic properties are established by appealing to the Strong Law of Large Numbers and the Central Limit Theorem. To apply these theorems, we require that the means and variance-covariance matrices of the components of the sample averages exist. Al provides the existence of the means. For the variance-covariance matrices we assume

__Assumption A3__: The variance-covariance matrices of the following functions exist under the density $p(X)$ and are positive definite:

1. $d_j(y,X)$ (Theorem 2)
2. $D(X)$ (Theorems 4, 5)
3. $a^*(y,X)$, $b^*(y,X;\mu_D^0,\Sigma_D)$ (Theorem 4)
4. $a(y,X)$, $b(y,X;\mu_D^0,\Sigma_D)$ (Theorem 8)

For utilizing completeness of the exponential family (3.20), we assume

__Assumption A4__: The expectation $E(\Delta^*(X)|\pi)$ defined using $p(X|\pi)$ of (3.20) exists for all $\pi\in\Pi$, where Π is a convex subset of R^Q containing an open neighborhood of $\pi=0$.

The closing remarks of Section 4 require the analogous assumption for $\Delta(X)$.

For the parametric characterization $p(X|\Lambda)$, we assume the existence of a CUAN estimator $\hat{\Lambda}$ obeying

__Assumption A5__: The estimator $\hat{\Lambda}$ can be written in the form

$$\hat{\Lambda} = \Lambda^0 + \frac{\Sigma_k \lambda(X_k;\Lambda^0)}{K} + o_p(1/\sqrt{K})$$

where $E(\lambda(X;\Lambda^0))=0$ and $Var(\lambda(X;\Lambda^0))=\Sigma_{\lambda\lambda}$ exists. The covariance between any two components of $\lambda(.)$, $a(.)$, $b(.)$ and $D(.)$ exists.

For establishing the consistency and asymptotic normality of $\tilde{a}$ and $\tilde{b}$, we require

Assumption A6: $a(.)$ and $b(.)$ are differentiable with respect to the components of Λ, and the derivatives have finite expectations. For each of the terms $|\lambda_i(.)\lambda_j(.)|$, $|a(.)|$, $|b_q(.)|$, $|a(.)\lambda_i(.)|$, $|b_q(.)\lambda_i(.)|$, $|[\partial a/\partial\Lambda]\lambda_i(.)|$ and $|[\partial b_q/\partial\Lambda]\lambda_i(.)|$, there exists a function $T^*(y,X)$ which is an upper bound for all Λ in an open neighborhood of Λ^0, such that the $1+\tau$ moment of T^* exists.

Finally, denote $A=E[\partial a/\partial\Lambda^0]$, $B=E[\partial b/\partial\Lambda^0]$, $\Sigma_{ab(D,\lambda)}$ as the covariance matrix of $(a(.)+A\lambda(.),(b(.)-\alpha D(X)+B\lambda(.))')'$ and $S_{ab(D,\lambda)}$ as the sample covariance matrix of $(\tilde{a}_k+\hat{A}\hat{\lambda}_k,(\tilde{b}_k-\hat{a}D(X_k)+\hat{B}\hat{\lambda}_k)')'$, where A, B and λ are equal to the corresponding functions evaluated at Λ.

Proof of Theorem 3: The proof parallels that of Theorem 1, where we utilize the representation $\text{Cov}(\partial F/\partial X_1, D_q(X)) = E[(\partial F/\partial X_1)(D_q(X) - \mu_{Dq}^0)]$. We begin by applying Fubini's Theorem for $i=1$ as

$$(A.1) \qquad \int_\Omega \frac{\partial F(X)}{\partial X_1} [D_q(X) - \mu_{Dq}^0] p(X) d\nu$$

$$= \int \left[\int_{\omega(X_0)} \frac{\partial F(X)}{\partial X_1} [D_q(X) - \mu_{Dq}^0] p(X) d\nu_1(X_1) \right] d\nu_0(X_0)$$

where, as in the proof of Theorem 1, $X = (X_1, X_0)$, with X_0 denoting the other components of X. Suppose first that $\omega(X_0) = [a,b]$, a bounded interval, and integrate the inside integral of (A.1) by parts as in

$$(A.2) \qquad \int_{\omega(X_0)} \frac{\partial F(X)}{\partial X_1} [D_q(X) - \mu_{Dq}^0] p(X) d\nu_1(X_1)$$

$$= - \int_{\omega(X_0)} F(X) \left[\frac{\partial \ln p(X)}{\partial X_1} [D_q(X) - \mu_{Dq}^0] + \frac{\partial D_q}{\partial X_1} \right] p(X) d\nu_1(X_1)$$

$$+ F(b, X_0)[D_q(b, X_0) - \mu_{Dq}^0] p(b, X_0) - F(a, X_0)[D_q(a, X_0) - \mu_{Dq}^0] p(a, X_0)$$

$$= \int_{\omega(X_0)} F(X) \left[\ell_1(X)[D_q(X) - \mu_{Dq}^0] - \frac{\partial D_q}{\partial X_1} \right] p(X) d\nu_1(X_1)$$

where the latter equality holds by Assumption 5. (A.2) is also valid when $\omega(X_0)$ is an unbounded interval by Assumption A2. Now, substitute (A.2) back into (A.1) to get

$$(A.3) \qquad E\left[\frac{\partial F}{\partial X_1} [D_q(X) - \mu_{Dq}^0] \right] = \int_\Omega F(X) \left[\ell_1(X)[D_q(X) - \mu_{Dq}^0] - \frac{\partial D_q}{\partial X_1} \right] p(X) d\nu$$

$$= E(F(X) \delta_{2iq}(y, X; \mu_D^0))$$

The covariance representation for $i=1$ follows from

$$(A.4) \quad \int \ell_1(X)[D_q(X)-\mu^0_{Dq}]p(X)d\nu = \int \frac{\partial D_q}{\partial X_1} p(X)d\nu$$

which again follows from integration-by-parts. The cases for $i=2,\ldots,M$ are identical. QED

Proof of Theorem 4: Consistency and asymptotic normality follow directly from the Strong Law of Large Numbers and the Central Limit Theorem. For the expression of the variance-covariance matrix, it is easy to see that $\sqrt{K}(\hat{b}^* - \beta^*)$ can be written as

$$(A.5) \quad \sqrt{K}(\hat{b}^* - \beta^*) = \Sigma_D^{-1} \sqrt{K}\left[\frac{\Sigma b^*(y_k,X_k;\mu_D,\Sigma_D)}{K} - \beta^*\right]$$

$$- \Sigma_D^{-1}\alpha^* \sqrt{K}(\bar{D} - \mu_D) + o_p(1)$$

so that $\Sigma^*_{ab(D)}$ is the limiting covariance matrix of $(\hat{a}^*,\hat{b}^{*\prime})'$. Consistency of $S^*_{ab(D)}$ follows as in Theorem 2, noting the standard properties of continuous functions of consistent estimators. QED

Proof Sketch of Theorem 7: All of the results are shown by applications of integration-by-parts, where all boundary terms vanish by Assumptions 4 and A2. Here we sketch the verification of (4.4a), as follows

$$(A.6) \quad E\left[H_{ij} \frac{\partial^2 F}{\partial X_i \partial X_j}\right] = \int H_{ij} \frac{\partial^2 F}{\partial X_i \partial X_j} p(X) d\nu$$

$$= - \int \frac{\partial F}{\partial X_i}\left[\frac{\partial H_{ij}}{\partial X_j} - H_{ij}\ell_j\right] p(X) d\nu$$

$$= \int F \left[\frac{\partial^2 H_{ij}}{\partial X_i \partial X_j} - \frac{\partial H_{ij}}{\partial X_i}\ell_j - \frac{\partial H_{ij}}{\partial X_j}\ell_i + H_{ij}[-\ell_{ij} + \ell_i\ell_j]\right] p(X)d\nu$$

$$= E(F(X)\eta_{1ij}(X))$$

where the second equality follows from integration-by-parts with respect to X_j and the third equality follows by integration-by-parts with respect to X_i. The covariance representation follows from $E(\eta_{1ij}(X))=0$, which is also verified by

integration-by-parts. QED

The proof of Theorem 8 is analogous to the proof of Theorem 4.

Proof of Theorem 9: Theorem 7 shows the result for n=1,2. (5.1a) is shown by induction, where we first assume the result for n-1, and show its validity for n. In particular, we have

$$(A.7) \quad \int G(X) \frac{\partial^n F(X)}{\partial X_{j_1} \dots \partial X_{j_n}} \, p(X)dv$$

$$= \int \frac{\partial^{n-1} F(X)}{\partial X_{j_1} \dots \partial X_{j_{n-1}}} \left[- \frac{\partial G}{\partial X_{j_n}} + G \, \ell_{j_n} \right] p(X)dv$$

where the boundary terms vanish from Assumptions 5 and A2 as before. The latter integral represents an expectation in the form (5.1) for derivatives of order n-1, for which the result is assumed. Consequently, by induction, the result is true for all positive n. (5.1b) can easily be verified, given (5.1a). QED

Proof of Theorem 10: Consistency follows from the consistency of $\hat{\Lambda}$, by standard arguments (among many others, see Stoker(1985), Theorem 7). Asymptotic normality follows from Assumptions A5-A6 in accordance with the expansions

$$(A.8) \quad \sqrt{K}(\tilde{a} - \alpha) = \frac{\Sigma(a(y_k, X_k; \Lambda^0) - \alpha)}{\sqrt{K}} + \left[E\left[\frac{\partial a}{\partial \Lambda}\right]' \right]\sqrt{K}(\hat{\Lambda} - \Lambda^0) + o_p(1)$$

$$(A.9) \quad \sqrt{K}(\tilde{b} - \beta) = \frac{\Sigma(b(y_k, \dot{X}_k; \mu_D^0, \Sigma_D, \Lambda^0) - \beta)}{\sqrt{K}} - \alpha \sqrt{K}(\bar{D} - \mu_D^0)$$

$$+ \left[E\left[\frac{\partial b}{\partial \Lambda}\right]' \right]\sqrt{K}(\hat{\Lambda} - \Lambda^0) + o_p(1)$$

$\Sigma_{ab(D,\lambda)}$ is the variance-covariance matrix of the RHS terms above, and $S_{ab(D,\lambda)}$ is the sample covariance matrix of the components of the RHS terms above. Consistency of $S_{ab(D,\lambda)}$ follows from the regularity conditions of Assumption A6. QED

Notes

1. These examples, as well as many others, are described in many economics textbooks in current use; c.f. Varian(1984a), among many others.

2. Note that by defining $u=y-F(X)$ we have that $y=F(X)+u$, so this interpretation holds in an artificial way for all models. The important point about ε is that it coincides with the specific modeling of individual differences, with the derivative $\partial f/\partial X$ defined holding ε constant. This issue arises in the correct understanding of the results of Zellner(1969), for example, where y is a linear function of X, with ε representing varying slope coefficients.

3. $\ell(X)$ is the score vector of translation family $p(X|\theta) \equiv p(X-\theta)$ evaluated at $\theta=0$; see Stoker(1984b) for details.

4. $\hat{d}_1$ is the "scaled coefficient" estimator proposed by Stoker(1984b); namely if $E(y|X)=F^*(X'\beta)$, then $\hat{d}_1$ consistently estimates $\gamma\beta$, where γ is a scalar. This can be seen by applying Theorem 1 here to F^*.

5. This connection between average derivatives and OLS coefficients when X is normally distributed is noted by Ruud(1984).

6. See Stoker(1984a) for the application of completeness to the study of aggregation problems in macroeconomic equations.

7. Whether functionals of kernel density estimates can provide $\sqrt{K}$ consistent estimators of the true functional value is an open question of substantial interest. While the $\sqrt{K}$ consistency results of Ahmad(1976) do not obviously extend to multivariate situations, the modification of the results of Stock(1984,1985) on multivariate regression functions may provide multivariate score function results as required here.

8. $N(X)$ can be taken to be a multivariate normal density, with any mean and covariance matrix value, such as mean 0 and covariance I. Notice that if $N(X)$ has mean μ_X and covariance Σ_X, then $\hat{d}_1$ of section 3 computed using $\hat{p}(X)$ of (6.1) is equivalent to the vector of OLS coefficients of y regressed on X plus the covariance between y and $-\partial \ln p/\partial X$, so that the nonnormality of the distribution of X is directly represented in $\hat{d}_1$ by its departure from the OLS coefficients.

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.

Ahmed, I.A.(1976), "On the Asymptotic Properties of an Estimate of a Functional of a Probability Density," _Scandinavian Actuarial Journal_, 176-181.

Barnett, W.A.(1984), "The Minflex-Laurent Translog Flexible Functional Form," Discussion Paper 84-7, Center for Economic Research, Department of Economics, University of Texas at Austin, forthcoming in _Journal of Econometrics_.

Barnett, W.A. and A. Jonas(1983), "The Müntz-Szatz Demand System: An Application of a Globally Well Behaved Series Expansion," _Economics Letters_, 11, 337-342.

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.

Diewert, W.E., and T.J. Wales(1984), "Flexible Functional Forms and Global Curvature Conditions," NBER Technical Working Paper No. 40, forthcoming in _Econometrica_.

Epstein, L.G. and A. Yatchew(1984), "Nonparametric Hypothesis Testing Procedures and Applications to Demand Analysis," Draft, Department of Economics, University of Toronto, May.

Fleming, W.(1977), <u>Functions of Several Variables</u>, 2nd edition, New York, Springer-Verlag.

Gallant, A.R.(1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," <u>Journal of Econometrics</u>, 15, 211-245.

Gallant, A.R.(1982), "Unbiased Determination of Production Technologies," <u>Journal of Econometrics</u>, 20, 285-323.

Gallant, A.R. and D.W. Nychka(1985), "Semi-Nonparametric Maximum Likelihood Estimation," Working Paper, Institute of Statistics, North Carolina State University.

Lehmann, E.L. and H. Scheffe(1950), "Completeness, Similar Regions and Unbiased Estimation, Part I," <u>Sankya</u>, 10, 305-340.

Lehmann, E.L. and H. Scheffe(1955), "Completeness, Similar Regions and Unbiased Estimation, Part II," <u>Sankya</u>, 15, 219-236.

Newey, W.K.(1984), "A Method of Moments Interpretation of Sequential Estimators," <u>Economics Letters</u>, 14, 201-206.

Prakasa-Rao, B.L.S.(1983), <u>Nonparametric Functional Estimation</u>, Orlando, Florida, Academic Press.

Ruud, P.A.(1984), "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," Draft, revised October.

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

Stock, J.H.(1984), "Nonparametric Estimation of the Effects of Distributional Changes in Exogenous Variables," Discussion Paper Series No. 127D, Kennedy School of Government, Harvard University, July.

Stock, J.H.(1985), "Semiparametric Hedonics," Draft, Harvard University, January.

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

Stoker, T.M.(1984a), "Completeness, Distribution Restrictions and the Form of Aggregate Functions," <u>Econometrica</u>, 52, 887-908.

Stoker, T.M.(1984b), "Consistent Estimation of Scaled Coefficients," MIT Sloan School of Management Working Paper No. 1583-84, July.

Stoker, T.M.(1985), "Aggregation, Efficiency and Cross Section Regression," MIT Sloan School of Management Working Paper No. 1453-83, forthcoming in <u>Econometrica</u>.

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.(1984a), Microeconomic Analysis, 2nd edition, New York, Norton.

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

Zellner, A.(1969), "On the Aggregation Problem, A New Approach to a Troublesome Problem," in Economic Models, Estimation and Risk Programming, K. Fox, J. K. Sengupta and G. V. L. Narasimham, New York, Springer-Verlag.