Nonparametric Least Squares Regression and
Testing in Economic Models[1]

Adonis Yatchew and Len Bos*

This paper proposes a tractable and consistent estimator of the (possibly multi-equation) nonparametric regression model. The estimator is based on least squares over sets of functions bounded in Sobolev norm and is closely related to penalized least squares. We establish consistency and rate of convergence results as well as asymptotic normality of the (suitably standardized) sum of squared residuals is established. These results are then used to produce a $T^{1/2}$-consistent, asymptotically normal estimator in the partial linear model $y = z'\beta + f(x) + \varepsilon$. Conditional moment tests are provided for a variety of hypotheses including specification, significance, additive and multiplicative separability, monotonicity, concavity and demand theory. The validity of bootstrap test procedures is proved.

Keywords: nonparametric regression, semiparametric regression, partial linear model, least squares, empirical processes, hypothesis testing, conditional moment tests, U-statistics, bootstrap, specification test, significance test, additive separability, multiplicative separability, homogeneity, monotonicity, concavity, demand theory, maximization hypothesis

October 23, 1997.

* Department of Economics, University of Toronto, yatchew@chass.utoronto.ca
Department of Mathematics and Statistics, University of Calgary.

TABLE OF CONTENTS

1. INTRODUCTION

This paper proposes a class of nonparametric estimators for regression models based on least squares over sufficiently smooth sets of functions. The estimators are closely related to penalized least squares, (see e.g., Wahba (1990)). They easily permit the imposition of additional constraints on estimation such as a partial linear specification, additive or multiplicative separability and specifications involving monotonicity, concavity or the implications of demand theory. Tests of a variety of hypotheses can be performed by inserting the constrained estimator in a conditional moment statistic, (an approach used by Fan and Li (1996) to produce nonparametric tests of specification and significance in a kernel regression setting).

There are technical benefits to working in the particular function space defined below. (Estimation takes place over balls of functions in a Sobolev space.) First, the space is a Hilbert space, thus allowing one to take projections, to decompose spaces into mutually orthogonal complements, and to transform the search for the best fitting function in an infinite dimensional space into a finite dimensional, (in the simplest cases quadratic) optimization problem. Computation is straightforward for nonparametric regression models that are functions of one or several variables. (Indeed, the computer code for the latter is hardly more complicated than for the former.)

Second, balls of functions in Sobolev space are bounded, (in supnorm), and have a number of bounded derivatives. From the point of view of asymptotic theory, this permits estimation over rich sets of functions with sufficiently low metric entropy so that uniform (over classes of functions) laws of large numbers, rate of convergence and central limit results apply. In the result, it is straightforward to obtain consistency, rate of convergence, and certain asymptotic normality results as well as $T^{1/2}$-consistency in the partial linear model by using the tools of empirical process theory. (See Dudley (1984), Pollard (1984) and the survey by Andrews (1994b)). We rely upon the results of Van de Geer (1987,1990) to prove a number of our results.

The logical structure of the paper proceeds as follows. Section 2 assembles and proves the necessary results on Sobolev spaces. Section 3 defines our nonparametric least squares problem, shows that it can be transformed into a finite dimensional problem, establishes the rate of convergence and demonstrates asymptotic normality of the (suitably standardized) sum of squared residuals. (The latter result is fundamental to demonstrating $T^{1/2}$-consistency and asymptotic normality in the partial linear model estimator of Section 4.) Section 3 continues with extensions to a multi-equation setting and cross-validation as a means of selecting the degree of smoothness of the class of functions over which optimization takes place. A triangular array convergence result, which is later used in demonstrating the validity of bootstrap inference (Efron (1979)), is also derived.

Section 4 produces a $T^{1/2}$-consistent estimator for the partial linear model $y = z'\beta + f(x) + \varepsilon$. The estimator follows Robinson (1988) except that in the first step, we use nonparametric least squares rather than kernel estimators.

Section 5 focusses on devising constrained estimators and on a conditional moment test procedure. A variety of hypotheses are considered including specification, significance, additive and multiplicative separability, monotonicity, concavity and demand theory.

The literature on non/semiparametric estimation is massive. Surveys in the economics profession include Delgado and Robinson (1992), Härdle and Linton (1994), Matzkin (1994) and Powell (1994), Yatchew (1997a), (see also Härdle (1990)). Isotonic regression is covered extensively in Barlow et al (1972) and Robertson et al (1988). Inequality constrained smooth regression is considered by Wright and Wegman (1980), Wong (1984), Utreras (1985), Villalobos and Wahba (1987), Mukarjee (1988), Mammen (1991) and Mukarjee and Stern (1994). Specification tests against nonparametric alternatives are proposed by Bierens (1982, 1990), Eubank and Spiegelman (1990), Whang and Andrews (1991), Lee (1991), Wooldridge (1992), Härdle and Mammen (1993), Gozalo (1993), Yatchew (1988, 1992), Horowitz and Härdle (1994), Li (1994), Hong and White (1995), Fan and Li (1996), (see also White (1994)). Robinson (1989) proposes tests for semi/nonparametric time series models. Stoker (1989,1991) proposes tests of significance, symmetry, homogeneity and linear separability. Gallant (1982), Varian (1985,1990) and Epstein and Yatchew (1985) propose nonparametric tests of the maximization hypothesis. Hausman and Newey (1995) estimate consumer surplus using kernel estimators. Pinske and Robinson (1995) propose a method for pooling nonparametric regressions of similar shape. Partial linear models have been studied extensively, see Wahba (1984, 1985), Engle et al (1986), Heckman (1986), Rice (1986), Robinson (1988), Chen (1988, 1991), Shiau and Wahba (1988), Speckman (1988), Eubank and Whitney (1989), Chen and Shiau (1994), Linton (1995a) , Yatchew (1997b), (see also Linton (1995b)), and Mammen and Van de Geer (1995).

A notational convention throughout the paper will be the use of a ' ^{B} ' superscript to denote a bootstrap sample, estimator, test statistic.... Subscripts are used to index elements of sequences, vectors and matrices, e.g., $[AB]_{st}$ is the st-th entry of the matrix AB .

2. SOBOLEV SPACE RESULTS

Let $\mathbb{N}$ be the non-negative natural numbers. Let $Q^q \subset \mathbb{R}^q$ be the unit cube which will be the domain of the nonparametric regression models below. (The proposed estimators are also valid if the domain is a rectangular cube.) Suppose $\alpha = (\alpha_1,\ldots,\alpha_q) \in \mathbb{N}^q$, define $|\alpha|_\infty = max\,|\alpha_i|$, and let $x = (x_1,\ldots,x_q) \in \mathbb{R}^q$. We will use the following standard derivative notation $D^\alpha f(x) = \partial^{\alpha_1+..+\alpha_q} f(x) \,/\, \partial x_1^{\alpha_1} \ldots \partial x_q^{\alpha_q}$.

Let C^m be the space of m-times continuously differentiable scalar functions, i.e., $C^m = \{f\colon Q^q \to \mathbb{R}^1 \mid D^\alpha f \in C^0,\ |\alpha|_\infty \leq m\}$ where $C^0 = \{f\colon Q^q \to \mathbb{R}^1 \mid f \text{ continuous on } Q^q\}$. On the space C^m define the norm, $\|f\|_{\infty,m} = \sum_{|\alpha|_\infty \leq m} max_{x \in Q^q} \mid D^\alpha f(x) \mid$ in which case C^m is a complete, normed, linear space, i.e., a Banach space. Consider the following inner product of scalar functions and the induced norm:

$$(2.1)\qquad \begin{aligned} <f,g>_{Sob} &= \sum_{|\alpha|_\infty \le m} \int_{Q^q} D^\alpha f D^\alpha g \\ \| f \|_{Sob} &= \left[\sum_{|\alpha|_\infty \le m} \int_{Q^q} [D^\alpha f]^2 \right]^{1/2} \end{aligned}$$

and define the Sobolev space $\mathcal{H}^m$ as the completion of $\{f \in C^m\}$ with respect to $\|f\|_{Sob}$. The following results on the Sobolev space $\mathcal{H}^m$ will be used extensively.

THEOREM 2.1: *$\mathcal{H}^m$ is a Hilbert space.* ∎

The Hilbert space property will allow us to take projections and to express $\mathcal{H}^m$ as a direct sum of subspaces that are orthogonal to one another. [2]

THEOREM 2.2: *Given $a \in Q^q$ and $b \in \mathbb{N}^q$, $|b|_\infty \le m-1$, there exists a function, $r_a^b \in \mathcal{H}^m$ called a representor s.t. $<r_a^b, f>_{Sob} = D^b f(a)$ for all $f \in \mathcal{H}^m$. Furthermore, $r_a^b(x) = \prod_{i=1}^{q} r_{a_i}^{b_i}(x_i)$ for all $x \in Q^q$, where $r_{a_i}^{b_i}(.)$ is the representor in the Sobolev space of functions of one variable on Q^1 with inner product $<f,g>_{Sob} = \sum_{\alpha=0}^{m} \int_{Q^1} \frac{d^\alpha f}{dx^\alpha} \frac{d^\alpha g}{dx^\alpha}$.* ∎

If b equals the zero vector, then we have representors of function evaluation which we will often write as $r_a = r_a^0$. Theorem 2.2 further assures us of the existence of representors for derivative evaluation (of order $|b|_\infty \le m-1$). The problem of solving for representors is well known in the literature, (see Wahba (1990)). For the inner product above, representors of function evaluation consist of two functions spliced together, each of which is a linear combination of trigonometric functions. Formulae may be derived using elementary methods, in particular integration by parts and the solution of a linear differential equation. Details may be found in Appendix 2. Finally, Theorem 2.2 states that representors in spaces of functions of several variables may be written as products of representors in spaces of functions of one variable, (a result which also applies if the domain is a rectangular cube rather than the unit cube which we assume for most of this paper). This particularly facilitates their implementation. When doing multiple (nonparametric) regression, we will want to calculate representors for functions of several variables, but this only requires writing a subroutine which calculates representors for functions of one variable and calling it repeatedly. We note that one usually sees the spaces C^m defined using $|\alpha| := \alpha_1 + \ldots + \alpha_q$. The version used in this paper yields the tensor product of the corresponding univariate spaces which, as may be seen from Theorem 2.2, greatly simplifies the computation.

THEOREM 2.3: *The imbedding* $\mathcal{H}^m \rightarrow C^{m-1}$ *is compact.* ■

Compactness of the imbedding means that given a ball of functions in $\mathcal{H}^m$ (with respect to $\|f\|_{Sob}$), its closure is compact in C^{m-1}, (with respect to $\|f\|_{\infty,m}$). This result ensures that functions in a bounded ball in $\mathcal{H}^m$ have all lower order derivatives bounded in supnorm. Our estimation will take place over such balls of functions in a Sobolev space. The juxtaposition of the Sobolev spaces and the C-spaces occurs here for the following reason. The Sobolev spaces, because they are Hilbert spaces, facilitate calculation of least squares projections. The C-spaces, on the other hand, facilitate the application of results from empirical process theory. [3]

THEOREM 2.4: *Divide* x *into two subsets* $x = (x_a, x_b)$. *If* $f(x_a, x_b)$ *is of the form* $f_a(x_a)\, f_b(x_b)$, *then* $\|f\|_{Sob}^2 = \|f_a\|_{Sob}^2\, \|f_b\|_{Sob}^2$. *If* $f(x_a, x_b)$ *is of the form* $f_a(x_a) + f_b(x_b)$ *and either* $\int f_a = 0$ *or* $\int f_b = 0$ *then* $\|f\|_{Sob}^2 = \|f_a\|_{Sob}^2 + \|f_b\|_{Sob}^2$. ■

These results will be useful for analyzing multiplicatively and additively separable models. If the domain is a rectangular cube rather than the unit cube, then in the multiplicative case, $\|f_a\|_{Sob}^2$ and $\|f_b\|_{Sob}^2$ are calculated over the domain of x_a and x_b respectively; in the additive case, $\|f_a\|_{Sob}^2$ and $\|f_b\|_{Sob}^2$ are both calculated over the domain of (x_a, x_b).

3. NONPARAMETRIC SOBOLEV LEAST SQUARES

3.1 Estimation

Consider the following model:

(3.1.1) $$y_t = f_o(x_t) + \upsilon_t \qquad t = 1, \dots, T$$

ASSUMPTIONS FOR THE SINGLE EQUATION MODEL: i) *x_t are q-dimensional random variables, i.i.d. with probability law P_x and density p_x bounded away from zero on the support Q^q, the unit cube in $\mathbb{R}^q$; ii) υ_t are i.i.d. random variables with probability law $P_{\upsilon o}$ which has mean 0 and variance $\sigma^2_{\upsilon o}$; $P_{\upsilon o} \in \mathcal{P}_\upsilon$ a collection of probability laws with mean 0 and support contained in a bounded interval of $\mathbb{R}^1$; x_t and υ_t are independent; iii) $\mathfrak{F}$ is a family of functions in the Sobolev space $\mathcal{H}^m$ from $\mathbb{R}^q$ to $\mathbb{R}^1$, $m > \frac{q}{2}$, $\mathfrak{F} = \left\{f \in \mathcal{H}^m \colon \|f\|^2_{Sob} \leq L\right\}$.* ■

The Sobolev ball $\mathfrak{F}$ is compact in C^{m-1}. (It is pre-compact using Theorem 2.3, and it can be shown to be closed.) The condition $m > \frac{q}{2}$ is a minimum condition which ensures consistency and asymptotic normality of the average sum of squared residuals. Generally we will also want $m \geq 2$ in which case , $\mathfrak{F}$ is an equicontinuous family of functions with respect to supnorm, since by Theorem 2.3 first derivatives are bounded. Indeed, the set $D^b\mathfrak{F} = \{D^b f \mid f \in \mathfrak{F}\}$ is equicontinuous with respect to supnorm for all b satisfying $|b|_\infty \leq m-2$. (If $\mathfrak{F}$ is the unit ball in $\mathcal{H}^m$, then $D^b\mathfrak{F}$ is a closed subset of the unit ball in $\mathcal{H}^{m-|b|_\infty}$ so one can just apply the imbedding theorem again.)

Next, let $r_{x_1}, \dots, r_{x_T}$ be the representors for function evaluation at $x_1,\dots,x_T$ respectively, i.e. $\langle r_{x_t}, f\rangle_{Sob} = f(x_t)$ for all $f \in \mathcal{H}^m$. Let R be the TxT representor matrix whose columns (and rows) equal the representors evaluated at $x_1,\dots,x_T$; i.e., $R_{ij} = \langle r_{x_i}, r_{x_j}\rangle_{Sob} = r_{x_i}(x_j) = r_{x_j}(x_i)$.

THEOREM 3.1.1: *Let* $y = (y_1,...,y_T)'$ *and define*

(3.1.2) $$\hat{\sigma}^2 = \min_f \frac{1}{T} \sum_t [y_t - f(x_t)]^2 \qquad \text{s.t.} \quad \| f \|^2_{Sob} \le L$$

(3.1.3) $$s^2 = \min_c \frac{1}{T} [y - Rc]' [y - Rc] \qquad \text{s.t.} \quad c'Rc \le L$$

where c *is a Tx1 vector and* R *is the representor matrix. Then* $\hat{\sigma}^2 = s^2$. *Furthermore, there exists a solution to (3.1.2) of the form* $\hat{f} = \sum_1^T \hat{c}_t r_{x_t}$ *where* $\hat{c} = (\hat{c}_1,...,\hat{c}_T)'$ *solves (3.1.3).* ■

This theorem provides for the transformation of the infinite dimensional problem into a finite dimensional (quadratic) optimization problem. The estimator $\hat{f}$, can be expressed as a linear combination of the representors with the number of terms equal to the number of observations. (Perfect fit is precluded, except by extraordinary coincidence, since the coefficients must satisfy the quadratic smoothness constraint.) However, unlike nonparametric series estimators that approximate an infinite dimensional function space by an expanding sequence of finite dimensional subsets, (i.e., sieve estimators), our estimating function is extracted from a *fixed* set of functions $\Im$, regardless of sample size. The estimator $\hat{f}$ is unique a.s., (since $\|\hat{f}\|^2_{Sob} = L$ a.s. and adding a function that is orthogonal to the space spanned by the representors will increase the norm). Finally, we note that our optimization problem is closely related to penalized least squares.[4]

THEOREM 3.1.2: *Let* $\hat{f}$ *satisfy* $s^2 = \min \frac{1}{T} \sum (y_t - f(x_t))^2$ *s.t.* $f \in \Im$. *Suppose* $f_o \in \Im$, *then:*

(a) $s^2 \xrightarrow{a.s.} \sigma^2_{\upsilon o}$

(b) $\frac{1}{T} \sum (\hat{f}(x_t) - f_o(x_t))^2 = O_p(T^{-r})$ where $r = \frac{2m}{2m+q}$

(c) $T^{1/2} \left[s^2 - \sigma^2_{\upsilon o}\right] \xrightarrow{D} N(0, Var(\upsilon^2))$. ■

$Var(\upsilon^2)$ may be estimated consistently using fourth order moments of the estimated residuals $\hat{\upsilon}_t = y_t - \hat{f}(x_t)$. The rate of convergence, which is derived using a lemma of Van de Geer (1990), is determined by the metric entropy of $\Im$. Let $N(\delta ; \Im)$ be the minimum number of balls of radius δ in supnorm required to cover the set of functions $\Im$. If $N(\delta ; \Im) \le A\delta^{-\zeta}$ for positive

A, ζ, then $\frac{1}{T}\sum\left(\hat{f}(x_t)-f_o(x_t)\right)^2$ is shown to be $O_p\left(T^{-\frac{2}{2+\zeta}}\right)$. For the Sobolev ball of Theorem 3.1.2, $\zeta=\frac{q}{m}$ hence the resulting rate of convergence. Furthermore, if first derivatives are bounded, $(m\geq 2)$, convergence in mean square implies convergence of $\hat{f}$ to f in supnorm. (In fact, $sup_x\left|D^b\hat{f}-D^bf_o\right|\overset{a.s.}{\to}0$ for $|b|_\infty\leq m-2$.)

The practical nonparametric estimation problem of equation (3.1.3) is straightforward as one is minimizing a quadratic function subject to a quadratic constraint -- see e.g., Golub and Van Loan (1989, p.564) for an efficient algorithm. Using Fortran code and a 90 MHz Pentium processor, calculation of the representor matrix R with $T=100$ takes about 60 seconds. Subsequent solution of the optimization problem can be performed about 20 times per second, thus the procedure is not only feasible but amenable to bootstrap resampling methods. (Code for the calculation of the representor functions may be tested by taking inner products with various functions to determine whether the value of the function is reproduced.)

3.2 Extensions to the Multi-Equation Setting

Consider now the multi-equation model with assumptions modified as follows:

(3.2.1) $$\underset{px1}{y_t}=\underset{px1}{f_o}\ \underset{qx1}{(x_t)}+\underset{px1}{\upsilon_t}\qquad t=1,\ldots,T$$

ASSUMPTIONS FOR THE MULTI-EQUATION MODEL: *i)* x_t *are q-dimensional random variables i.i.d. with probability law* P_x *and density* p_x *bounded away from zero on the support* Q^q, *the unit cube in* $\mathbb{R}^q$; *ii)* υ_t *are p-dimensional i.i.d. random variables with probability law* $P_{\upsilon o}$ *which has mean 0 and covariance matrix* $\Sigma_{\upsilon o}$; $P_{\upsilon o}\in\mathcal{P}_\upsilon$ *a collection of probability laws with mean 0 and support contained in a bounded closed cube in* $\mathbb{R}^q$; x_t *and* υ_t *are independent; iii)* $\mathfrak{F}$, *a family of functions from* $\mathbb{R}^q$ *to* $\mathbb{R}^p$, *is a cross-product of Sobolev balls:* $\mathfrak{F}=\{f=(f_1,\ldots f_p)'\,|\ f_i\in\mathcal{H}^m,\ \|f_i\|^2_{Sob}\leq L_i,\ i=1,\ldots,p\}$ *where* $m>\frac{q}{2}$. ■

THEOREM 3.2.1: *Let* Λ *be a positive definite matrix and define:*

(3.2.1) $$\hat{\sigma}^2 = \min_{f} \frac{1}{T} \sum_{t} [y_t - f(x_t)]' \Lambda [y_t - f(x_t)] \quad s.t.\ \|f_i\|^2_{Sob} \le L_i,\ i=1,\dots,p$$

(3.2.2) $$s^2 = \min_{C} \frac{1}{T} \sum_{t} \left[y_t - C' \begin{pmatrix} R_{t1} \\ \vdots \\ R_{tT} \end{pmatrix} \right]' \Lambda \left[y_t - C' \begin{pmatrix} R_{t1} \\ \vdots \\ R_{tT} \end{pmatrix} \right] \quad s.t.\ (C_{1i},\dots,C_{Ti}) R \begin{pmatrix} C_{1i} \\ \vdots \\ C_{Ti} \end{pmatrix} \le L_i\ ,\ i=1,\dots,p$$

where C *is a* Txp *matrix,* $r_{x_1}, \dots, r_{x_T}$ *are the representors for function evaluation and* R *is the representor matrix of inner products of the* r_{x_t}. *Then* $\hat{\sigma}^2 = s^2$. *Furthermore, there exists a solution to the infinite dimensional problem (3.2.1) of the form* $\hat{f} = (\hat{f}_1,\dots,\hat{f}_p)'$, $\hat{f}_i = \sum_{t=1}^{T} \hat{C}_{ti} r_{x_t}$, $i=1,\dots,p$, *where* $\hat{C}$ *solves the finite dimensional problem (3.2.2).* ■

The estimator proposed here is an obvious generalization of the single equation estimator, (indeed it reduces to it, equation by equation, if we replace Λ with the identity matrix). Now define the natural estimator of the covariance matrix:

(3.2.3) $$\hat{\Sigma}_\upsilon = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{f}(x_t))(y_t - \hat{f}(x_t))'$$

For any symmetric pxp matrix A, define $uvec(A)$ to be the $p(p+1)/2$ dimensional column vector consisting of the upper triangular elements $\{\{A_{ij} | 1 \le i \le j\}; 1 \le j \le p\}$ of A.

THEOREM 3.2.2: *Let* $\hat{f}$ *satisfy the minimization problem of equation (3.2.1). If* $f_o \in \mathfrak{F}$, *then:*

(a) $\hat{\Sigma}_\upsilon \overset{a.s.}{\to} \Sigma_{\upsilon o}$

(b) $\frac{1}{T} \sum (\hat{f}_i(x_t) - f_{oi}(x_t))^2 = O_P(T^{-r})$ *for* $i = 1,\dots,p$ *where* $r = \frac{2m}{2m+q}$

(c) $T^{1/2} uvec[\hat{\Sigma}_\upsilon - \Sigma_{\upsilon o}] \overset{D}{\to} N(0, \Pi_o)$, $\Pi_o = Cov[uvec(\upsilon\upsilon')]$. ■

Apart from potential gains in efficiency, our interest in multi-equation nonparametric regression models is three-fold. First, as Robinson (1988) demonstrated, estimation and inference on the *parametric* component of the partial linear model can be reduced to analyzing the residuals of an appropriate multi-equation *nonparametric* regression model. In our setting, $T^{1/2}$-consistency and asymptotic normality of our estimator of the parametric component -- which is a function of $\hat{\Sigma}_{\upsilon}$ -- follow easily from Theorem 3.2.2(c). Second, the results of this theorem permit inference on other functions of the elements of $\Sigma_{\upsilon o}$, a subject which commands a substantial statistical literature.[5] Third is the application to systems of demand equations.

It is straightforward to verify that the results of Theorem 3.2.2 hold if Λ is replaced by a consistent estimate of $\Sigma_{\upsilon o}^{-1}$. Thus, estimation can proceed by first estimating the system equation by equation, using the residuals to estimate $\Sigma_{\upsilon o}$ consistently, then performing the finite dimensional multi-equation regression of Theorem 3.2.1.

3.3 Triangular Array Convergence

Staying with the multi-equation model of Section 3.2, fix a sequence $x_t, t = 1, \ldots, \infty$ that is dense in Q^q the support of x. Each ordered pair $\theta = (f, P_{\upsilon})$ that is an element of $\Theta = \mathfrak{F} \otimes \mathcal{P}_{\upsilon}$ defines a data generating mechanism (DGM), and the set Θ collects all possible DGMs. Where there may be ambiguity, we will indicate which DGM is being used to generate the data as in the following. Define a triangular array of random variables where each row corresponds to a (possibly) different DGM:

$$
(3.3.1) \qquad \begin{array}{ll}
\theta_1 & y_1(\theta_1) \\
\theta_2 & y_1(\theta_2), y_2(\theta_2) \\
\vdots & \vdots \\
\theta_T & y_1(\theta_T) \ldots\ldots y_T(\theta_T) \\
\vdots & \vdots
\end{array}
$$

$y_t(\theta_T) = f_T(x_t) + \upsilon_t(P_{\upsilon T})$, $t=1,\ldots,T$. Note that the dependence of υ_t on the DGM θ_T is only through $P_{\upsilon T}$. Let $\hat{f}$ be the estimator of Theorem 3.1.1 applied equation by equation. For a given DGM θ_T define the estimated residuals $\hat{\upsilon}_t(\theta_T) = y_t(\theta_T) - \hat{f}(x_t;\theta_T)$, $t=1,\ldots,T$. Then,

THEOREM 3.3.1 : *For any sequence of data generating mechanisms* $\{\theta_T\} = \{(f_T, P_{\upsilon T})\} \subset \Theta$,

$$T^{\frac{1}{2}}\ uvec\left(\frac{1}{T}\sum_{t=1}^{T} \hat{\upsilon}_t(\theta_T)\ \hat{\upsilon}_t'(\theta_T) - \frac{1}{T}\sum_{t=1}^{T} \upsilon_t(P_{\upsilon T})\,\upsilon_t'(P_{\upsilon T})\right) \xrightarrow{P} 0 \quad \blacksquare \tag{3.3.2}$$

Consider a sequence $\theta_T = (f_T, P_{\upsilon T})$ with associated covariance matrices $\Sigma_{\upsilon T}$ which converges to $\theta_o = (f_o, P_{\upsilon o})$. Then under our assumptions, (indeed uniformly bounded fourth order moments are sufficient), the Lindberg-Feller Theorem implies that:

$$T^{\frac{1}{2}}\ uvec\left(\frac{1}{T}\sum_{t=1}^{T} \upsilon_t(P_{\upsilon T})\ \upsilon_t'(P_{\upsilon T}) - \Sigma_{\upsilon T}\right) \xrightarrow{D} N(0, \Pi_o) \tag{3.3.3}$$

where $\Pi_o = Cov\left[uvec\left(\upsilon(P_{\upsilon o})\ \upsilon(P_{\upsilon o})'\right)\right]$. Combining this with Theorem 3.3.1, we have the following triangular array distributional result. Let $\theta_T = (f_T, P_{\upsilon T}) \rightarrow \theta_o = (f_o, P_{\upsilon o})$, then

$$T^{\frac{1}{2}}\ uvec\left(\frac{1}{T}\sum_{t=1}^{T} \hat{\upsilon}_t(\theta_T)\ \hat{\upsilon}_t'(\theta_T) - \Sigma_{\upsilon T}\right) \xrightarrow{D} N(0, \Pi_o) \tag{3.3.4}$$

We also note the following triangular array consistency result which follows immediately:

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\upsilon}_t(\theta_T)\ \hat{\upsilon}_t'(\theta_T) \xrightarrow{P} \Sigma_{\upsilon o} \tag{3.3.5}$$

Below, we will use these triangular array convergence results to justify bootstrap procedures.

3.4 Selection of Sobolev Norm Bounds Through Cross Validation

Cross validation is used extensively in nonparametric estimation for the purpose of selecting smoothing parameters. Examples include selection of the penalty parameter in penalized least squares and bandwidth selection in kernel regression. In our case, we are interested in selecting Sobolev norm bounds for the regression function f_o. If we select bounds that are much larger than the true norm, then heuristically, we should expect our estimators to be less efficient though they will be consistent. On the other hand, if we select bounds that are smaller than the true norm, then our estimator will in general be inconsistent. Consider the cross validation function:

$$CV(L) = \frac{1}{T} \sum_{t=1}^{T} \left[y_t - \hat{f}_{-t}(x_t)\right]^2 \tag{3.4.1}$$

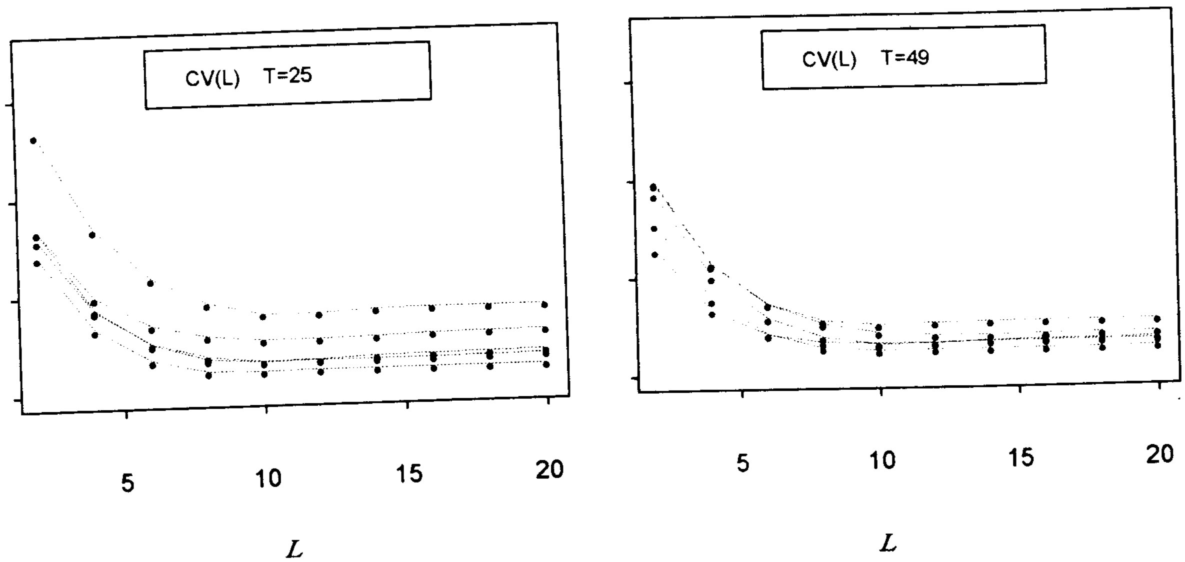
where $\hat{f}_{-t}$ is obtained by solving

$$\min_{f} \sum_{s \neq t}^{T} \left[y_s - f(x_s)\right]^2 \qquad s.t. \quad \|f\|_{Sob}^2 \leq L \tag{3.4.2}$$

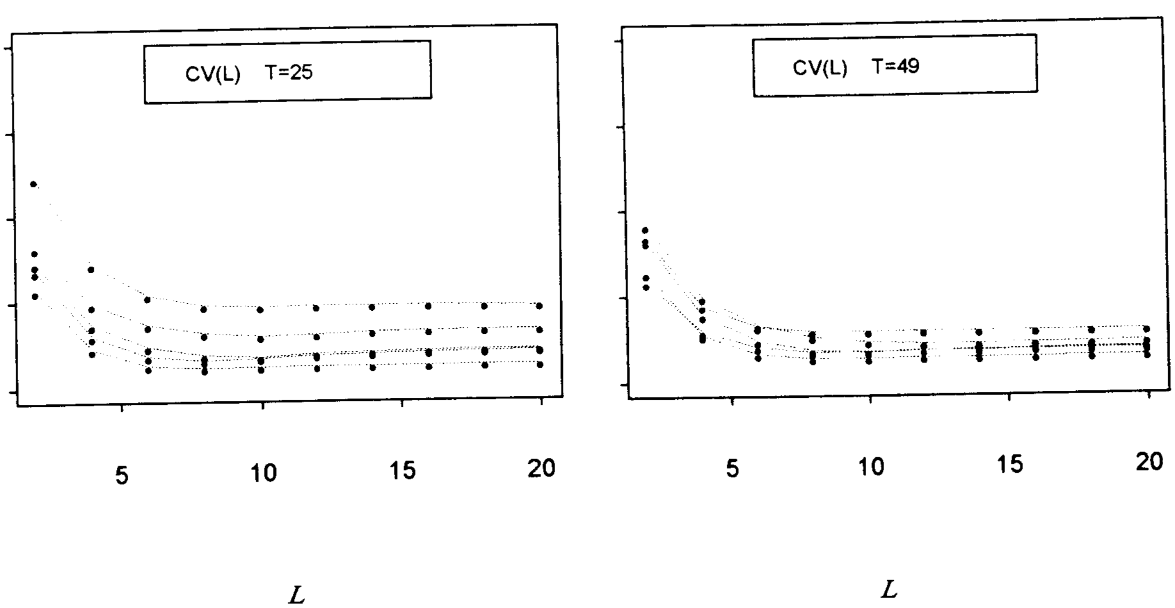
As in other applications of cross validation, the idea is to select the smoothing parameter based on its ability to predict outside the sample, (hence the omission of the t-th observation from estimation when the t-th observation is being predicted). The difference here is that the smoothing parameter is the actual squared norm of the regression function.

Elsewhere the minimum of the cross validation function has been used to estimate the smoothing parameter and is known to have certain optimality properties. While this paper does not contain theoretical results of this nature, (see Li (1986, 1987) for related results), simulations are encouraging as may be seen from Figure 1. The data were generated using linear and multiplicative specifications. In each case the simulated cross validation function, CV(L), exhibits a rapid initial decline. The minima are in the neighbourhood of the square of the true norm.

FIGURE 1: CROSS VALIDATION OF SOVOLEV NORM BOUNDS



LINEAR MODEL: $y=x_1+x_2+\varepsilon$ $x_1,x_2\epsilon[1,2]^2$ $\varepsilon \sim N(0,\sigma^2)$ $\|.\|^2_{Sob} = 11.167$ $\sigma^2=.64$



MULTIPLICATIVE MODEL: $y=x_1x_2+\varepsilon$ $x_1,x_2\epsilon[1,2]^2$ $\varepsilon \sim N(0,\sigma^2)$ $\|.\|^2_{Sob} =11.111$ $\sigma^2=.36$

Cross validation functions CV(L) defined in equations (3.4.1) and (3.4.2). Sobolev fourth order norm is used (m=4 in equation (2.1)).

4. THE PARTIAL LINEAR MODEL

Consider the partial linear model $y = z'\beta_o + f_o(x) + \varepsilon$, where x and z are independent of ε; z is a p-dimensional vector. We begin by specifying the conditional distribution of $z, y \mid x$:

$$\begin{array}{lclclcl} z & = & E(z|x) & + & u & = & g_o(x) + u \\ y & = & E(y|x) & + & v & = & h_o(x) + v \end{array} \tag{4.1}$$

where $h_o(x) = g_o(x)'\beta_o + f_o(x)$, $v = u'\beta_o + \varepsilon$. Equation (4.1) is a multi-equation regression examined in Sections (3.2) and (3.3) except that the dependent variable now consists of the $p+1$ vector (z, y). Let P_{uvo} denote the true joint distribution of u, v and define

$$Cov\begin{pmatrix} u \\ v \end{pmatrix} \equiv \begin{bmatrix} \Sigma_{zz|x} & \sigma_{zy|x} \\ \cdots & \sigma^2_{y|x} \end{bmatrix} = \begin{bmatrix} \Sigma_{zz|x} & \Sigma_{zz|x}\beta_o \\ \cdots & \sigma^2_\varepsilon + \beta_o{}'\Sigma_{zz|x}\beta_o \end{bmatrix} \tag{4.2}$$

Assumptions for the Partial Linear Model: *i) x_t, $t=1,\dots,\infty$ is dense in Q^q the unit cube in $\mathbb{R}^q$; ii) $\beta_o \in \mathfrak{Z}$, a compact subset of $\mathbb{R}^p$; iii) let $\mathcal{H}^m$ be the Sobolev space from $\mathbb{R}^q$ to $\mathbb{R}^1$, $m > \frac{q}{2}$, fix constants $L_{g_1},\dots,L_{g_p}, L_f$, then*

$$g_o \in G = \left\{ g(x) = (g_1(x),\dots,g_p(x))' \mid g_i \in \mathcal{H}^m, \|g_i\|^2_{Sob} \le L_{g_i}, i=1,\dots,p \right\}$$
$$f_o \in F = \left\{ f(x) \mid f \in \mathcal{H}^m, \|f\|^2_{Sob} \le L_f \right\}$$
$$h_o \in H = \left\{ h(x) \mid h \in \mathcal{H}^m, \|h\|^2_{Sob} \le L_h \right\}, \quad L_h = sup\left\{ \|g(x)'\beta + f(x)\|^2_{Sob} \mid g \in G, f \in F, \beta \in \mathfrak{Z} \right\}$$

iv) conditional on x, u and ε are independent of each other; the probability laws P_{uo} and $P_{\varepsilon o}$ have zero means and bounded support. ∎

Note that the DGM for (4.1) is completely determined by the elements $f_o, g_o, P_{uo}, P_{\varepsilon o}, \beta_o$ or equivalently by $g_o, h_o, \beta_o, P_{uvo}$. The results below form the basis for asymptotic as well as bootstrap inference on β. Given independent observations, apply the estimator of Theorem 3.1.1 to each equation to obtain estimates of $h_o, g_{oi}, i=1,\dots,p$. (These estimates may then be used to produce a multi-equation estimator as in Section 3.2.) Calculate sample moments of the estimated residuals to obtain $\hat\Sigma^{-1}_{zz|x}, \hat\sigma_{zy|x}$. Define $\hat\beta = \hat\Sigma^{-1}_{zz|x}\hat\sigma_{zy|x}$ and $\hat\sigma^2_\varepsilon = \frac{1}{T}\sum_t\left[(y_t - \hat h(x_t)) - (z_t - \hat g(x_t))'\hat\beta\right]^2$. Finally, re-centre the estimated residuals from each equation of (4.1) and construct the empirical distribution function $\hat P_{uv}$. The estimator $\hat\beta$ can be shown

FIGURE 2: PARTIAL LINEAR MODEL

DATA GENERATING MECHANISM:

$f_o(x) = x$, $g_o(x) = x$, $\beta_o = 1$, $\varepsilon \sim N(0,1)$, $u \sim N(0,9)$, x equispaced on $[0,10]$

Hence using equation (4.1):

$$y = z\beta_o + f_o(x) + \varepsilon \;= 2x + u + \varepsilon$$
$$z = g_o(x) + u \;=\; x + u$$

$$\| f_o \|^2_{Sob} = \| g_o \|^2_{Sob} = 343.33 \qquad \| h_o \|^2_{Sob} = 4*343.33$$

ESTIMATION:

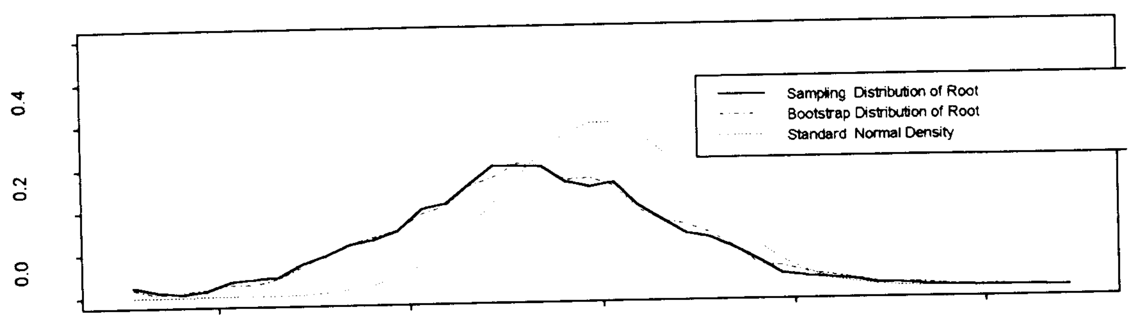
Sobolev least squares of y on x subject to $\| g \|^2_{Sob} = 5*343.33$ to obtain $\hat{g}$ and $\hat{u}$.
Sobolev least squares of z on x subject to $\| h \|^2_{Sob} = 5*4*343.33$ to obtain $\hat{h}$ and $\hat{v}$.

$$\hat{\beta} = \Sigma \hat{u}_t \hat{v}_t / \Sigma \hat{u}_t^2 \ , \qquad \hat{\sigma}_u^2 = \Sigma \hat{u}_t^2 / T \ , \qquad \hat{\sigma}_\varepsilon^2 = \frac{1}{T} \sum_t \left[(y_t - \hat{h}(x_t)) - (z_t - \hat{g}(x_t)) \hat{\beta} \right]^2$$

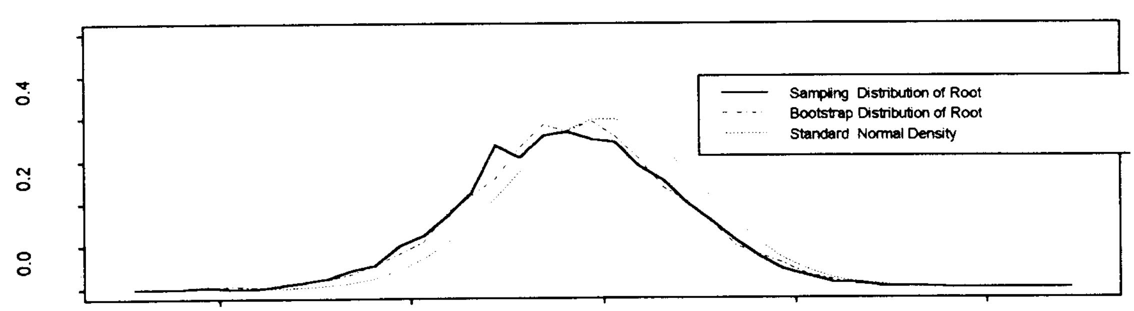
ROOT: (see Theorem 4.1 where $\sigma_u^2 = \Sigma_{zz|x}$).

$$T^{1/2} \frac{\hat{\beta} - 1}{\hat{\sigma}_\varepsilon / \hat{\sigma}_u} \sim N(0,1)$$

$T = 25$



$T = 100$



Sobolev fourth order norm is used (m=4 in equation (2.1)). Sampling distribution of the root based on 5000 samples. Bootstrap distribution based on 5000 resamples from a single initial sample. Simulations performed using Fortran code on a 90 MHz Pentium processor.

to satisfy the stochastic equicontinuity efficiency conditions of Andrews (1994a).

Let A be a full rank rxp matrix, $r \leq p$ and define the root, (which is an asymptotic pivot) $\eta = T\left[A\hat{\beta} - A\beta_o\right]'\left[A\ \hat{\Sigma}^{-1}_{zz|x}\ A'\right]^{-1}\left[A\hat{\beta} - A\beta_o\right] / \hat{\sigma}^2_\varepsilon$. Then,

THEOREM 4.1: $T^{1/2}\left(\hat{\beta} - \beta_o\right) \xrightarrow{D} N\left(0, \sigma^2_\varepsilon \Sigma^{-1}_{zz|x}\right), \quad \eta \xrightarrow{D} \chi^2_r$ ∎

Bootstrap inference may be conducted as follows: sample with replacement from $\hat{P}_{uv}$ to obtain $\left(u_1^B, v_1^B\right),\ldots,\left(u_T^B, v_T^B\right)$. Add these to the estimated regression functions in (4.1) to obtain $(y_1^B, z_1^B, x_1)\ldots(y_T^B, z_T^B, x_T)$. Re-estimate g_o, h_o and from the estimated residuals, calculate $\hat{\Sigma}^B_{zz|x}$, $\hat{\sigma}^B_{zy|x}$, β^B, $\hat{\sigma}^{2B}_\varepsilon$ and $\eta^B = T\left[A\,\beta^B - A\,\hat{\beta}\right]'\left[A\ \left(\hat{\Sigma}^B_{zz|x}\right)^{-1} A'\right]^{-1}\left[A\,\beta^B - A\,\hat{\beta}\right] / \hat{\sigma}^{2B}_\varepsilon$. Then,

THEOREM 4.2: $T^{1/2}\left(\beta^B - \hat{\beta}\right) \xrightarrow{D} N\left(0, \sigma^2_\varepsilon \Sigma^{-1}_{zz|x}\right), \quad \eta^B \xrightarrow{D} \chi^2_r$ ∎

The results in this section can be extended straightforwardly to the situation where z (but not x) is correlated with ε. An instrumental variable estimator may be constructed by first regressing the instruments on x using a nonparametric regression as in (4.1). Test statistics may be bootstrapped and a Hausman (1978) type test procedure for testing endogeneity of the z's may be constructed. Alternate resampling methodologies such as the wild bootstrap may also be used. Simulation results for a simple example may be found in Figure 2. To our knowledge, the only other partial linear model for which the bootstrap has been shown to be valid is Mammen and Van de Geer (1995).

5. CONSTRAINED ESTIMATION AND TESTING

5.1 Constrained Estimation

In section 3.1 we discussed estimation of the single equation model $y = f_o(x) + \upsilon$ subject to a smoothness constraint $f \in \mathfrak{F} = \{f \in \mathcal{H}^m : \|f\|^2_{Sob} \leq L\}$. In the remaining sections we focus on the imposition of additional constraints on f_o and on the testing of such constraints. In particular, we want to estimate subject to $f \in \bar{\mathfrak{F}} \subseteq \mathfrak{F}$ where $\bar{\mathfrak{F}}$ combines smoothness with further functional properties and to test $H_o : f_o \in \bar{\mathfrak{F}}$.

ASSUMPTIONS FOR THE CONSTRAINED SINGLE EQUATION MODEL: *i) invoke the Assumptions For the Single Equation Model, (section 3.1); ii)* $\bar{\mathfrak{F}} \subseteq \mathfrak{F}$ *is a closed set of functions such that the metric entropy* $\log N(\delta;\bar{\mathfrak{F}}) \leq A\,\delta^{-\zeta}$ *for some* $A > 0, \zeta > 0$; *iii)* $\{\bar{\mathfrak{F}}_T\}_{T=1}^{\infty}$ *is a descending sequence of closed and possibly random sets of functions* $\mathfrak{F} \supset \bar{\mathfrak{F}}_1 \supset \ldots \supset \bar{\mathfrak{F}}_T \supset \ldots \supset \bar{\mathfrak{F}}$ *such that* $\bigcap_1^{\infty} \bar{\mathfrak{F}}_T = \bar{\mathfrak{F}}$ *a.s. and* $\log N(\delta;\bar{\mathfrak{F}}_T) \leq A'\delta^{-\zeta}$, $T = 1, \ldots \infty$ *for some* $A' > 0$. ■

Think of $\{\bar{\mathfrak{F}}_T\}_{T=1}^{\infty}$ as a sequence of sets of functions that incorporates progressively more of the restrictions of $\bar{\mathfrak{F}}$. (Since all are subsets of $\mathfrak{F}$, $\zeta \leq \frac{q}{m}$ where q is the dimension of x, m is the order of the Sobolev norm. See section 3.1.) The elaboration involving descending sets is useful in models involving monotonicity, concavity or the implications of demand theory. (In such cases, optimization over $\bar{\mathfrak{F}}$ is technically difficult except in a limiting sense.) The elaboration will be redundant when performing significance or specification tests or when the restricted model is separable, (in these cases we set $\bar{\mathfrak{F}}_1 = \ldots = \supset \bar{\mathfrak{F}}_T = \bar{\mathfrak{F}}$).

PROPOSITION 5.1.1: *Let* $\hat{f}$ *satisfy* $s^2 = \min \frac{1}{T}\sum (y_t - f(x_t))^2$ *s.t.* $f \in \bar{\mathfrak{F}}_T$. *If* $f_o \in \bar{\mathfrak{F}}$ *then the conclusions of Th. 3.1.2 continue to hold with rate of convergence* $r = \frac{2}{2+\zeta}$. *Suppose* $f_o \notin \bar{\mathfrak{F}}$, $\|f_o\|^2_{Sob}$ *is finite and there exists a unique* $\bar{f}_o \in \bar{\mathfrak{F}}$ *satisfying* $\min_{f \in \bar{\mathfrak{F}}} \int (f_o - f)^2 \, dP_x$. *Then* $s^2 \overset{a.s.}{\to} \sigma^2_{\upsilon o} + \int (f_o - \bar{f}_o)^2 \, dP_x$. ■

5.2 Conditional Moment Tests [6]

We outline a testing approach which generalizes Fan and Li (1996) and Li (1994).[7] Let $\bar{f}_o$ be the 'closest' function to f_o in $\bar{\mathfrak{F}}$, (as in Proposition 5.1.1). The motivation for the test is the conditional moment condition:

$$(5.2.1) \qquad E_{\upsilon,x}\left[\left(y-\bar{f}_o(x)\right) E_{\upsilon}\left[y-\bar{f}_o(x) \mid x\right] p_x(x)\right] = E_x\left[\left(f_o(x)-\bar{f}_o(x)\right)^2 p_x(x)\right] \geq 0$$

where the inequality becomes an equality only if $f_o \in \bar{\mathfrak{F}}$ in which case $\bar{f}_o = f_o$. Let [8]

$$(5.2.2) \qquad U = \frac{1}{T}\sum_t \left(y_t - \hat{f}(x_t)\right)\left[\frac{1}{\lambda^q T}\sum_{s \neq t} \left(y_s - \hat{f}(x_s)\right) K\left(\frac{x_s - x_t}{\lambda}\right)\right]$$

where $K(\cdot)$ is a product kernel with common smoothing parameter λ. We assume that the underlying univariate kernel is symmetric having support $[-\frac{1}{2},\frac{1}{2}]$. The term in square brackets may be thought of as an estimator of $\left(f_o(x_t)-\bar{f}_o(x_t)\right)p_x(x_t)$. Expansion yields:

$$(5.2.3) \qquad \begin{aligned} U &= U_1 + U_2 + U_3 \\ &= \frac{1}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t \upsilon_s K\left(\frac{x_s - x_t}{\lambda}\right) \\ &+ \frac{1}{\lambda^q T^2}\sum_t \sum_{s\neq t} \left(f_o(x_t)-\hat{f}(x_t)\right)\left(f_o(x_s)-\hat{f}(x_s)\right) K\left(\frac{x_s - x_t}{\lambda}\right) \\ &+ \frac{2}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t\left(f_o(x_s)-\hat{f}(x_s)\right) K\left(\frac{x_s - x_t}{\lambda}\right) \end{aligned}$$

As in Fan and Li (1996), the essence of the proof of the theorem below is to first show that U_1 (suitably standardized), is approximately normally distributed. If the null hypothesis is true, then the restricted estimator $\hat{f} \rightarrow f_o$. If this convergence is sufficiently rapid, causing U_2 and U_3 to converge to zero rapidly, then the distribution of U (also standardized) is determined by the distribution of U_1. (If the null is false, then U diverges to $+\infty$.)

Assert the Assumptions for the Constrained Single Equation Model, (section 5.1) and define $\hat{f}$ as in Proposition 5.1.1. Then, THEOREM 5.2.1: *Suppose* $f_o \in \bar{\mathfrak{I}}$, $\lambda^q T \to \infty$ *and* $\lambda^{q/2} T^{1-r} \to 0$ *where* $r = \frac{2}{2+\zeta}$ *is the rate of convergence of the restricted estimator. Then*

$$\lambda^{q/2} T U \xrightarrow{D} N\left(0\, ,\, 2\sigma_\upsilon^4 \int p^2(x) \int K^2(u)\right) \tag{5.2.4}$$

Let the estimated variance of U *be given by:*

$$\hat{\sigma}_U^2 = \frac{2}{\lambda^{2q} T^4} \sum_t \sum_{s \neq t} \left(y_t - \hat{f}(x_t)\right)^2 \left(y_s - \hat{f}(x_s)\right)^2 K^2\left(\frac{x_s - x_t}{\lambda}\right) \tag{5.2.5}$$

Then $\lambda^q T^2 \hat{\sigma}_U^2 \xrightarrow{P} 2\sigma_\upsilon^4 \int p^2(x) \int K^2(u)$. *Hence* $U / \hat{\sigma}_U \xrightarrow{D} N(0,1)$. ■

It is important to keep in mind that in applying Theorem 5.2.1, q is the dimension of x in the *unrestricted* model. We bootstrap critical values as follows. Estimate f_o using a nonparametric estimator which is consistent under both the null and the alternative, (e.g., the estimator of section 3.1). (This ensures that the empirical distribution function of the estimated residuals $\hat{P}_\upsilon$ is consistent for P_υ whether or not the null hypothesis is true.) Construct recentered estimated residuals $\hat{\upsilon}_1, \dots, \hat{\upsilon}_T$. Sample from these and construct the bootstrap dataset $\left(y_t^B, x_t\right) = \left(\hat{f}(x_t) + \hat{\upsilon}_t^B,\, x_t\right)$ where $\hat{f}$ is the restricted estimator. Using these data reestimate f_o under the null hypothesis, (i.e., recalculate the restricted estimator) to obtain $\hat{f}^B$ and calculate the bootstrap value of the standardized test statistic $U^B / \hat{\sigma}_U^B$ by inserting $\hat{f}^B$ and the y_t^B into (5.2.2) and (5.2.5). Repeat many times to obtain the simulated distribution, (calculate the critical value of the test statistic from the right hand tail). Continuing with the assumptions of Theorem 5.2.1, we have:

THEOREM 5.2.2: $U^B / \hat{\sigma}_U^B \xrightarrow{D} N(0,1)$ ■

Significance Test: Define $\bar{\mathfrak{I}} = \{f(x_1, x_2) \in \mathcal{H}^m : \|f\|_{Sob}^2 \leq L\, ,\ f \text{ is constant wrt } x_2\}$ to be the restricted set of functions, x_1 and x_2 are scalars. From Theorem 3.1.2 we know that the nonparametric least squares estimator over $\bar{\mathfrak{I}}$ converges at a rate $r = 2m/(2m+1)$. Theorem 5.2.1 may then be used to obtain an appropriate rate of convergence for the kernel bandwidth λ, (e.g., if $m = 2$, then $\lambda = T^{-1/4} = o(T^{-1/5})$ suffices).

Specification Test: suppose we want to test the partial linear specification $y = z'\beta + f(x) + \varepsilon$. Let q_x be the dimension of x. A number of variants of semi/nonparametric least squares procedures exist which, if the specification is correct, yield the optimal rate of convergence $\frac{1}{T}\sum\left(z_t'\hat{\beta} + \hat{f}(x_t) - z_t'\beta_o - f_o(x_t)\right)^2 = O_p\left(T^{-2m/(2m+q_x)}\right)$. Theorem 5.2.1 may then be used to produce a test against the alternative hypothesis that the regression function is of the general nonparametric form $f(z,x)$.

5.3 *Additive Separability*

The nonparametrics literature has devoted considerable attention to improving the rate of convergence of nonparametric estimators using additive models, (see e.g., Stone (1985, 1986), Hastie and Tibshirani (1990)). Indeed, Stone demonstrates that under general conditions, the optimal rate of convergence for a spline regression that is additively separable in each explanatory variable does not deteriorate as one increases the number of variables. Partition $x = (x_a, x_b)$ with dimensions q_a, q_b respectively and $x \in Q^{q_a+q_b} = [0,1]^{q_a+q_b}$. (Extensions of the results in this section to more general additively separable specifications is straightforward.) Define $\mathfrak{F} = \left\{f(x_a, x_b) \in \mathcal{H}^m : f(x_a, x_b) = f_a(x_a) + f_b(x_b),\ \|f_a + f_b\|^2_{Sob} \le L,\ \int f_b = 0\right\}$ where the integral constraint will be seen to be an identification condition.

THEOREM 5.3.1: *Given data* $(y_1, x_{a1}, x_{b1}), \ldots, (y_T, x_{aT}, x_{bT})$ *let* $y = (y_1, \ldots, y_T)'$ *and define*

$$(5.3.1) \qquad \hat{\sigma}^2 = \min_{f_a, f_b} \frac{1}{T}\sum_t \left[y_t - f_a(x_{at}) - f_b(x_{bt})\right]^2 \quad \text{s.t.} \quad \|f_a + f_b\|^2_{Sob} \le L, \quad \int f_b = 0$$

$$(5.3.2) \qquad s^2 = \min_{c_a, c_b} \frac{1}{T}\left[y - R_a c_a - R_b c_b\right]' \left[y - R_a c_a - R_b c_b\right] \quad \text{s.t.} \quad c_a' R_a c_a + c_b' R_b c_b \le L, \quad \sum_t c_{bt} = 0$$

where c_a, c_b *are Tx1 vectors,* R_a, R_b *are the representor matrices on* $[0,1]^{q_a}$ *at* $x_{a1}, \ldots, x_{aT}$ *and on* $[0,1]^{q_b}$ *at* $x_{b1}, \ldots, x_{bT}$ *respectively. Then* $\hat{\sigma}^2 = s^2$. *Furthermore, there exists a solution to the infinite dimensional problem (5.3.1) of the form* $\hat{f}_a(x_a) + \hat{f}_b(x_b) = \sum_1^T \hat{c}_{at} r_{x_{at}}(x_a) + \hat{c}_{bt} r_{x_{bt}}(x_b)$ *where* $\hat{c}_a = (\hat{c}_{a1}, \ldots, \hat{c}_{aT})'$, $\hat{c}_b = (\hat{c}_{b1}, \ldots, \hat{c}_{bT})'$ *solve the finite dimensional problem (5.3.2).* ∎

FIGURE 3: CONDITIONAL MOMENT TEST OF ADDITIVE SEPARABILITY

DATA GENERATING MECHANISM:

$y = f_a(x_a) + f_b(x_b) + \upsilon = x_a + x_b + \upsilon \quad \upsilon \sim N(0,.25)$

x_a, x_b lie on a uniform grid in $[1,2]x[1,2]$

$\| f_a + f_b \|^2_{Sob} = 11.167$

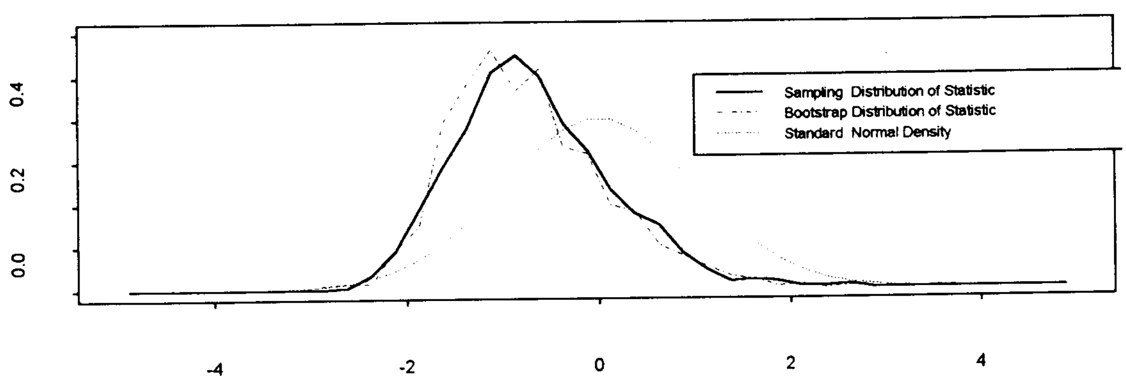
HYPOTHESES:

$H_o: y = f_a(x_a) + f_b(x_b) + \upsilon \quad H_1: y = f(x_a, x_b) + \upsilon$

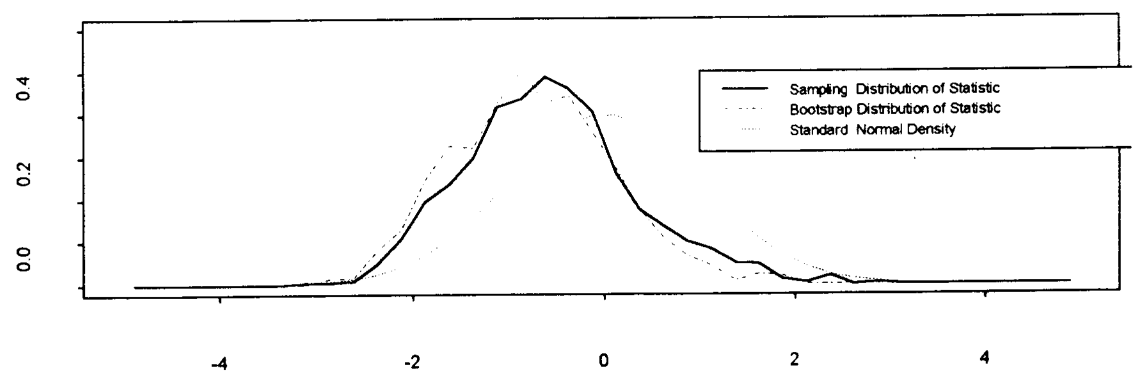
RESTRICTED ESTIMATOR: Sobolev least squares subject to $\| f_a + f_b \|^2_{Sob} \leq 11.167$ and an identification restriction $\int f_b = 0$.

TEST STATISTIC: $U / \hat{\sigma}_U \xrightarrow{D} N(0,1)$ under H_o (see equations 5.2.2 and Theorems 5.2.1 and 5.2.2).

T=25 , $\lambda = T^{-1/5} = .525$



T=49 , $\lambda = T^{-1/5} = .459$



Sobolev fourth order norm is used (m=4 in equation (2.1)). A uniform product kernel is used in the computation of U. From Theorem 5.2.1 we need $\lambda^{q/2} T^{1-r} \rightarrow 0$. With m=4, $q = 2$, $q_a = q_b = 1$ the rate of convergence of the restricted model is $r = 2m/(2m + \max\{q_a, q_b\}) = 8/9$, (see section 5.3). We set $\lambda = T^{-1/5}$. The sampling distribution of the test statistic is based on 1000 samples. The bootstrap distribution is based on 1000 resamples from a single initial sample. Fortran code was used to calculate the representor matrix. GAMS, (Brooke et al (1992)), was used to solve the restricted optimization problem (about 3 seconds per optimization for T=25 on a 90 MHz Pentium processor, 12 seconds for T=49).

The sets of functions $\{f_a(x_a)\}$ and $\left\{f_b(x_b) \mid \int f_b = 0\right\}$ are orthogonal in the Sobolev space $\mathcal{H}^m$ on $Q^{q_a+q_b}$. Thus, using the Hilbert space property of $\mathcal{H}^m$, (Theorem 2.1), a function f_a+f_b that satisfies the infinite dimensional optimization problem has a unique representation as a sum of functions in each of the two subpaces. The metric entropy for the set over which estimation takes place is given by $\log N(\delta;\bar{\mathfrak{F}}) \leq A\, \delta^{-\max\{q_a,q_b\}/m}$. Suppose the true regression function is given by $f_{ao}(x_a)+f_{bo}(x_b)$, then Proposition 5.1.1 yields a rate of convergence $r = \frac{2m}{2m+\max\{q_a,q_b\}}$. (Note that asymptotic normality of s^2 holds as long as $m > ½\max\{q_a,q_b\}$.) Theorem 5.2.1 then provides a test of additive separability. (Figure 3 contains simulation results.) As long as first derivatives are bounded then the component regressions are separately identified:

PROPOSITION 5.3.2 *if* $m \geq 2$ *then* $\sup_{x_a}|\hat{f}_a - f_{ao}| \overset{a.s.}{\to} 0$ *and* $\sup_{x_b}|\hat{f}_b - f_{bo}| \overset{a.s.}{\to} 0$. ■

5.4 Multiplicative Separability

An alternative structure which mitigates the curse of dimensionality is multiplicative. Define $\bar{\mathfrak{F}} = \left\{f(x_a,x_b) \in \mathcal{H}^m : f(x_a,x_b) = f_a(x_a)\cdot f_b(x_b)\,,\ \| f_a \cdot f_b \|^2_{Sob} \leq L,\right\}$ where $x = (x_a,x_b)$ with dimensions q_a, q_b respectively and $x \in Q^{q_a+q_b} = [0,1]^{q_a+q_b}$. Recall from Theorem 2.4 that in this case $\|f(x_a,x_b)\|^2_{Sob} = \|f_a(x_a)\|^2_{Sob} \cdot \|f_b(x_b)\|^2_{Sob}$. Recall also that for an arbitrary vector w, $[w]_t$ denotes its t-th element.

THEOREM 5.4.1: *Given data* $(y_1,x_{a1},x_{b1}),\dots,(y_T,x_{aT},x_{bT})$ *let* $y = (y_1,\dots,y_T)'$ *and define*

$$\hat{\sigma}^2 = \min_{f_a,f_b} \frac{1}{T}\sum_t\, [y_t - f_a(x_{at})\cdot f_b(x_{bt})]^2 \quad \text{s.t.} \quad \| f_a \cdot f_b \|^2_{Sob} \leq L \tag{5.4.1}$$

$$s^2 = \min_{c_a,\,c_b} \frac{1}{T}\sum \left(y_t - [R_a c_a]_t \cdot [R_b c_b]_t\right)^2 \quad s.t. \quad c_a' R_a c_a \cdot c_b' R_b c_b \leq L \tag{5.4.2}$$

where c_a, c_b *are Tx1 vectors,* R_a, R_b *are the representor matrices on* $[0,1]^{q_a}$ *at* $x_{a1},\dots,x_{aT}$ *and on* $[0,1]^{q_b}$ *at* $x_{b1},\dots,x_{bt}$ *respectively. Then* $\hat{\sigma}^2 = s^2$. *Furthermore, there exists a solution to the* *infinite dimensional problem (5.4.1) of the form* $\hat{f}_a(x_a)\cdot\hat{f}_b(x_b) = \left(\sum_1^T \hat{c}_{at} r_{x_{at}}(x_a)\right)\cdot\left(\sum_1^T \hat{c}_{bt} r_{x_{bt}}(x_b)\right)$ *where* $\hat{c}_a = (\hat{c}_{a1},\ldots,\hat{c}_{aT})'$, $\hat{c}_b = (\hat{c}_{b1},\ldots,\hat{c}_{bT})'$ *solve the finite dimensional problem (5.4.2).* ■

In order that f_a and f_b may be identified individually, f_a and f_b must be non-zero somewhere in their respective domains. Even so, identification problems result from rescaling, (since $f_a\cdot f_b = cf_a\cdot f_b/c$). We will assume $f_a(0) = 1$ which may be implemented by imposing $\hat{f}_a(0) = \sum_1^T \hat{c}_{at} r_{x_{at}}(0) = 1$. Other identifying restrictions are possible, such as $\|f_a\|_{Sob}^2 = 1$. However, in general, it is necessary to know some region where either f_a or f_b is of known sign, otherwise f_a, f_b are indistinguishable from $-f_a, -f_b$. Suppose the true regression function is given by $f_{ao}(x_a)\cdot f_{bo}(x_b)$, then:

PROPOSITION 5.4.2: *augment optimization problem (5.4.2) of Theorem 5.4.1 with the constraint* $\hat{f}_a(0) = \sum_1^T \hat{c}_{at} r_{x_{at}}(0) = 1$. *If* $m \geq 2$ *then* $sup_{x_a}|\hat{f}_a - f_{ao}| \overset{a.s.}{\to} 0$, $sup_{x_b}|\hat{f}_b - f_{bo}| \overset{a.s.}{\to} 0$. ■

The metric entropy for $\bar{\mathfrak{F}}$ is given by $\log N(\delta;\bar{\mathfrak{F}}) \leq A\ \delta^{-\max\{q_a,q_b\}/m}$. Proposition 5.1.1 may be applied to obtain a rate of convergence $r = \frac{2m}{2m+\max\{q_a,q_b\}}$ which in turn may be used in Theorem 5.2.1 to obtain a test of multiplicative separability.

A natural application in economics is to models that are homogeneous of degree k. Let x_a, x_b be scalars corresponding to levels of inputs and y the observed level of output. Define unrestricted and restricted function sets: $\mathfrak{F} = \{f(x_a, x_b/x_a) \in \mathcal{H}^m : \|f(x_a, x_b/x_a)\|_{Sob}^2 \leq L\}$ and $\bar{\mathfrak{F}} = \{ f(x_a, x_b/x_a) \in \mathcal{H}^m :\ f(x_a, x_b/x_a) = x_a^k g(x_b/x_a),\ \|x_a^k g(x_b/x_a)\|_{Sob}^2 \leq L\}$. Then the restricted optimization becomes: $\min_g 1/T \sum \left(y_t - x_{at}^k g(x_{bt}/x_{at})\right)^2$, $s.t.\ \|x_a^k\|_{Sob}^2\ \|g(x_b/x_a)\|_{Sob}^2 \leq L$. (See Theorem 2.4.)

5.5 Monotonicity and Concavity

The isotonic regression literature which in the simplest case considers least squares regression subject only to monotonicity constraints, goes back several decades, (see e.g., Barlow et al (1972) and Robertson et al (1988)). It differs from our setup in that we impose additional smoothness constraints. Monotonicity combined with smoothness assumptions has been studied by Mukarjee (1988), Mammen (1991) and Mukarjee and Stern (1994).[9] Consider the model $y = f(x) + \upsilon$, x is a scalar. Define $\mathfrak{S} = \{f \in \mathcal{H}^m: \|f\|^2_{Sob} \leq L\}$ and $\bar{\mathfrak{S}} = \{f \in \mathfrak{S}, f \textit{ non-decreasing}\}$. In examples considered in previous sections, it has been a simple matter to ensure that the restricted estimator $\hat{f}$ satisfies the additional properties everywhere in its domain. However, defining an estimator that is both smooth and monotone everywhere is more difficult. Instead, we will have it lie in descending sets $\bar{\mathfrak{S}}_T$ that converge to $\bar{\mathfrak{S}}$ as sample size increases.

Let $\bar{\mathfrak{S}}_T = closure\left\{f \in \mathcal{H}^m: \|f\|^2_{Sob} \leq L, f(x_s) \leq f(x_t), x_s \leq x_t, s, t, =1,...,T\right\}$ where $\{x_t\}_{t=1}^{\infty}$ is an infinite random sequence. Estimation takes place over the sets $\bar{\mathfrak{S}}_T$ which are random and which will, in general, include functions that are not monotone. Note that with probability one the sequence $\{x_t\}$ is dense in the domain and consider such a sequence. Bounded first derivatives combined with the imposition of monotonicity inequalities at $x_1,...,x_T$ ensure that for large enough T, $\bar{\mathfrak{S}}_T$ will contain only monotone functions and certain functions that are arbitrarily close (in supnorm) to a monotone function. Furthermore, $\bigcap_1^{\infty} \bar{\mathfrak{S}}_T = \bar{\mathfrak{S}}$ since if monotonicity restrictions are satisfied on a dense set, then given the smoothness of the family, they are satisfied everywhere. Proposition 5.1.1 may be applied. In addition, the following result holds:

PROPOSITION 5.5.2: *Let $\hat{f}_{unr}$ be the (smoothness constrained) estimator of Th. 3.1.1 and let $\hat{f}$ be the monotonicity constrained smooth estimator satisfying* $\min \frac{1}{T} \sum (y_t - f(x_t))^2$ *s.t.* $f \in \bar{\mathfrak{S}}_T$. *If f_o is strictly monotone on Q^1 and $m > 2$, then $Prob[\hat{f} = \hat{f}_{unr}] \to 1$ as $T \to \infty$.* ∎

Essentially, the corollary states that if the true regression function is strictly monotone then the monotonicity restrictions become nonbinding as sample size increases. In such cases, the constrained estimator has the same convergence rate as the unconstrained estimator in Theorem

3.1.2, i.e., $O_p(T^{-2m/(2m+1)})$. (Utreras (1984) and (Mammen (1991) find similar results using alternate monotonicity constrained nonparametric estimators.) Let R be the representor matrix at $x_1,\ldots,x_T$, then the monotonicity restricted optimization problem may be implemented by solving:

(5.5.1)
$$s^2 = \min_{c_1,\ldots c_T} \frac{1}{T}\sum [y_t - \hat{y}_t]^2 \quad \hat{y}_t = [Rc]_t \quad c'Rc \le L$$
$$\hat{y}_s \le \hat{y}_t \;\; for\; x_s \le x_t \quad s,t = 1,\ldots,T$$

where $c = (c_1,\ldots,c_T)$. Concavity restrictions may be implemented by solving:

(5.5.2)
$$s^2 = \min_{c_1,\ldots,c_T} \frac{1}{T}\sum [y_t - \hat{y}_t]^2 \quad \hat{y}_t = [Rc]_t \quad c'Rc \le L$$
$$\hat{y}_s \ge \frac{x_s - x_r}{x_t - x_r}\hat{y}_r + \frac{x_t - x_s}{x_t - x_r}\hat{y}_t \quad for\; x_r \le x_s \le x_t \quad r,s,t = 1,\ldots,T$$

If the regression function f_o is strictly concave, and if second derivatives are estimated consistently, (which will be the case if $m>3$), then the concavity constraints will become nonbinding in large samples and the rate of convergence will be no faster than in the absence of such functional structure. (See Matzkin (1994) for additional references on concave estimation.)

5.6 Homothetic Demand

Consider a two-equation regression model:

(5.6.1)
$$\begin{bmatrix} y_a \\ y_b \end{bmatrix} = \begin{bmatrix} f_a(x) \\ f_b(x) \end{bmatrix} + \begin{bmatrix} \upsilon_a \\ \upsilon_b \end{bmatrix} \qquad \begin{bmatrix} \upsilon_a \\ \upsilon_b \end{bmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; \Sigma_\upsilon = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \ldots & \sigma_b^2 \end{bmatrix}\right]$$

In section 3.2 we discussed estimation of such models subject to smoothness constraints. In the current section we add the constraints embodied by the homothetic demand model, (other constraints may also be implemented) but first we provide the obvious generalizations of asymptotic results on constrained estimation and testing.

ASSUMPTIONS FOR THE CONSTRAINED TWO EQUATION MODEL: *i) invoke the Assumptions for the Multi-Equation Model setting* $p = 2$, *(section 3.2);* $\mathfrak{F} = \mathfrak{F}_a \otimes \mathfrak{F}_b$, *is a cross-product of Sobolev balls* $\mathfrak{F}_a = \{f_a \in \mathcal{H}^m, \|f_a\|^2_{Sob} \le L_a\}$ *and* $\mathfrak{F}_b = \{f_b \in \mathcal{H}^m, \|f_b\|^2_{Sob} \le L_b\}$ *where* $m > \frac{q}{2}$; *ii)* $\bar{\mathfrak{F}} \subseteq \mathfrak{F}$ *is a closed set of vector functions; let* $\bar{\mathfrak{F}}_a = \{f_a | (f_a, f_b) \in \bar{\mathfrak{F}}\}$ *and* $\bar{\mathfrak{F}}_b = \{f_b | (f_a, f_b) \in \bar{\mathfrak{F}}\}$, *with metric entropies* $\log N(\delta; \bar{\mathfrak{F}}_a) \le A\, \delta^{-\zeta}$ *and* $\log N(\delta; \bar{\mathfrak{F}}_b) \le A\, \delta^{-\zeta}$ *for some* $A > 0, \zeta > 0$; *iii)* $\{\bar{\mathfrak{F}}_T\}_{T=1}^{\infty}$ *is a descending sequence of closed and possibly random sets of functions* $\mathfrak{F} \supset \bar{\mathfrak{F}}_1 \supset \ldots \supset \bar{\mathfrak{F}}_T \supset \ldots \supset \bar{\mathfrak{F}}$ *such that* $\bigcap_1^{\infty} \bar{\mathfrak{F}}_T = \bar{\mathfrak{F}}$ *a.s.; let* $\bar{\mathfrak{F}}_{aT} = \{f_a | (f_a, f_b) \in \bar{\mathfrak{F}}_T\}$, $\bar{\mathfrak{F}}_{bT} = \{f_b | (f_a, f_b) \in \bar{\mathfrak{F}}_T\}$ *with metric entropies satisfying* $\log N(\delta; \bar{\mathfrak{F}}_{aT}) \le A' \delta^{-\zeta}$ *and* $\log N(\delta; \bar{\mathfrak{F}}_{bT}) \le A' \delta^{-\zeta}$, $T = 1, \ldots \infty$ *for some* $A' > 0$. ■

Since $\bar{\mathfrak{F}}_{aT}$, $\bar{\mathfrak{F}}_{bT}$ are subsets of $\mathfrak{F}_a$, $\mathfrak{F}_b$ respectively, $\zeta \le \frac{q}{m}$ where q is the dimension of x , m is the order of the Sobolev norm. (See section 3.1.) Let $\hat{f}_a, \hat{f}_b$ satisfy:

$$\min_{(f_a, f_b) \in \bar{\mathfrak{F}}_T} \frac{1}{T} \sum_t \left(y_{at} - f_a(x_t), y_{bt} - f_b(x_t)\right) \Lambda \left(y_{at} - f_a(x_t), y_{bt} - f_b(x_t)\right)' \tag{5.6.2}$$

where Λ is a positive definite matrix and let $\hat{\Sigma}_{\upsilon}$ be the matrix of moments of the estimated residuals.

PROPOSITION 5.6.1: *If* $f_o \in \bar{\mathfrak{F}}$ *then the conclusions of Theorem 3.2.2 hold with rate of convergence* $r = \frac{2}{2+\zeta}$. *Furthermore, suppose* $f_o \notin \bar{\mathfrak{F}}$, $\|f_{ao}\|^2_{Sob}, \|f_{bo}\|^2_{Sob}$ *are finite and there exists a unique* $(\bar{f}_{ao}, \bar{f}_{bo}) \in \bar{\mathfrak{F}}$ *satisfying* $\min_{f \in \bar{\mathfrak{F}}} \int (f_{ao} - f_a, f_{bo} - f_b) \Lambda (f_{ao} - f_a, f_{bo} - f_b)' dP_x$. *Then* $\hat{\Sigma}_{\upsilon} \overset{a.s.}{\to} \Sigma_{\upsilon o} + \int (f_{ao} - \bar{f}_{ao}, f_{bo} - \bar{f}_{bo})'(f_{ao} - \bar{f}_{ao}, f_{bo} - \bar{f}_{bo}) dP_x$. ■

Analogously to equation (5.2.2) define the statistics:

$$\begin{aligned} U_a &= \frac{1}{T} \sum_t \left(y_{at} - \hat{f}_a(x_t)\right) \left[\frac{1}{\lambda^q T} \sum_{s \neq t} \left(y_{as} - \hat{f}_a(x_s)\right) K\left(\frac{x_s - x_t}{\lambda}\right) \right] \\ U_b &= \frac{1}{T} \sum_t \left(y_{bt} - \hat{f}_b(x_t)\right) \left[\frac{1}{\lambda^q T} \sum_{s \neq t} \left(y_{bs} - \hat{f}_b(x_s)\right) K\left(\frac{x_s - x_t}{\lambda}\right) \right] \end{aligned} \tag{5.6.3}$$

THEOREM 5.6.2: *Suppose* $(f_{ao}, f_{bo}) \in \bar{\mathfrak{S}}$, $\lambda^q T \to \infty$ *and* $\lambda^{q/2} T^{1-r} \to 0$ *where* $r = \frac{2}{2+\zeta}$. *Then*

$$(5.6.4) \qquad \lambda^{q/2} T \begin{bmatrix} U_a \\ U_b \end{bmatrix} \to N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \left(2\int p_x^2 \int K^2\right) \cdot \begin{bmatrix} \sigma_a^4 & \sigma_{ab}^2 \\ \dots & \sigma_b^4 \end{bmatrix}\right] \equiv N(0, \Omega)$$

Define estimators of the three distinct elements of Ω *by:*

$$(5.6.5) \qquad \begin{aligned} \hat{\Omega}_{aa} &= \frac{2}{\lambda^q T^2} \sum_t \sum_{s \neq t} \left(y_{at} - \hat{f}_a(x_t)\right)^2 \left(y_{as} - \hat{f}_a(x_s)\right)^2 K^2\left(\frac{x_s - x_t}{\lambda}\right) \\ \hat{\Omega}_{bb} &= \frac{2}{\lambda^q T^2} \sum_t \sum_{s \neq t} \left(y_{bt} - \hat{f}_b(x_t)\right)^2 \left(y_{bs} - \hat{f}_b(x_s)\right)^2 K^2\left(\frac{x_s - x_t}{\lambda}\right) \\ \hat{\Omega}_{ab} &= \frac{2}{\lambda^q T^2} \sum_t \sum_{s \neq t} \left(y_{at} - \hat{f}_a(x_t)\right)\left(y_{bt} - \hat{f}_b(x_t)\right)\left(y_{as} - \hat{f}_a(x_s)\right)\left(y_{bs} - \hat{f}_b(x_s)\right) K^2\left(\frac{x_s - x_t}{\lambda}\right) \end{aligned}$$

Then $\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_{aa} & \hat{\Omega}_{ab} \\ \hat{\Omega}_{ab} & \hat{\Omega}_{bb} \end{bmatrix} \xrightarrow{P} \Omega$ *and* $\lambda^q T^2 [U_a, \; U_b] \, \hat{\Omega}^{-1} \begin{bmatrix} U_a \\ U_b \end{bmatrix} \xrightarrow{D} \chi_2^2$. ∎

Generalizing the bootstrapping procedure described in section 5.2 in the obvious fashion (note that one samples from the joint empirical distribution of the estimated residuals), we have:

THEOREM 5.6.3: $\lambda^q T^2 \left[U_a^B, \; U_b^B\right] \left[\hat{\Omega}^B\right]^{-1} \begin{bmatrix} U_a^B \\ U_b^B \end{bmatrix} \xrightarrow{D} \chi_2^2$ ∎

We now apply these results to a homothetic demand model where y_a, y_b are demands for each of two goods, $x = (x_a, x_b)$ are corresponding prices and income is fixed at 10. Homotheticity implies that expenditure share functions $\pi_a(\cdot), \pi_b(\cdot)$ depend only on relative prices. The budget constraint requires $\pi_a + \pi_b = 1$. The remaining constraints implied by the maximization hypothesis require that the ratios of demands are a monotone function of relative prices. [10] Thus, the restricted model incorporates multiplicative separability, a cross-equation monotonicity restriction and smoothness. Results of simulations are in Figure 4.

Figure 4: Conditional Moment Test of a Homothetic Demand Model

Data Generating Mechanism:

$$y_a = f_a(x_a, x_b) = .5*10/x_a + \upsilon_a$$
$$y_b = f_b(x_a, x_b) = .5*10/x_b + \upsilon_b$$
$$\begin{bmatrix}\upsilon_a \\ \upsilon_b\end{bmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}.25 & 0. \\ 0. & .25\end{pmatrix}\right]$$

price vectors (x_a, x_b) lie on a uniform grid in $[2,4]x[2,4]$; x_a/x_b lies in the interval $[.5, 2]$; $\|f_a\|^2_{Sob} = \|f_b\|^2_{Sob} = 23.76$;[i]

Hypotheses:

$H_o: f_a, f_b$ are smooth homothetic demand functions

$H_1: f_a, f_b$ are smooth functions

Restricted Estimator: multivariate Sobolev least squares on the model:[ii]

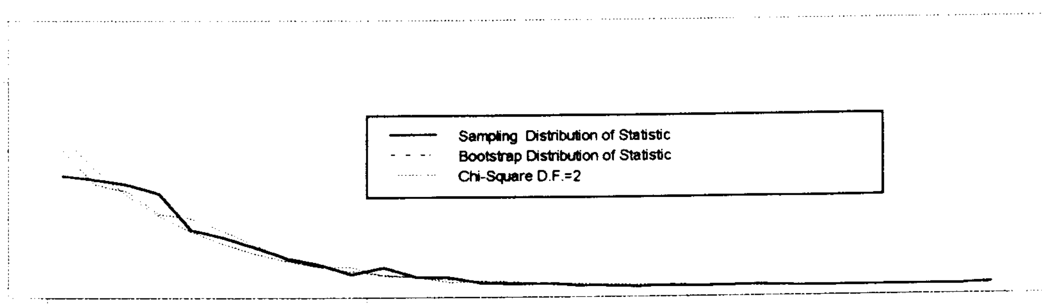
$$\begin{bmatrix}y_a \\ y_b\end{bmatrix} = \begin{bmatrix}\pi_a(x_a/x_b)\,10/x_a + \upsilon_a \\ \pi_b(x_a/x_b)\,10/x_b + \upsilon_b\end{bmatrix}$$

subject to: smoothness constraints: $\|\pi_a\|^2_{Sob} \le .4$ $\|\pi_b\|^2_{Sob} \le .4$; budget constraints: $\pi_a(x_a/x_b) + \pi_b(x_a/x_b) = 1$; and monotonicity constraints:

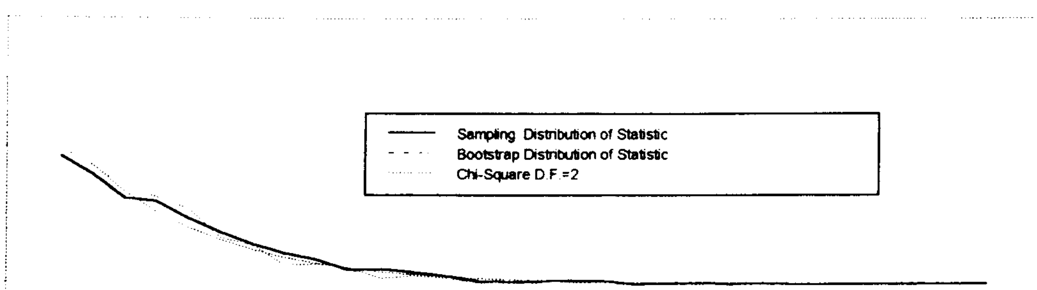
$$\frac{\pi_a(x_{at}/x_{bt})\,10/x_{at}}{\pi_b(x_{at}/x_{bt})\,10/x_{bt}} \ge \frac{\pi_a(x_{as}/x_{bs})\,10/x_{as}}{\pi_b(x_{as}/x_{bs})\,10/x_{bs}} \quad \text{whenever} \quad \frac{x_{bt}}{x_{at}} \ge \frac{x_{bs}}{x_{as}}$$

Test Statistic: $\lambda^2 T^2 [U_a, \ U_b]\, \hat{\Omega}^{-1} \begin{bmatrix}U_a \\ U_b\end{bmatrix} \sim \chi^2_2$ under H_o.[iii]

T=25, $\lambda = 2T^{-1/5} = 1.051$



T=49, $\lambda = 2T^{-1/5} = .918$



[i] For the fourth order Sobolev norm, calculation yields $\|1/z\|^2_{Sob} = .4752$, if $2 < z < 4$. Hence $\|f_a\|^2_{Sob} = \|f_b\|^2_{Sob} = .4752 \int_2^4 (.5*10)^2 = 23.76$. See comments following Th.2.4.

[ii] π_a, π_b are expenditure shares, $\pi_a(\cdot)=.5$, $\pi_b(\cdot)=.5$ on [.5, 2]; $\|\pi_a\|^2_{Sob} = \|\pi_b\|^2_{Sob} = .5^2(2-.5) = .375$ Objective function: $1/T\sum (y_{at}-\pi_a(x_{at}/x_{bt}))^2 + (y_{bt}-\pi_b(x_{at}/x_{bt}))^2$.

[iii] See Th. 5.6.2. A uniform product kernel is used in the computation of U_a, U_b. Since $\Omega_{ab}=0$ we set $\hat{\Omega}_{ab}=0$. Sobolev fourth order norm is used (m=4 in equation (2.1)). From Theorem 5.6.2 we need $\lambda^{q/2}T^{1-r} \to 0$. Since each equation depends nonparametrically on only one variable, (relative prices), $r = 2m/(2m+1) = 8/9$, (Proposition 5.6.1). We set $\lambda = 2T^{-1/5}$. The sampling distribution of the test statistic is based on 1000 samples. The bootstrap distribution is based on 1000 resamples from a single initial sample. Fortran code was used to calculate the representor matrix. GAMS, (Brooke et al (1992)), was used to solve the restricted optimization problem (about 1.5 seconds per optimization for T=25 on a 90 MHz Pentium processor, 8 seconds for T=49).

More general versions of the maximization hypothesis may be imposed by adding the Afriat (1967,1973) inequalities to the optimization problem, (see Epstein and Yatchew (1985)).[11] Estimation takes place over a descending sequence of sets of functions which satisfy Afriat inequalities at points where data are observed.

6. CONCLUDING REMARKS

Many bodies of theory are incomplete without empirically determined constants. Economic theory is in an unenviable position in that these unknowns are often functions, constrained only by certain functional properties. Hence (constrained) nonparametric estimation finds a natural home in economics. Given the curse of dimensionality, the single most important contribution that economic theory makes to the nonparametric empirical exercise is in limiting the number of distinct variables entering into the regression function. Other restrictions that enhance the rate of convergence include forms of separability and various semiparametric specifications. In large samples and given sufficient smoothness assumptions, functional restrictions such as monotonicity and concavity are of lesser importance in that they do not improve the rate of convergence, but in small samples or in the absence of smoothness assumptions, they can be beneficial.

One of the reasons that nonparametric techniques have not seen wider use in applied work is the absence of a unified framework for constrained estimation and particularly testing of hypotheses. This paper describes a class of estimators, closely related to spline estimators, which can accommodate a variety of restrictions. A conditional moment test procedure which generalizes the work of other authors, and is useful for testing a broad range of hypotheses, is also provided.

FOOTNOTES

1 The authors are grateful to the SSHRC for its support and to many individuals but especially to Larry Epstein, Zvi Griliches, Angelo Melino, Per Mykland, Peter Robinson and Sara Van de Geer. Thanks are also due to Alex Meeraus and Peter Steacy at the GAMS Development Corporation. This paper is dedicated to Mrs. Jeanette DeHaan.

2 For a recent review of Hilbert space methods in probability and statistics see Small and McLeish (1994).

3 See e.g., Kolmogorov's theorem (Dudley (1984, p.51)), relating metric entropy of a class of functions to the number of bounded derivatives and dimension of the explanatory variable.

4 Penalized least squares optimizes a criterion function which balances fidelity of the regression function to the data against smoothness of the regression function. The 'penalty' parameter for smoothness is typically selected using crossvalidation. In our set-up, we select a bound on the smoothness of the regression function using cross validation, then maximize fit subject to the constraint that the smoothness of the estimated regression function not exceed the selected bound. A further difference between the approaches lies in the measure of smoothness. In penalized least squares, smoothness is typically measured using a semi-norm such as the integrated squared second derivative. We use a Sobolev norm of sufficient order to permit the derivation of certain asymptotic distribution results. If one solves our Lagrangian problem $\min_{f,\lambda} 1/T\,\Sigma(y_t - f(x_t))^2 - \lambda\left[\|f\|^2_{Sob} - L\right]$, then uses the resulting $\hat{\lambda}$ in a penalized least squares procedure with the same Sobolev norm, the identical minimum sum of squares will result.

5 Specific examples include tests of sphericity, confidence regions for eigenvalues of Σ, (which may be used in principal component analysis), factor analysis and various particular structures arising from heterogeneity of the residuals. See Shao and Tu (1995, pp.373-383) for references and an overview of related bootstrap inference.

6 The authors have investigated an alternative test procedure. For the purposes of this footnote define s^2_{unr} to be the solution to the 'unrestricted' optimization problem of Th. 3.1.1. Define s^2_{res} to be the solution to any of the 'restricted' optimization problems in Section 5, (e.g., additive or multiplicative separability, monotonicity, concavity or homothetic demand). Then $T^{1/2}\left(s^2_{res} - s^2_{unr}\right) \to 0$ which precludes the use of this statistic (without re-normalization) for testing purposes. The degeneracy can be circumvented by sample splitting (see e.g. Yatchew (1992), but not without loss in efficiency, (Hong and White (1995)). In simulations performed by the authors, the analogue to the conventional F- statistic, given by $F = T\left(s^2_{res} - s^2_{unr}\right)/\,s^2_{unr}$ was found to perform well when critical values were obtained using bootstrap procedures, (s^2_{unr} and s^2_{res} were computed without sample splitting). Simulations were performed for tests of specification, significance, additive separability, monotonicity and the homothetic demand model. Results are available from the authors. For related work see Cleveland and Devlin (1988).

7 See Fan and Li (1996) for significance and specification tests using the statistic in equation (5.2.2) where the restricted estimator is based on kernel methods.

8 Although (5.2.2) is a kind of U-statistic, studied by Hall (1984) and De Jong (1987), we will not use those results but instead rely directly on triangular array results for dependent processes to obtain our distributional results. This will simplify subsequent demonstration of the validity of the bootstrap.

9 See also Wright and Wegman (1980), Wong (1984), Villalobas and Wahba (1987), Ramsay (1988) and Goldman and Ruud (1992).

10 To see that these conditions are sufficient, first note that with two goods, data that satisfy the weak axiom of revealed preference are consistent with the maximization hypothesis (Varian (1984, p.143)). Then verify that with the budget constraints and the monotonicity constraints in place, the weak axiom cannot be violated.

11 See also Diewert (1973) and Varian (1982, 1983, 1988). Other work in this area includes Yatchew (1985), Matzkin (1986, 1991, 1994) and Matzkin and Richter (1991). For references on a parallel literature in producer theory, see Matzkin (1994).

BIBLIOGRAPHY

Adams, R. (1975): *Sobolev Spaces*, Academic Press, N.Y.

Afriat, S. (1967): "The Construction of a Utility Function From Expenditure Data", *International Economic Review*, 8, 66-77.

Afriat, S. (1973): "On a System of Inequalities in Demand Analysis", *International Economic Review*, 14, 460-472.

Andrews, D.W.K., (1994a): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity", *Econometrica*, 62, 43-72.

Andrews, D.W.K., (1994b): "Empirical Process Methods in Econometrics", in the *Handbook of Econometrics, Vol. IV*, R. Engle and D. McFadden, eds., North Holland, 2247-2294.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972): *Statistical Inference Under Order Restrictions*, New York: John Wiley.

Bierens, H.J. (1982): "Consistent Model Specification Tests", *Journal of Econometrics*, 20, 105-134.

Bierens, H. (1990): "A Consistent Conditional Moment Test of Functional Form", *Econometrica*, 58, 1443-1458.

Brooke, A., D. Kendrick and A. Meeraus (1992): *GAMS*, Scientific Press, Redwood City, California.

Chen H. (1988): "Convergence Rates for Parametric Components in a Partly Linear Model", *Annals of Statistics*, 16, 136-146.

Chen H. (1991): "A Two Stage Spline Smoothing Method for Partially Linear Models", *Journal of Statistical Planning and Inference*, 27, 187-201.

Chen H. and J.-J. H. Shiau (1994): "Data Driven Efficient Estimators for a Partially Linear Model", *Annals of Statistics*, 22, 211-237.

Cleveland W.S. and S.J. Devlin (1988): "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association*, 83, 596-610.

De Jong, P., (1987), "A Central Limit Theorem for Generalized Quadratic Forms", *Probab. Th. Rel. Fields*, 75, 261-277.

Delgado, M.A. and P.M. Robinson (1992) "Nonparametric and Semiparametric Methods for Economic Research", *Journal of Economic Surveys*, Vol.6, No.3, 201-249.

Diewert, E.W., (1973), "Afriat and Revealed Preference Theory", *Review of Economic Studies*, 40, 419-426.

Dudley, R.M. (1984): "A Course on Empirical Processes", *Lecture Notes in Mathematics*, Ecole d`Été de Probabilités de Saint-Flour XII-1982, New York: Springer-Verlag.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, pp.1-26.

Engle, R., C.Granger, J.Rice and A.Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales", *Journal of the American Statistical Association*, 81, 310-320.

Epstein, L. and A.J. Yatchew (1985): "Nonparametric Hypothesis Testing Procedures and Applications to Demand Analysis", *Journal of Econometrics*, 30, pp. 150-169.

Eubank L. and C.H. Spiegelman (1990): "Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Techniques", *Journal of the American Statistical Association*, 85, 387-392.

Eubank R.L. and P. Whitney (1989) "Convergence Rates for Estimation for Certain Partially Linear Models", *Journal of Statistical Planning and Inference*, 23, 33-43.

Fan Y. and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms", *Econometrica*, 64, 865-90.

Gallant, A.R. (1982): "Unbiased Determination of Production Technologies", *Journal of Econometrics*, 20, 285-323.

Goldman S. and P. Ruud, (1992), "Nonparametric Multivariate Regression Subject to Monotonicity and Convexity Constraints", manuscript, University of California, Berkeley.

Golub, G.H., and C.F. Van Loan, (1989): *Matrix Computations*, Second Edition, Baltimore and London: John Hopkins University Press.

Gozalo, P.L., (1993): "A Consistent Specification Test for Nonparametric Estimation of regression Function Models", *Econometric Theory*, 9, 451-77.

Hall, P., (1984): "Central Limit Theorem For Integrated Square Error of Multivariate Nonparametric Density Estimators", *Journal of Multivariate Analysis*, 14, 1-16.

Härdle W. (1990): *Applied Nonparametric Regression*, Econometric Society Monograph Series, 19, Cambridge University Press.

Härdle W. and O. Linton, (1994), "Applied Nonparametric Methods", in the *Handbook of Econometrics, Vol. IV*, R. Engle and D. McFadden, eds., North Holland, 2297-2334.

Härdle W. and E. Mammen, (1993), "Comparing Nonparametric vs Parametric Regression Fits", *Annals of Statistics*, 21, 1926-47.

Hastie, T. J. and R. Tibshirani, (1990), *Generalized Additive Models*, London: Chapman and Hall.

Hausman J., (1978), "Specification Tests in Econometrics", *Econometrica*, 46, 1251-72.

Hausman J., W. Newey, (1995), "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss", *Econometrica*, 63, 1445-76.

Heckman, N.E. (1986): "Spline Smoothing in a Partly Linear Model", *Journal of the Royal Statistical Society, Series B*, 48, 244-248.

Hong, Y. and H. White, (1995): "Consistent Specification Testing via Nonparametric Series Regression" *Econometrica*, 63, 1133-1160.

Horowitz, J. and W. Härdle, (1994): "Testing a Parametric Model Against a Semiparametric Alternative", *Econometric Theory*, 10, 821-48.

Kolmogorov, A.N. and Tichomirov, V.M., (1959): "ε-Entropy and ε-Capacity of Sets in Function Spaces", *Uspechi Math. Nauk*, , 14, 3-86, (English Translation: American Mathematical Society Transl. 2 (1961), 17, 277-364).

Lee, B.J. (1991): "A Model Specification Test Against the Nonparametric Alternative", Department of Economics, University of Colorado, manuscript.

Li, Ker-Chau (1986): "Asymptotic Optimality of C_L and Generalized Cross-Validation in Ridge Regression With Application to Spline Smoothing", *Annals of Statistics*, 14, 1101-12.

Li, Ker-Chau (1987): "Asymptotic Optimality for C_P, C_L, Cross-Validation and Generalized Cross-Validation: Discrete Index Set", *Annals of Statistics*, 15, 958-75.

Li, Qi (1994): "Some Simple Consistent Tests for a Parametric Regression Function versus Semiparametric or Nonparametric Alternatives", Department of Economics, University of Guelph, manuscript.

Linton, O. (1995a): "Second Order Approximation in the Partially Linear Regression Model", *Econometrica*, 63, 1079-1112.

Linton, O. (1995b): "Estimation in Semiparametric Models: A Review", in *Advances in Econometrics and Quantitative Economics, Essays in Honour of Professor C.R. Rao*, 146-171.

Mammen, E. (1991): "Estimating a Smooth Monotone Regression Function", *Annals of Statistics*, 19, 724-740.

Mammen E., S. Van de Geer (1995): "Penalized Quasi-likelihood Estimation in Partial Linear Models", manuscript, Humboldt-Universitat Berlin and University of Leiden.

Matzkin, R.L., (1986): "Mathematical and Statistical Inferences from Demand Data", Ph.D. Dissertation, University of Minnesota.

Matzkin, R.L., (1991): "Axioms of Revealed Preference for Nonlinear Choice Sets", *Econometrica*, 59, 1779-1786.

Matzkin, R.L., and M.K. Richter, (1991): "Testing Strictly Concave Rationality", *Journal of Economic Theory*, 53, 287-303.

Matzkin, R.L., (1994), "Restrictions of Economic Theory in Nonparametric Methods", in the *Handbook of Econometrics, Vol. IV*, R. Engle and D. McFadden, eds., North Holland, 2524-2559.

McLeish, D.L. (1974): "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability*, 2, 620-8.

H. Mukarjee, (1988): "Monotone Nonparametric Regression", *Annals of Statistics*, 16, 741-750.

H. Mukarjee and S. Stern (1994): "Feasible Nonparametric Estimation of Multi-argument Monotone Functions", *Journal of the American Statistical Association*, 89, 77-80.

Pinske C., and P. Robinson (1995): "Pooling Nonparametric Estimates of Regression Functions with a Similar Shape", in *Advances in Econometrics and Quantitative Economics, Essays in Honour of Professor C.R. Rao*, 172-197.

Pollard, D. (1984): *Convergence of Stochastic Processes*, New York: Springer-Verlag.

Powell, J. (1994), "Estimation of Semiparametric Models", in the *Handbook of Econometrics, Vol. IV*, R. Engle and D. McFadden, eds., North Holland, 2444-2523.

Ramsay, J. (1988): "Monotone Regression Splines in Action", *Statistical Science*, 3, 425-61.

Ranga Rao, R. (1962): "Relations Between Weak and Uniform Convergence of Measures With Applications", *Annals of Mathematical Statistics*, 33, pp. 659-680.

Rice, J. (1986): "Convergence Rates for Partially Splined Models", *Statistical and Probability Letters*, 4, 203-208.

Robertson, T., F.T. Wright and R.L. Dykstra (1988): *Order Restricted Statistical Inference*, New York: John Wiley and Sons.

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

Robinson, P.M. (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Economic Time Series", *Review of Economic Studies*, 56, pp. 511-534.

Shao, J. and D. Tu (1995): *The Jackknife and Bootstrap*, Springer.

Shiau, J.-J. H. and G. Wahba (1988): "Rates of Convergence of Some Estimators for a Semiparametric Model", *Communications in Statistics -- Simulation*, 17(4), 1117-1113.

Small, C. and D. McLeish, (1994): *Hilbert Space Methods in Probability and Statistical Inference*, New York: John Wiley and Sons.

Speckman, P. (1988): "Kernel Smoothing in Partial Linear Models", *Journal of the Royal Statistical Society B*, 50, 413-436.

Stoker, T. (1989): "Tests of Additive Derivative Constraints", *Review of Economic Studies*, 56, 535-552.

Stoker, T. (1991): *Lectures on Semiparametric Econometrics*, CORE Foundation Louvain-La-Neuve.

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

Stone, C. (1985): "Additive Regression and Other Nonparametric Models", *Annals of Statistics*, 13, 689-705.

Stone, C. (1986): "The Dimensionality Reduction Principle for Generalized Additive Models", *Annals of Statistics*, 14,590-606.

Utreras, F. (1984): "Smoothing Noisy Data Under Monotonicity Constraints: Existence, Characterization and Convergence Rates", *Numer. Math.*, 47, 611-625.

Van de Geer, S., (1987): "A New Approach to Least Squares Estimation With Applications", *Annals of Statistics*, Vol. 15, No. 2, pp. 587-602.

Van de Geer, S., (1990): "Estimating a Regression Function", *Annals of Statistics*, Vol. 18, No. 2, pp. 907-924.

Varian, H. (1982): "The Nonparametric Approach to Demand Analysis", *Econometrica*, 50, 945-972 .

Varian, H. (1983): "Nonparametric Test of Models of Consumer Behaviour", *Review of Economic Studies*, 50, 99-110.

Varian, H. (1984): *Microeconomic Analysis*, Second Edition, New York, London: W.W. Norton and Company.

Varian, H. (1985): "Nonparametric Analysis of Optimizing Behaviour with Measurement Error", *Journal of Econometrics*, 30.

Varian, H. (1988): "Revealed Preference With a Subset of Goods", *Journal of Economic Theory*, 46,179-185.

Varian, H. (1990): "Goodness of Fit in Optimizing Models", *Journal of Econometrics*, 46, 141-163.

Villalobos, M. and G. Wahba (1987): "Inequality-Constrained Multivariate Smoothing Splines With Application to the Estimation of Posterior Probabilities", *Journal of the American Statistical Association*", 82, 239-248.

Wahba, G. (1984): "Partial Spline Models for the Semiparametic Estimation of Functions of Several Variables", in *Statistical Analysis of Time Series Data*, Proceedings of the Japan U.S. Joint Seminar, Tokyo, 319-329.

Wahba, G. (1985): "Multivariate Thin Plate Spline Smoothing with Positivity and Other Linear Inequality Constraints", in *Statistical Image Processing and Graphics*, E. Wegman and D. dePriest, eds., Marcel Dekker, New York, 275-290.

Wahba, G. (1990): *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, #59, Society for Industrial and Applied Mathematics.

Whang Y.-J. and D.W.K. Andrews (1991): "Tests of Specification for Parametric and Semiparametric Models", *Journal of Econometrics*, 57, 277-318.

White, H. (1994): *Estimation, Inference and Specification Analysis*, Cambridge University Press.

Wooldridge, J.M. (1992): "A Test for Functional Form Against Nonparametric Alternatives", *Econometric Theory*, 8, 452-75.

Wong, W. (1984): "On Constrained Multivariate Spline and Their Approximations", *Numer. Math.*, 43, 141-152.

Wright, I. and E. Wegman, (1980): "Isotonic, Convex and Related Splines", *Annals of Statistics*", 8, 1023-1035.
Yatchew, A., (1985): "A Note on Nonparametric Tests of Consumer Behaviour", *Economic Letters*, 18, 45-48.
Yatchew, A.J. (1988): "Some Tests of Nonparametric Regressions Models", *Dynamic Econometric Modelling*, Proceedings of the Third International Symposium in Economic Theory and Econometrics, W. Barnett, E. Berndt and H. White (eds.), Cambridge University Press, 121-135.
Yatchew, A.J., (1992): "Nonparametric Regression Model Tests Based on Least Squares", *Econometric Theory*, 8, 435-451.
Yatchew, A.J., (1997a): "Nonparametric Regression Techniques in Economics", forthcoming, *Journal of Economic Literature*.
Yatchew, A.J., (1997b): "An Elementary Estimator of the Partial Linear Model", forthcoming, *Economics Letters*.

APPENDIX 1 - PROOFS

PROOF OF THEOREM 2.1: It is straightforward to verify that $\mathcal{H}^m$ is a vector space and that equation (2.1) defines an inner product on it. Furthermore, $\mathcal{H}^m$ is complete by construction. ∎

PROOF OF THEOREM 2.2: For functions of one variable, derivations of representors are well known. Details are contained in Appendix 2. For functions of several variables, the proof proceeds by producing the representor r_a^b.

Let $r_a^b(x) = \prod_{i=1}^{q} r_{a_i}^{b_i}(x_i)$ for all $x \in Q^q$, where $r_{a_i}^{b_i}(x_i)$ is the representor at a_i in $\mathcal{H}^m(Q^1)$. Since C^m is dense in $\mathcal{H}^m$ it is sufficient to show the result for $f \in C^m$. We have then

$$\left\langle r_a^b, f \right\rangle_{Sob} = \left\langle \prod r_{a_i}^{b_i}, f \right\rangle_{Sob} = \sum_{|\alpha|_\infty \le m} \int_{Q^q} \frac{\partial^{\alpha_1} r_{a_1}^{b_1}(x_1)}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_q} r_{a_q}^{b_q}(x_q)}{\partial x_q^{\alpha_q}} D^\alpha f(x)\, dx$$

which may be rewritten as

$$= \sum_{i_1,\ldots,i_q = 0,\ldots m} \int_{Q^q} \frac{\partial^{i_1} r_{a_1}^{b_1}(x_1)}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_q} r_{a_q}^{b_q}(x_q)}{\partial x_q^{i_q}} \frac{\partial^{i_1+\ldots+i_q} f(x)}{\partial x_1^{i_1} \ldots \partial x_q^{i_q}}\, dx$$

$$= \left[\sum_{i_1=0}^{m} \int_0^1 \frac{\partial^{i_1} r_{a_1}^{b_1}(x_1)}{\partial x_1^{i_1}} \left[\sum_{i_2=0}^{m} \int_0^1 \frac{\partial^{i_2} r_{a_2}^{b_2}(x_2)}{\partial x_2^{i_2}} \left[\cdots \left[\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q} r_{a_q}^{b_q}(x_q)}{\partial x_q^{i_q}} \cdot \frac{\partial^{i_1+\ldots+i_q} f(x)}{\partial x_1^{i_1} \ldots \partial x_q^{i_q}}\, dx_q \right] .. dx_2 \right] dx_1 \right] \right]$$

Consider the centermost bracket. Since $\partial^{i_1+\ldots+i_{q-1}} f(x) / \partial x_1^{i_1} \ldots \partial x_{q-1}^{i_{q-1}}$ viewed as a function of x_q, holding $x_1,\ldots,x_{q-1}$ constant is a function in $\mathcal{H}^m(Q^1)$ and $r_{a_q}^{b_q}(x_q)$ is the representor for the b_q-th derivative at a_q in $\mathcal{H}^m(Q^1)$, the quantity within the centermost square bracket is equal to $\partial^{b_q} f(x_1,\ldots,x_{q-1},a_q) / \partial x_q^{b_q}$. Proceeding in this way we obtain the value for the whole expression to be equal to $D^b f(a)$. ∎

PROOF OF THEOREM 2.3: In one variable, we may appeal to the standard case, (Adams (1975), Theorem 6.2, p. 144), which states that the imbedding $\mathcal{H}^m(Q^1) \to C^{m-1}(Q^1)$ is compact. In particular, for $0 \le b \le m-1$, and $a \in Q^1$, the mapping $f \to d^b f(a)/dx^b$ is a bounded linear functional on $\mathcal{H}^m(Q^1)$. By the Riesz-Frechet Representation Theorem there is a representor $r_a^b \in \mathcal{H}^m(Q^1)$ such that for all $f \in \mathcal{H}^m(Q^1)$, $<r_a^b, f>_{Sob} = d^b f(a)/dx^b$. Moreover, since compact sets are bounded, there is an M>0 such that if $\|f\|_{Sob} \le 1$ then $\|f\|_{\infty,m-1} \le M$. Hence, we have $\|f\|_{Sob} \le 1 \Rightarrow |<r_a^b, f>_{Sob}| \le M$ for $0 \le b \le m-1$ and all $a \in Q^1$. In which case:

(A.1)
$$\left| \left\langle r_a^b, \frac{r_a^b}{\|r_a^b\|_{Sob}} \right\rangle_{Sob} \right| = \| r_a^b \|_{Sob} \le M$$

Furthermore, it is not difficult to see by the construction of these representors, (see Appendix 2), that the mapping: $Q^1 \to \mathcal{H}^m(Q^1)$ given by $a \to r_a^b$ is continuous for $0 \le b \le m-1$.

Consider the several variable case and define $r_a^b(x) := r_{a_1}^{b_1}(x_1) \ldots r_{a_n}^{b_n}(x_n)$. By Theorem 2.2 $<r_a^b, f>_{Sob} = D^b f(a)$ for all $f \in \mathcal{H}^m(Q)$ and b such that $|b|_\infty \le m-1$. It is straightforward to show that:

(A.2)
$$\|r_a^b(x)\|_{Sob} = \|r_{a_1}^{b_1}(x_1)\|_{Sob} \cdot \cdot \cdot \cdot \|r_{a_q}^{b_q}(x_q)\|_{Sob}$$

Furthermore, as we have indicated in the text, $C^m(Q^q)$ is a complete normed linear space, i.e., a Banach space. We will first show that $\mathcal{H}^m(Q^q)$ is imbedded in $C^{m-1}(Q^q)$, i.e., $f \in \mathcal{H}^m(Q^q)$ implies that $f \in C^{m-1}(Q^q)$. Let $f \in \mathcal{H}^m(Q^q)$. By definition there is a sequence $\{f_t\} \subset C^m(Q^q)$ such that $\|f - f_t\|_{Sob} \to 0$. We wish to show that $\{f_t\}$ is convergent in $C^{m-1}(Q^q)$ which we do by showing that it is Cauchy in this space. But we know $\{f_t\}$ is Cauchy in $\mathcal{H}^m(Q^q)$. Hence, given $\varepsilon > 0$, choose T large enough so that $t_1, t_2 > T$ implies $\|f_{t_1} - f_{t_2}\|_{Sob} < \dfrac{\varepsilon}{m^q M^q}$. Then for $t_1, t_2 > T$ we have:

$$\begin{aligned}
\|f_{t_1} - f_{t_2}\|_{\infty,m-1} &= \sum_{|\alpha|_\infty \le m-1} \max_{x \in Q^q} |D^\alpha f_{t_1}(x) - D^\alpha f_{t_2}(x)| \\
&= \sum_{|\alpha|_\infty \le m-1} \max_{x \in Q^q} |<r_x^\alpha, f_{t_1} - f_{t_2}>_{Sob}| \qquad \textit{using Theorem } 2.2
\end{aligned}$$

$$\leq \sum_{|\alpha|_\infty \leq m-1} \max_{x \in Q^q} \|r_x^\alpha\|_{Sob} \|f_{t_1} - f_{t_2}\|_{Sob} \quad \textit{using Cauchy Schwartz}$$

$$< m^q M^q \frac{\varepsilon}{m^q M^q} = \varepsilon$$

The last inequality follows by noting that $\|r_x^\alpha\|_{Sob} \leq M^q$, using (A.1) and (A.2) above and noting that there are m^q elements in the summation, (each α_i taking on values $0,...,m-1$). Hence, we have that $\{f_t\}$ is indeed Cauchy in $C^{m-1}(Q^q)$. But as $C^{m-1}(Q^q)$ is complete, there exists $g \in C^{m-1}(Q^q)$ such that $f_t \to g$ in $C^{m-1}(Q^q)$, that is, $D^b f_t \to D^b g$ uniformly for all $|b|_\infty \leq m-1$. Then $D^b f_t \to D^b g$ in $L_2(Q^q)$ for all $|b|_\infty \leq m-1$, or in other words, $f_t \to g$ in $\mathcal{H}^{m-1}(Q^q)$. But by definition, $f_t \to f$ in $\mathcal{H}^m(Q^q)$ and so automatically, $f_t \to f$ in $\mathcal{H}^{m-1}(Q^q)$. Hence, by uniqueness, $f = g$, $f \in C^{m-1}(Q^q)$.

To show compactness of the imbedding, we proceed by induction on m. Consider $m=1$. In this case, we must show that if $\{f_t\}$ is a bounded sequence in $\mathcal{H}^1(Q^q)$ then there is a subsequence that converges in $C^0(Q^q)$. But by the Arzela-Ascoli Theorem it suffices to show that $\{f_t\}$ is equicontinuous. This follows easily from the fact that

$$|f_t(a) - f_t(a')| = |<r_a - r_{a'}, f_t>_{Sob}| \leq \|r_a - r_{a'}\|_{Sob} \|f_t\|_{Sob}$$

and that r_a depends continuously on a. Now suppose that the imbedding $\mathcal{H}^m(Q^q) \to C^{m-1}(Q^q)$ is compact for particular m and consider $m+1$. Let $\{f_t\}$ be a bounded sequence in $\mathcal{H}^{m+1}(Q^q)$. Then $\{D^\alpha f_t\}$ is bounded in $\mathcal{H}^m(Q^q)$ for each α with $|\alpha|_\infty \leq 1$. Hence for *each* $|\alpha|_\infty \leq 1$ there is a subsequence of $\{D^\alpha f_t\}$ which converges in $C^{m-1}(Q^q)$. By passing to subsequences (of these 2^n sequences) we may extract a single subsequence $\{D^\alpha f_{t'}\}$ which converges in $C^{m-1}(Q^q)$ for all $|\alpha|_\infty \leq 1$. That is, $\{f_{t'}\}$ converges in $C^{m-1}(Q^q)$. ■

PROOF OF THEOREM 2.4: Both results can be proved by expanding the norm defined in (2.1). Note that for additively separable functions, $\{f_a(x_a)\}$ and $\{f_b(x_b) \mid \int f_b = 0\}$ are orthogonal in the Sobolev space $\mathcal{H}^m$ of functions $f(x_a, x_b)$. ■

PROOF OF THEOREM 3.1.1: Let $M = span\{r_{x_t} : t=1,...,T\}$, $M^{\perp} = \{h \in \mathcal{H}^m : <r_{x_t}, h>_{Sob} = 0, \forall t\}$. Representors exist by Theorem 2.2 and we can write $\mathcal{H}^m = M \oplus M^{\perp}$ since $\mathcal{H}^m$ is a Hilbert space. Note that the functions $h \in M^{\perp}$ take on the value zero at $x_1,...,x_T$. Consider a function $f \in \mathcal{H}^m$. Write $f = \sum_{j=1}^{T} c_j r_{x_j} + h$, $h \in M^{\perp}$. Then

$$\begin{aligned}\sum_t [y_t - f(x_t)]^2 &= \sum_t \Big[y_t - < r_{x_t}, \sum_j c_j r_{x_j} + h >_{Sob}\Big]^2 = \sum_t \Big[y_t - \sum_j < r_{x_t}, c_j r_{x_j} >_{Sob}\Big]^2 \\ &= \sum_t \Big[y_t - \sum_j c_j < r_{x_t}, r_{x_j} >_{Sob}\Big]^2 = \sum_t \Big[y_t - \sum_j R_{tj} c_j\Big]^2 = [y - Rc]'[y - Rc]\end{aligned}$$

Note further that

$$\| f \|_{Sob}^2 = < f, f >_{Sob} = \left\langle \sum_j c_j r_{x_j}, \sum_j c_j r_{x_j} \right\rangle_{Sob} + < h, h >_{Sob} = c'Rc + < h, h >_{Sob}$$

Suppose that $f = \sum_{j=1}^{T} c_j r_{x_j} + h$ minimizes $\frac{1}{T} \sum [y_t - f(x_t)]^2$ $s.t.$ $\| f \|_{Sob}^2 \leq L$ then so does $f^* = f - h$ since h is zero at $x_1,...,x_T$. Hence, there exists a function f^* minimizing the infinite dimensional optimization problem that is a linear combination of the representors. Furthermore, $\| f^* \|_{Sob}^2 \leq \| f^* \|_{Sob}^2 + \| h \|_{Sob}^2 = \| f \|_{Sob}^2 \leq L$. We note also that $\| f^* \|_{Sob}^2 = c'Rc$. Finally, we observe that the following two problems are equivalent:

$$\min_{f^*} \frac{1}{T} \sum \big[y_t - f^*(x_t)\big]^2 \quad s.t. \quad \| f^* \|_{Sob}^2 \leq L, \quad f^* \in M$$

$$and \quad \min_c \frac{1}{T} [y - Rc]'[y - Rc] \quad s.t. \quad c'Rc \leq L \qquad \blacksquare$$

COMMENT: Lemmas 1 and 2 are used to prove rate of convergence and triangular array convergence results. The latter are used to justify bootstrap procedures. Both Lemmas are straightforward adaptations of Van de Geer (1990, Lemma 3.5). (Van de Geer uses a sub-gaussianity assumption on the residuals. We assume bounded support. She also uses a somewhat different definition of the set over which estimation takes place.) The rate of convergence results also yield asymptotic normality of the estimated average sum of squared residuals. For an alternate approach to these asymptotic normality results, see Yatchew (1992) who uses arguments developed by Pollard (1984, p.140). ■

LEMMA 1: Invoke Assumptions for the Single Equation Model, (section 3.1). Assume $f_o \in \mathfrak{F}$ and let $\hat{f}$ satisfy $\min_{f \in \mathfrak{F}} \frac{1}{T} \sum (y_t - f(x_t))^2$. Then $\exists$ positive constants A, C_o, K_o such that $\forall K > K_o$

$$\text{(L1.1)} \qquad Prob\left[\frac{1}{T} \Sigma \left(\hat{f}(x_t) - f_o(x_t)\right)^2 \geq \left(\frac{K^2 A}{T}\right)^{\frac{2m}{2m+q}} \right] \leq \exp\left[-C_o K^2\right] \quad \blacksquare$$

PROOF: Let $N(\delta\,; \mathfrak{F})$ be the minimum number of balls of radius δ in supnorm required to cover the set of functions $\mathfrak{F}$. Using Kolmogorov and Tichomirov (1959), it can be shown that $\exists$ A>0 such that for δ>0, we have $\log N(\delta\,; \mathfrak{F}) \leq A\, \delta^{-q/m}$. Use this result and Van de Geer (1990) Lemma 3.5 to establish that there exist positive constants C_o, K_o such that $\forall$ K>K_o,

$$\text{(L1.2)} \qquad Prob\left[\sup_{\|f\|^2_{Sob} \leq L} T^{1/2} \frac{\left| -\frac{2}{T} \Sigma\ \upsilon_t\left(f_o(x_t) - f(x_t)\right) \right|}{\left(\frac{1}{T} \Sigma\ (f_o(x_t) - f(x_t))^2\right)^{\frac{1}{2} - \frac{q}{4m}}} \geq K A^{1/2} \right] \leq \exp\left[-C_o K^2\right]$$

Since $f_o \in \mathfrak{F}$ and $\hat{f}$ minimizes the sum of squared residuals over $f \in \mathfrak{F}$,

$$\text{(L1.3)} \qquad \frac{1}{T} \Sigma\ (f_o(x_t) - \hat{f}(x_t))^2 \leq -\frac{2}{T} \Sigma\ \upsilon_t\left(f_o(x_t) - \hat{f}(x_t)\right)$$

Now combine (L1.2) with (L1.3) to obtain the result that $\forall$ K>K_o

$$\text{(L1.4)} \qquad Prob\left[T^{1/2} \left(\frac{1}{T} \Sigma\ \left(\hat{f}(x_t) - f_o(x_t)\right)^2\right)^{1-\left(\frac{1}{2} - \frac{q}{4m}\right)} \geq K A^{1/2} \right] \leq \exp\left[-C_o K^2\right]$$

which after straightforward manipulation yields (L1.1). $\blacksquare$

LEMMA 2: Suppose that in addition to the assumptions of Lemma 1, we have a scalar random variable ω_t where $(\upsilon_t, \omega_t)|x_t$ are i.i.d. with probability law $P_{\upsilon\omega} \in \mathcal{P}_{\upsilon\omega}$, a collection of probability laws with mean 0 and support a bounded subset of $\mathbb{R}^2$. Then $\exists$ positive constants A, C_o, K_o such that $\forall K > K_o$

$$\text{(L2.1)} \qquad Prob\left[\left| \frac{1}{T} \Sigma\ \omega_t\left(\hat{f}(x_t) - f_o(x_t)\right) \right| \geq \left(\frac{K^2 A}{T}\right)^{\frac{2m}{2m+q}} \right] \leq \exp\left[-C_o K^2\right] \quad \blacksquare$$

PROOF: Following the proof of Lemma 1, there are positive constants A, $C_o^{/}, K_o^{/}$ such that $\forall K > K_o^{/}$,

$$\text{(L2.2)} \quad Prob\left[\sup_{\|f\|^2_{Sob} \leq L} T^{1/2} \frac{\left| \frac{1}{T} \Sigma\, \omega_t \left(f_o(x_t) - f(x_t)\right) \right|}{\left(\frac{1}{T} \Sigma \left(f_o(x_t) - f(x_t)\right)^2 \right)^{\frac{1}{2} - \frac{q}{4m}}} \geq K A^{1/2} \right] \leq \exp\left[-C_o^{/} K^2\right]$$

which implies

$$\text{(L2.3)} \quad Prob\left[\left| \frac{1}{T} \Sigma\, \omega_t \left(f_o(x_t) - \hat{f}(x_t)\right) \right| \geq \left(\frac{K^2 A}{T} \right)^{1/2} \left(\frac{1}{T} \Sigma \left(f_o(x_t) - \hat{f}(x_t)\right)^2 \right)^{\frac{1}{2} - \frac{q}{4m}} \right] \leq \exp\left[-C_o^{/} K^2\right]$$

Equation (L1.1) implies $\exists\, C_o^{//}, K_o^{//}$ s.t. $\forall K > K_o^{//}$

$$\text{(L2.4)} \qquad Prob\left[\left(\frac{1}{T} \Sigma \left(\hat{f}(x_t) - f_o(x_t)\right)^2 \right)^{\frac{1}{2} - \frac{q}{4m}} \geq \left(\frac{K^2 A}{T} \right)^{\frac{2m-q}{2(2m+q)}} \right] \leq \exp\left[-C_o^{//} K^2\right]$$

Combining (L2.3) and (L2.4) yields (L2.1). ■

REMARKS ON LEMMAS 1 AND 2:

1. From equation (L1.1) it is immediate that $\frac{1}{T} \Sigma \left(\hat{f}(x_t) - f_o(x_t)\right)^2$ is $O_P\left(T^{\frac{-2m}{2m+q}}\right)$. But Lemma 1 implies a stronger triangular array result. In particular, let $\{\theta_T\} = \{(f_T, P_{\upsilon T})\}$ be a sequence of DGM's such that $\forall\, T, f_T \in \mathfrak{F}, P_{\upsilon T} \in \mathcal{P}$. For each T, $y_t(\theta_T) = f_T(x_t) + \upsilon_t(P_{\upsilon T})$ $t = 1, \ldots, T$. Let $\hat{f}_T$ satisfy $\min_{f \in \mathfrak{F}} \frac{1}{T} \sum (y_t(\theta_T) - f(x_t))^2$. Since the constants in equation (L1.1) are not dependent on the specific DGM, $\frac{1}{T} \Sigma \left(\hat{f}_T(x_t;\theta_T) - f_T(x_t)\right)^2$ is also $O_P\left(T^{-2m/(2m+q)}\right)$. Similarly, Lemma 2 implies $\frac{1}{T} \Sigma\, \omega_t \left(\hat{f}(x_t;\theta_T) - f_T(x_t)\right)$ is $O_P\left(T^{-2m/(2m+q)}\right)$. Note that we may set $\omega = \upsilon$ in which case $\frac{1}{T} \Sigma\, \upsilon_t \left(\hat{f}(x_t;\theta_T) - f_T(x_t)\right)$ is $O_P\left(T^{-2m/(2m+q)}\right)$.

2. Lemmas 1 and 2 may be generalized to the constrained estimation setting. Invoke the Assumptions for the Constrained Single Equation Model, (section 5.1). Suppose $f_o \in \bar{\mathfrak{F}}$ and let $\hat{f}$ satisfy $\min_{f \in \bar{\mathfrak{F}}_T} \frac{1}{T} \sum (y_t - f(x_t))^2$. Let $r = \frac{2}{2+\zeta}$. Then $\exists$ positive constants A, C_o, K_o such

that $\forall$ K>K$_o$ we have

$$Prob\left[\frac{1}{T}\sum\left(\hat{f}(x_t)-f_o(x_t)\right)^2 \geq \left(\frac{K^2A}{T}\right)^r\right] \leq \exp\left[-C_oK^2\right]$$

$$and\quad Prob\left[\left|\frac{1}{T}\sum\omega_t\left(\hat{f}(x_t)-f_o(x_t)\right)\right| \geq \left(\frac{K^2A}{T}\right)^r\right] \leq \exp\left[-C_oK^2\right]$$

Thus $\frac{1}{T}\sum\left(\hat{f}(x_t)-f_o(x_t)\right)^2$ and $\frac{1}{T}\sum\omega_t\left(\hat{f}(x_t)-f_o(x_t)\right)$ are $O_P(T^{-r})$. Suppose $\{\theta_T\}=\{(f_T,P_{\upsilon T})\}$ is a sequence of DGM's such that $\forall\ T, f_T\in\bar{\mathfrak{F}}_T, P_{\upsilon T}\in\mathcal{P}$. Then $\frac{1}{T}\sum\left(\hat{f}(x_t;\theta_T)-f_T(x_t)\right)^2$ and $\frac{1}{T}\sum\omega_t\left(\hat{f}(x_t;\theta_T)-f_T(x_t)\right)$ are $O_P(T^{-r})$. Note that we may again set $\omega=\upsilon$ in which case $\frac{1}{T}\sum\upsilon_t\left(\hat{f}(x_t;\theta_T)-f_T(x_t)\right)$ is $O_P(T^{-r})$.

3. Generalization to a multi-equation setting is also straightforward. Invoke the Assumptions for the Multi-Equation Model (section 3.2) where for simplicity we will focus on the 2-equation model $f_o(x)=(f_{ao}(x),f_{bo}(x))'$, $y=(y_a,y_b)'$. Let $\hat{f}$ satisfy $\min_{f\in\mathfrak{F}}\frac{1}{T}\sum_t[y_t-f(x_t)]'\Lambda[y_t-f(x_t)]$, suppose $f_o\in\mathfrak{F}$ and let $r=\frac{2m}{2m+q}$. Then $\exists$ positive constants A, C_o, K_o such that $\forall$ K>K$_o$

$$Prob\left[\frac{1}{T}\sum_t\left[\hat{f}(x_t)-f_o(x_t)\right]'\Lambda\left[\hat{f}(x_t)-f_o(x_t)\right] \geq \left(\frac{K^2A}{T}\right)^r\right] \leq \exp\left[-C_oK^2\right]$$

hence $\frac{1}{T}\sum_t\left(\hat{f}_a(x_t)-f_{ao}(x_t)\right)^2$, $\frac{1}{T}\sum_t\left(\hat{f}_b(x_t)-f_{bo}(x_t)\right)^2$ and $\frac{1}{T}\sum_t\left(\hat{f}_a(x_t)-f_{ao}(x_t)\right)\left(\hat{f}_b(x_t)-f_{bo}(x_t)\right)$ are $O_P(T^{-r})$.

PROOF OF THEOREM 3.1.2: Part b), the rate of convergence result, follows directly from remark 1 following Lemma 2 above. To prove part c), expand s^2 to obtain:

$$T^{1/2}\left(s^2-\sigma^2_{\upsilon o}\right)=T^{1/2}\left(\frac{1}{T}\sum\upsilon_t^2-\sigma^2_{\upsilon o}\right)+T^{1/2}\frac{1}{T}\sum\left(f_o(x_t)-\hat{f}(x_t)\right)^2+T^{1/2}\frac{2}{T}\sum\upsilon_t\left(f_o(x_t)-\hat{f}(x_t)\right)$$

The first term is asymptotically $N(0,Var(\upsilon^2))$. The second and third terms go to zero. (See remark 1 following Lemma 2 above and recall that $m>\frac{q}{2}$.) Finally, part a), the convergence of s^2 is implicit in part c). (Alternatively and under weaker assumptions consistency can be proved using Ranga Rao (1962, Th. 6.2, p.672). See e.g., Epstein and Yatchew (1985).) ■

PROOF OF THEOREM 3.2.1: Let $M = span\{r_{x_t} : t=1,\ldots,T\}$, $M^{\perp} = \{h \in \mathcal{H}^m : <r_{x_t}, h>_{Sob} = 0, \forall t\}$. Representors exist by Theorem 2.2 and we can write $\mathcal{H}^m = M \oplus M^{\perp}$ since $\mathcal{H}^m$ is a Hilbert space. The functions $h \in M^{\perp}$ take on the value zero at $x_1,\ldots,x_T$. Consider a vector function $f = (f_1,\ldots,f_p)'$ where $f_i \in \mathcal{H}^m$. Write $f_i = \sum_{t=1}^{T} C_{ti}\, r_{x_t} + h_i$, $h_i \in M^{\perp}$. Then

$$f_i(x_t) = <r_{x_t}, \sum_{j=1}^{T} C_{ji} r_{x_j} + h_i>_{Sob} = \sum_{j=1}^{T} <r_{x_t}, C_{ji} r_{x_j}>_{Sob} = \sum_{j=1}^{T} C_{ji} <r_{x_t}, r_{x_j}>_{Sob} = \sum_{j=1}^{T} R_{tj} C_{ji}$$

Note further that

$$\| f_i \|^2_{Sob} = <f_i, f_i>_{Sob} = <\sum_{j=1}^{T} C_{ji} r_{x_j}, \sum_{j=1}^{T} C_{ji} r_{x_j}>_{Sob} + <h_i, h_i>_{Sob} = (C_{1i},\ldots,C_{Ti}) R \begin{pmatrix} C_{1i} \\ \vdots \\ C_{Ti} \end{pmatrix} + <h_i, h_i>_{Sob}$$

Suppose that f minimizes $\frac{1}{T} \sum_t \, [y_t - f(x_t)]' \, \Lambda \, [y_t - f(x_t)]$ s.t. $\|f_i\|^2_{Sob} \le L_i$, $i=1,\ldots,p$, then so does $f^* = f - h$ where

$$f = C' \begin{bmatrix} r_{x_1} \\ \vdots \\ r_{x_T} \end{bmatrix} + \begin{bmatrix} h_1 \\ \vdots \\ h_p \end{bmatrix} \qquad f^* = C' \begin{bmatrix} r_{x_1} \\ \vdots \\ r_{x_T} \end{bmatrix}$$

since the h_i are zero at $x_1,\ldots,x_T$. Hence, there exists a function f^* minimizing the infinite dimensional optimization problem whose components are linear combinations of the representors. Furthermore, $\| f_i^* \|^2_{Sob} \le \| f_i^* \|^2_{Sob} + \| h_i \|^2_{Sob} \le \| f_i \|^2_{Sob} \le L_i$. We note also that $\|f_i^*\|^2_{Sob} = (C_{1i},\ldots,C_{Ti}) R \, (C_{1i},\ldots,C_{Ti})'$. Finally, we observe that

$$\min_{f^*} \frac{1}{T} \sum_t \, [y_t - f^*(x_t)]' \, \Lambda \, [y_t - f^*(x_t)] \quad s.t. \; \|f_i^*\|^2_{Sob} \le L_i, \; i=1,\ldots,p$$

is equivalent to the finite dimensional optimization problem (3.2.2). ▪

PROOF OF THEOREM 3.2.2: The proof is similar to that of Th. 3.1.2. See remark 3 following Proof of Lemma 2 for rate of convergence results which may be then used to demonstrate asymptotic normality and convergence of $\hat{\Sigma}_{\upsilon}$. ▪

PROOF OF THEOREM 3.3.1: Let $\{\theta_T\} = \{(f_T, P_{\upsilon T})\} \subset \Theta$ be a sequence of DGM's. First show:

$$T^{\frac{1}{2}} \frac{1}{T} \sum_{t=1}^{T} \left(f_{Ti}(x_t) - \hat{f}_{Ti}(x_t;\theta_T)\right) \left(f_{Tj}(x_t) - \hat{f}_{Tj}(x_t;\theta_T)\right) \overset{P}{\to} 0 \qquad i,j = 1,\ldots,p$$

If $i=j$, equation (L1.1) of Lemma 1 is sufficient since the constants of (L1.1) do not depend on the specific DGM and $\frac{1}{2} - \frac{2m}{2m+q} < 0$. (See remark 1 following Proof of Lemma 2.) If $i \neq j$, use Cauchy-Schwartz then apply equation (L1.1). Next show that

$$T^{\frac{1}{2}} \frac{1}{T} \sum_{t=1}^{T} \left(f_{Ti}(x_t) - \hat{f}_{Ti}(x_t;\theta_T)\right) \upsilon_{jt}(P_{\upsilon T}) \overset{P}{\to} 0 \qquad i,j = 1,\ldots,p$$

for which equation (L2.1) is sufficient. (Lemma 2 is valid if $\omega = \upsilon$, see remark 1 following Proof of Lemma 2). Finally, note that the components of the vector in equation (3.3.2), after expansion, are composed of sums of terms like the ones above. ∎

PROOF OF THEOREM 4.1: Define $\hat{\Sigma}^{-1}_{zz|x}$, $\hat{\sigma}_{zy|x}$ and $\hat{\beta} = \hat{\Sigma}^{-1}_{zz|x}\hat{\sigma}_{zy|x}$ as in the text by using the estimated residuals from the nonparametric regression (4.1). Define $\underline{\beta} = \underline{\Sigma}^{-1}_{zz|x}\underline{\sigma}_{zy|x}$ where $\underline{\Sigma}^{-1}_{zz|x}$, $\underline{\sigma}_{zy|x}$ are the sample moments of the true (unobserved) residuals u_t, v_t. Now consider

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\beta} - \underline{\beta}) &= T^{\frac{1}{2}}\left(\hat{\Sigma}^{-1}_{zz|x}\hat{\sigma}_{zy|x} - \underline{\Sigma}^{-1}_{zz|x}\underline{\sigma}_{zy|x}\right) \\ &= \underline{\Sigma}^{-1}_{zz|x}\, T^{\frac{1}{2}}\left(\hat{\sigma}_{zy|x} - \underline{\sigma}_{zy|x}\right) + T^{\frac{1}{2}}\left(\hat{\Sigma}^{-1}_{zz|x} - \underline{\Sigma}^{-1}_{zz|x}\right)\hat{\sigma}_{zy|x} \end{aligned}$$

Th. 3.3.1 implies that $T^{\frac{1}{2}}\left(\hat{\sigma}_{zy|x} - \underline{\sigma}_{zy|x}\right) \to 0$ and $T^{\frac{1}{2}}\left(\hat{\Sigma}^{-1}_{zz|x} - \underline{\Sigma}^{-1}_{zz|x}\right) \to 0$. Consistency of $\underline{\Sigma}^{-1}_{zz|x}$, $\hat{\sigma}_{zy|x}$ then implies that $T^{\frac{1}{2}}(\hat{\beta} - \underline{\beta}) \to 0$. Since $T^{\frac{1}{2}}(\underline{\beta} - \beta_o) \overset{D}{\to} N\left(0, \sigma^2_\varepsilon \Sigma^{-1}_{zz|x}\right)$, so does $T^{\frac{1}{2}}(\hat{\beta} - \beta_o)$. Finally, note that η is a continuous function of $T^{\frac{1}{2}}(\hat{\beta} - \beta_o)$, hence its limiting distribution is χ^2_r ∎

PROOF OF THEOREM 4.2: Consistent estimation of g_o, h_o in (4.1) is established by Th. 3.1.2, and consistent estimation of β_o is implicit in Th. 4.1. Glivenko-Cantelli then ensures consistent estimation of P_{uvo}. We need to show that along any path θ_T of DGM's converging to θ_o, (the true DGM), the exact sampling distribution of the statistic converges to the specified normal. For given θ_T define $\hat{\Sigma}^{-1}_{zz|x}(\theta_T)$, $\hat{\sigma}_{zy|x}(\theta_T)$ and $\hat{\beta}(\theta_T) = \hat{\Sigma}^{-1}_{zz|x}(\theta_T)\hat{\sigma}_{zy|x}(\theta_T)$ using the estimated residuals from nonparametric regression estimation of (4.1). Define $\underline{\beta}(\theta_T) = \underline{\Sigma}^{-1}_{zz|x}(\theta_T)\underline{\sigma}_{zy|x}(\theta_T)$ where $\underline{\Sigma}^{-1}_{zz|x}(\theta_T)$, $\underline{\sigma}_{zy|x}(\theta_T)$ are the sample moments of the true (unobserved) residuals $u_t(\theta_T), v_t(\theta_T)$, (the DGM for which is simply P_{uvT}). Now consider

$$
\begin{aligned}
T^{1/2}\left(\hat{\beta}(\theta_T) - \underline{\beta}(\theta_T)\right) &= T^{1/2}\left(\hat{\Sigma}^{-1}_{zz|x}(\theta_T)\hat{\sigma}_{zy|x}(\theta_T) - \underline{\Sigma}^{-1}_{zz|x}(\theta_T)\underline{\sigma}_{zy|x}(\theta_T)\right) \\
&= \underline{\Sigma}^{-1}_{zz|x}(\theta_T)\, T^{1/2}\left(\hat{\sigma}_{zy|x}(\theta_T) - \underline{\sigma}_{zy|x}(\theta_T)\right) + T^{1/2}\left(\hat{\Sigma}^{-1}_{zz|x}(\theta_T) - \underline{\Sigma}^{-1}_{zz|x}(\theta_T)\right)\hat{\sigma}_{zy|x}(\theta_T)
\end{aligned}
$$

Th. 3.3.1 implies that $T^{1/2}\left(\hat{\sigma}_{zy|x}(\theta_T) - \underline{\sigma}_{zy|x}(\theta_T)\right) \to 0$ and $T^{1/2}\left(\left(\hat{\Sigma}_{zz|x}(\theta_T)\right)^{-1} - \left(\underline{\Sigma}_{zz|x}(\theta_T)\right)^{-1}\right) \to 0$. Consistency of $\underline{\Sigma}^{-1}_{zz|x}(\theta_T)$ follows from the triangular array law of large numbers and that of $\hat{\sigma}_{zy|x}(\theta_T)$ from equation (3.3.5). Thus $T^{1/2}\left(\hat{\beta}(\theta_T) - \underline{\beta}(\theta_T)\right) \to 0$. Since $T^{1/2}\left(\underline{\beta}(\theta_T) - \beta_T\right) \overset{D}{\to} N\left(0, \sigma^2_\varepsilon(\theta_o)\,\Sigma^{-1}_{zz|x}(\theta_o)\right)$, so does $T^{1/2}\left(\hat{\beta}(\theta_T) - \beta_T\right)$.

Hence, $T^{1/2}(\beta^B - \beta) \overset{D}{\to} N\left(0, \sigma^2_\varepsilon(\theta_o)\,\Sigma^{-1}_{zz|x}(\theta_o)\right)$ Finally, η^B is a continuous function of $T^{1/2}(\hat{\beta}^B - \hat{\beta})$, hence its limiting distribution is χ^2_r. ■

PROOF OF PROPOSITION 5.1.1: If $f_o \in \bar{\Im}$, the proof is analogous to that of Th. 3.1.2. See remark 2 following Proof of Lemma 2. If $f_o \notin \bar{\Im}$ proceed as follows. Since each $\bar{\Im}_T$ is a subset of a fixed Sobolev ball and closed by construction, the compact imbedding theorem, (Theorem 2.3) implies that it is compact with respect to supnorm. Let $\bar{f}_{oT}$ satisfy $\min_{f \in \bar{\Im}_T} \int (f_o - f)^2 dP_x$. To show that $sup_x |\bar{f}_{oT} - \bar{f}_o| \to 0$ a.s., restrict attention to those sequences $\{x_t\}$ which are dense in the domain. The set of such sequences has probability one. We have $\bigcap_1^\infty \bar{\Im}_T = \bar{\Im}$ by assumption. Consider the sequence $\bar{f}_{oT}$. Each element of this sequence is in $\bar{\Im}_1$ since $\bar{\Im}_1 \supset \ldots \supset \bar{\Im}_T \ldots$. Furthermore, $\bar{\Im}_1$ is compact with respect to supnorm. Hence, there exists a convergent subsequence which we label f_T^* which converges to say $\bar{f}^*$ in $\bar{\Im}_1$. Next consider the following collection of sequences: $\{f_1^*, f_2^*, \ldots\} \subset \bar{\Im}_1$, $\{f_2^*, f_3^*, \ldots\} \subset \bar{\Im}_2, \ldots,$ $\{f_T^*, f_{T+1}^*, \ldots\} \subset \bar{\Im}_T, \ldots.$ Each sequence converges to $\bar{f}^*$ and since $\bar{\Im}_T$ is compact, $\bar{f}^* \in \bar{\Im}_T$ $\forall$ T, in which case $\bar{f}^* \in \bar{\Im} = \bigcap_1^\infty \bar{\Im}_T$. Consider the sequence $I_T = \int (f_o - f_T^*)^2 \to \int (f_o - \bar{f}^*)^2$. I_T is a nondecreasing sequence. Furthermore, $\int (f_o - f_T^*)^2 \leq \int (f_o - \bar{f}_o)^2 \leq \int (f_o - \bar{f}^*)^2$ since $\bar{f}^* \in \bar{\Im}$. But $\bar{f}_o$ is the unique function minimizing the integral over $\bar{\mathfrak{F}}$. Hence, $\bar{f}^* = \bar{f}_o$. Return to the original sequence $\bar{f}_{oT}$. Since the selection of the convergent sequence above was arbitrary, we have every convergent subsequence of $\bar{f}_{oT}$ converging to $\bar{f}_o$. Suppose $\bar{f}_{oT}$ does not converge to $\bar{f}_o$. By compactness there exists a subsequence converging to some function not equal to $\bar{f}_o$ which is a contradiction. Hence, $\bar{f}_{oT}$ converges to $\bar{f}_o$. Now, consider the following inequalities:

$$\frac{1}{T}\sum\left[y_t-\hat{f}(x_t)\right]^2-\sigma_\upsilon^2-\int\left[f_0-\hat{f}\right]^2 \le \frac{1}{T}\sum\left[y_t-\hat{f}(x_t)\right]^2-\sigma_\upsilon^2-\int\left[f_o-\bar{f}_{oT}\right]^2 \le \frac{1}{T}\sum\left[y_t-\bar{f}_o(x_t)\right]^2-\sigma_\upsilon^2-\int\left[f_o-\bar{f}_{oT}\right]^2$$

The left hand side converges to 0 using Ranga Rao (1962, Th. 6.2, p. 672). The right hand side converges to 0 using the law of large numbers and the fact that $\bar{f}_{oT}$ converges to $\bar{f}_o$. Hence $s^2 \overset{a.s.}{\to} \sigma_{\upsilon o}^2 + \int\left[f_o-\bar{f}_o\right]^2$. ∎

PROOF OF THEOREM 5.2.1:

K will refer to both the kernel function and to the corresponding TxT matrix whose entries are $K_{st} = K((x_s-x_t)/\lambda)$, $s,t=1,\dots,T$. The matrix $\frac{K}{\lambda^q T}$ when multiplied by a column vector of ones, yields the kernel estimate of the density of x at $x_1,\dots,x_T$. The matrix is symmetric and may be decomposed as $\frac{K}{\lambda^q T} = A\,\Upsilon A'$ where A is orthogonal and Υ is the diagonal matrix of eigenvalues. For given T the matrix K and hence the eigenvalues are determined by the sequence $x_1,\dots,x_T$. Under our assumptions about the DGM for the x's, $max\,|\Upsilon_{tt}|$ is bounded in probability. Similarly, if K^2 is the matrix whose elements are the squares of the elements of K, then $\frac{K^2}{\lambda^q T}$ is symmetric with largest eigenvalue bounded in probability.

a) Variance of U_1: Recall $U_1 = \frac{1}{\lambda^q T^2}\sum_t\sum_{s\neq t}\upsilon_t\upsilon_s K\left(\frac{x_s-x_t}{\lambda}\right)$, and note that $E(U_1)=0$. To obtain $Var(U_1)$, proceed as follows where if x is a vector then the limits of integration 0 and 1 are the corresponding vectors of zeros and ones:

$$E\left(\frac{\upsilon_t^2\,\upsilon_s^2}{\lambda^q}\,K^2\left(\frac{x_s-x_t}{\lambda}\right)\right)$$
$$= \sigma_\upsilon^4\int_0^1\int_0^1 \frac{1}{\lambda^q}K^2\left(\frac{x_s-x_t}{\lambda}\right)p_x(x_t)\,p_x(x_s)\,dx_s dx_t$$
$$= \sigma_\upsilon^4\int_0^1 p_x(x_t)\left(\int_{-x_t/\lambda}^{(1-x_t)/\lambda} K^2(u)\,p_x(\lambda u+x_t)\,du\right)dx_t \qquad u=(x_s-x_t)/\lambda \quad du=\lambda^{-q}dx_s$$
$$\rightarrow \sigma_\upsilon^4\int_0^1 p_x(x_t)\left(\int_{-1/2}^{1/2} K^2(u)\,p_x(\lambda u+x_t)\,du\right)dx_t \qquad \lambda\rightarrow 0,\ K(\cdot)=0 \ \textit{outside}\ [-1/2,1/2]$$
$$\rightarrow \sigma_\upsilon^4\int_0^1 p_x^2(x_t)\left(\int_{-1/2}^{1/2} K^2(u)\,du\right)dx_t \qquad \textit{by dominated convergence}$$

That is,

$$(*) \qquad E\left(\frac{\upsilon_t^2\,\upsilon_s^2}{\lambda^q}\,K^2\left(\frac{x_s-x_t}{\lambda}\right)\right) \rightarrow \sigma_\upsilon^4\int p_x^2\int K^2$$

Hence $Var\left(\lambda^{q/2}\,TU_1\right) \rightarrow 2\sigma_\upsilon^4\int p_x^2\int K^2$.

b) Distribution of U_1: this portion of the proof may be obtained using Hall (1984, Theorem 1), see e.g., Fan and Li (1996). We use a triangular array central limit result due to McLeish (1974) which simplifies demonstration of bootstrap related results below. Rewrite $U_1 = \frac{1}{T}\sum_t \upsilon_t \frac{2}{\lambda^q T}\left[\sum_{s<t}\upsilon_s\,K\left(\frac{x_s-x_t}{\lambda}\right)\right]$ and define

$$z_{tT} = \frac{\lambda^{q/2}}{\sigma_\upsilon^2\sqrt{2\int p_x^2\int K^2}}\,\upsilon_t\left[\frac{2}{\lambda^q T}\sum_{s<t}\upsilon_s\,K\left(\frac{x_s-x_t}{\lambda}\right)\right]$$

where λ and $K((x_s-x_t)/\lambda)$ are (implicitly) functions of T. Note that $\sum_t^T z_{tT} = \lambda^{q/2}\,TU_1/\left(\sigma_\upsilon^2\sqrt{2\int p_x^2\int K^2}\right)$. Now apply McLeish (1974, Corollary (2.6), p. 622) to conclude $\sum_t^T z_{tT} \overset{D}{\rightarrow} N(0,1)$, i.e., $\lambda^{q/2}TU_1 \overset{D}{\rightarrow} N\left(0\ ,\ 2\sigma_\upsilon^4\int p_x^2\int K^2\right)$.

c) Rate of Convergence of U_2 and U_3: with moderate abuse of notation, let $\left[f_o(x)-\hat{f}(x)\right]$ be the $Tx1$ vector with elements $f_o(x_t)-\hat{f}(x_t)$. Since $\exists\, k$ s.t. $Prob\left[max\,|\Upsilon_{tt}| > k\,\right]\rightarrow 0$ we have with probability going to one:

$$U_2 = \frac{1}{T}\left[f_o(x) - \hat{f}(x)\right]' A \Upsilon A'\left[f_o(x) - \hat{f}(x)\right] \leq \frac{1}{T}\left[f_o(x) - \hat{f}(x)\right]' A k I A'\left[f_o(x) - \hat{f}(x)\right]$$
$$= k\frac{1}{T}\sum_t \left(f_o(x_t) - \hat{f}(x_t)\right)^2$$

in which case, $U_2 = O_P\left(\frac{1}{T}\sum_t \left(f_o(x_t) - \hat{f}(x_t)\right)^2\right)$. Similarly, $U_3 = O_P\left(\frac{1}{T}\sum_t \upsilon_t\left(f_o(x_t) - \hat{f}(x_t)\right)\right)$. Thus, U_2 and U_3 are $O_P(T^{-r})$, (see remark 2 following Lemma 2).

d) Since by assumption, $\lambda^{q/2}T^{1-r} \to 0$, we have $\lambda^{q/2}TU_2 \xrightarrow{P} 0$ and $\lambda^{q/2}TU_3 \xrightarrow{P} 0$, in which case $\lambda^{q/2}TU \xrightarrow{D} N\left(0\ ,\ 2\sigma_\upsilon^4\int p_x^2 \int K^2\right)$.

e) To show that $\lambda^q T^2 \hat{\sigma}_U^2 \xrightarrow{P} 2\sigma_\upsilon^4 \int p^2(x)\int K^2(u)$, expand to obtain:

$$\begin{aligned}
\lambda^q T^2 \hat{\sigma}_U^2 &= \frac{2}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t^2 \upsilon_s^2 K^2\left(\frac{x_s - x_t}{\lambda}\right) \\
&+ \frac{2}{\lambda^q T^2}\sum_t \sum_{s\neq t} \left(f_o(x_t) - \hat{f}(x_t)\right)^2 \left(f_o(x_s) - \hat{f}(x_s)\right)^2 K^2\left(\frac{x_s - x_t}{\lambda}\right) \\
&+ \frac{4}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t^2 \left(f_o(x_s) - \hat{f}(x_s)\right)^2 K^2\left(\frac{x_s - x_t}{\lambda}\right)
\end{aligned}$$

Using equation (*) and a law of large numbers the first term converges to $2\sigma_\upsilon^4\int p^2(x)\int K^2(u)$. Using arguments similar to part c) above, the second and third terms converge to zero, in which case $\lambda^q T^2 \hat{\sigma}_U^2 \xrightarrow{P} 2\sigma_\upsilon^4 \int p^2(x) \int K^2(u)$. Combining these results, we have $U / \hat{\sigma}_U \xrightarrow{D} N(0\,,1)$. ■

PROOF OF THEOREM 5.2.2: The proof is similar to that of Th. 5.2.1. We need to show that along any path of DGM's $\{\theta_T\} = \{(f_T, P_{\upsilon T})\} \subset \Theta$ converging to $\{\theta_o\} = \{(f_o, P_{\upsilon o})\}$ the distribution of the test statistic converges to $N(0,1)$. Define $K_{st} = K((x_s - x_t)/\lambda)$ and:

$$\begin{aligned}
U(\theta_T) &= U_1(P_{\upsilon T}) + U_2(\theta_T) + U_3(\theta_T) \\
&= \frac{1}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t(P_{\upsilon T})\, \upsilon_s(P_{\upsilon T})\, K_{st} \\
&+ \frac{1}{\lambda^q T^2}\sum_t \sum_{s\neq t} \left(f_T(x_t) - \hat{f}(x_t;\theta_T)\right)\left(f_T(x_s) - \hat{f}(x_s,\theta_T)\right) K_{st} \\
&+ \frac{2}{\lambda^q T^2}\sum_t \sum_{s\neq t} \upsilon_t(P_{\upsilon T})\left(f_T(x_s) - \hat{f}(x_s,\theta_T)\right) K_{st}
\end{aligned}$$

Note that $U_1(\cdot)$ depends only on $P_{\upsilon T}$ while $U_2(\cdot)$ and $U_3(\cdot)$ depend on the whole DGM θ_T.

Take any sequence $\{\theta_T\} = \{(f_T, P_{\upsilon T})\} \subset \Theta$. To show that $\lambda^{q/2} T U_1(P_{\upsilon T})$ converges to a $N\left(0,\, 2\sigma_\upsilon^4 \int p_x^2 \int K^2\right)$ refer to the Proof of Th. 5.2.1 part b) and note that the McLeish result is a triangular array central limit theorem so that

$$\sum_t^T z_{tT}(\theta_T) \equiv \frac{\lambda^{q/2}\, T U_1(\theta_T)}{\sigma_\upsilon^2 \sqrt{2\int p_x^2 \int K^2}} \overset{D}{\to} N(0,1)$$

Next show that $\lambda^{q/2} T U_2(\theta_T)$ converges to zero. Refer to the Proof of Th. 5.2.1 part c) and noting that since the x's are not resampled, we have for fixed k with probability going to one:

$$U_2(\theta_T) \leq k\frac{1}{T}\sum_t \left(f_T(x_t) - \hat{f}(x_t;\theta_T)\right)^2$$

hence $U(\theta_T) = O_P\left(\frac{1}{T}\sum_t \left(f_T(x_t) - \hat{f}(x_t;\theta_T)\right)^2\right)$. From remark 2 following Proof of Lemma 2 we have $\frac{1}{T}\sum_t\left(f_T(x_t) - \hat{f}(x_t;\theta_T)\right)^2 = O_P(T^{-r})$. Since $\lambda^{q/2} T^{1-r} \to 0$ we have $\lambda^{q/2} T U_2(\theta_T) \overset{P}{\to} 0$. By a similar argument, $\lambda^{q/2} T U_3(\theta_T) \overset{P}{\to} 0$. Hence $\lambda^{q/2} T U(\theta_T) \overset{D}{\to} N\left(0,\, 2\sigma_\upsilon^4 \int p_x^2 \int K^2\right)$. ■

PROOF OF THEOREM 5.3.1: In $\mathcal{H}^m(Q^{q_a+q_b})$ consider the subspace $G_a \oplus G_b$ where $G_a = \{f_a \in \mathcal{H}^m \mid f_a(x_a, x_b) = f_a(x_a)\}$, $G_b = \{f_b \in \mathcal{H}^m \mid f_b(x_a, x_b) = f_b(x_b),\ \int f_b(x_b) = 0\}$. G_a, G_b are orthogonal subspaces, consisting of functions that depend only on x_a, x_b respectively. Within G_a let $r_{x_{a1}}(x_a) \ldots r_{x_{aT}}(x_a)$ be representors of function evaluation at $x_{a1} \ldots, x_{aT}$, let $M_a \subset G_a$ be the span of these representors and write $G_a = M_a \oplus M_a^\perp$. Similarly, within G_b let $r_{x_{b1}}(x_b) \ldots r_{x_{bT}}(x_b)$ be representors of function evaluation at $x_{b1} \ldots, x_{bT}$, let $M_b \subset G_b$ be the span of these representors and write $G_b = M_b \oplus M_b^\perp$. Note that the spaces $M_a, M_a^\perp, M_b, M_b^\perp$ are mutually orthogonal, hence any function in $G_a \oplus G_b$ can be expressed uniquely in the form: $f_a + f_b = \sum_{j=1}^T c_{aj} r_{x_{aj}} + h_a + \sum_{j=1}^T c_{bj} r_{x_{bj}} + h_b$ where $h_a \in M_a^\perp$, $h_b \in M_b^\perp$. The objective function for the infinite dimensional optimization problem may be rewritten as:

$$\frac{1}{T}\sum_t \left[y_t - f_a(x_{at}) - f_b(x_{bt})\right]^2$$
$$= \frac{1}{T}\sum_t \left[y_t - \langle r_{x_{at}}, \sum_j c_{aj} r_{x_{aj}} + h_a \rangle_{Sob} - \langle r_{x_{bt}}, \sum_j c_{bj} r_{x_{bj}} + h_b \rangle_{Sob}\right]^2 = \frac{1}{T}\left[y - R_a c_a - R_b c_b\right]'\left[y - R_a c_a - R_b c_b\right]$$

where R_a, R_b are the representor matrices on Q^{q_a} and Q^{q_b} at $x_{a1} \ldots, x_{aT}$ and $x_{b1} \ldots, x_{bT}$ respectively. Suppose $f_a + f_b = \sum c_{aj} r_{x_{aj}} + h_a + \sum c_{bj} r_{x_{bj}} + h_b$ minimizes $\frac{1}{T}\sum \left[y_t - f_a(x_{at}) - f_b(x_{bt})\right]^2$

$s.t. \parallel f_a + f_b \parallel^2_{Sob} \leq L$. Then so does $f_a^* + f_b^* = f_a - h_a + f_b - h_b = \sum c_{aj} r_{x_{aj}} + \sum c_{bj} r_{x_{bj}}$ since h_a and h_b are zero at $(x_{a1}, x_{b1}), \ldots, (x_{aT}, x_{bT})$. Hence, there exists a function $f_a^* + f_b^*$ minimizing the infinite dimensional optimization problem that is a linear combination of the representors. Furthermore, $\parallel f_a^* + f_b^* \parallel^2_{Sob} \leq \parallel f_a^* + f_b^* \parallel^2_{Sob} + \parallel h_a + h_b \parallel^2_{Sob} = \parallel f_a + f_b \parallel^2_{Sob} \leq L$. Using Th.2.4 $\parallel f_a^* + f_b^* \parallel^2_{Sob} = \parallel f_a^* \parallel^2_{Sob} + \parallel f_b^* \parallel^2_{Sob} = c_a' R_a c_a + c_b' R_b c_b$. Finally, note that for $f_b^* \in M_b$, $0 = \int f_b^*(x_b) = \langle f_b^*, 1 \rangle_{Sob} = \Sigma c_{bt}$, and observe that the following two problems

$$\min_{f_a^*, f_b^*} \frac{1}{T} \sum \left[y_t - f_a^*(x_{at}) - f_b^*(x_{bt}) \right]^2 \quad s.t. \quad \parallel f_a^* + f_b^* \parallel^2_{Sob} \leq L, \quad f_a^* \in M_a, \quad f_b^* \in M_b$$

$$\min_{c_a, c_b} \frac{1}{T} \left[y - R_a c_a - R_b c_b \right]' \left[y - R_a c_a - R_b c_b \right] \quad s.t. \quad c_a' R_b c_a + c_b' R_b c_b \leq L, \quad \Sigma c_{bt} = 0$$

are equivalent. ∎

PROOF OF PROPOSITION 5.3.2: First note that Th. 3.1.2(b) combined with bounded first derivatives implies $sup_{x_a, x_b} |\hat{f}_a(x_a) + \hat{f}_b(x_b) - f_{ao}(x_a) - f_{bo}(x_b)| \overset{a.s.}{\to} 0$. To show that f_{ao} and f_{bo} are separately identified fix x_a at x_a^* and note that:

$$\int \hat{f}_a(x_a^*) - f_{ao}(x_a^*) + \hat{f}_b(x_b) - f_{bo}(x_b) \, dx_b = \hat{f}_a(x_a^*) - f_{ao}(x_a^*) + \int \hat{f}_b(x_b) - f_{bo}(x_b) \, dx_b \overset{a.s.}{\to} 0$$

But $\int \hat{f}_b(x_b) dx_b = 0$ and $\int f_{bo}(x_b) dx_b = 0$. Hence, $|\hat{f}_a(x_a^*) - f_{ao}(x_a^*)| \overset{a.s.}{\to} 0$. In which case $sup_{x_b} |\hat{f}_b(x_b) - f_{bo}(x_b)| \overset{a.s.}{\to} 0$ and $sup_{x_a} |\hat{f}_a(x_a) - f_{ao}(x_a)| \overset{a.s.}{\to} 0$. ∎

PROOF OF THEOREM 5.4.1: In $\mathcal{H}^m(Q^{q_a+q_b})$ consider the subspaces:

$$G_a = \{ f_a \in \mathcal{H}^m \,|\, f_a(x_a, x_b) = f_a(x_a) \} \qquad G_b = \{ f_b \in \mathcal{H}^m \,|\, f_b(x_a, x_b) = f_b(x_b) \}$$

consisting of functions that depend only on x_a, x_b respectively. Within G_a let $r_{x_{a1}}(x_a) \ldots r_{x_{aT}}(x_a)$ be representors of function evaluation at $x_{a1} \ldots, x_{aT}$, let $M_a \subset G_a$ be the span of these representors and write $G_a = M_a \oplus M_a^{\perp}$. Similarly, within G_b let $r_{x_{b1}}(x_b) \ldots r_{x_{bT}}(x_b)$ be representors of function evaluation at $x_{b1} \ldots, x_{bT}$, let $M_b \subset G_b$ be the span of these representors and write $G_b = M_b \oplus M_b^{\perp}$. Consider $G_{ab} = \{ f_a \cdot f_b \in \mathcal{H}^m(Q^{q_a+q_b}) : f_a \in G_a, f_b \in G_b \}$. Any function in G_{ab} can be expressed in the form:

$$f_a \cdot f_b = \left(\sum_{j=1}^{T} c_{aj} r_{x_{aj}} + h_a \right) \cdot \left(\sum_{j=1}^{T} c_{bj} r_{x_{bj}} + h_b \right)$$

where $h_a \in M_a^{\perp}, h_b \in M_b^{\perp}$. The objective function for the infinite dimensional optimization problem may be rewritten as:

$$\frac{1}{T} \sum_t \left[y_t - f_a(x_{at}) \cdot f_b(x_{bt}) \right]^2$$

$$= \frac{1}{T} \sum_t \left[y_t - \left(< r_{x_{at}}, \sum_j c_{aj} r_{x_{aj}} + h_a >_{Sob} \right) \cdot \left(< r_{x_{bt}}, \sum_j c_{bj} r_{x_{bj}} + h_b >_{Sob} \right) \right]^2 = \frac{1}{T} \sum_t \left[y_t - [R_a c_a]_t \cdot [R_b c_b]_t \right]^2$$

where R_a, R_b are the representor matrices on Q^{q_a} and Q^{q_b} at $x_{a1} \ldots, x_{aT}$ and $x_{b1} \ldots, x_{bT}$ respectively. Suppose $f_a \cdot f_b = \left(\sum_t c_{aj} r_{x_{aj}} + h_a \right) \cdot \left(\sum_t c_{bj} r_{x_{bj}} + h_b \right)$ minimizes $\frac{1}{T} \sum \left[y_t - f_a(x_{at}) \cdot f_b(x_{bt}) \right]^2$ s.t. $\| f_a \cdot f_b \|^2_{Sob} \leq L$. Then so does $f_a^* \cdot f_b^* = (f_a - h_a) \cdot (f_b - h_b) = \left(\sum_t c_{aj} r_{x_{aj}} \right) \cdot \left(\sum_t c_{bj} r_{x_{bj}} \right)$ since h_a and h_b are zero at $(x_{a1}, x_{b1}), \ldots, (x_{aT}, x_{bT})$. Hence, there exists a function $f_a^* \cdot f_b^*$ minimizing the infinite dimensional optimization problem such that $f_a^* \in M_a$, $f_b^* \in M_b$. Furthermore, using Th. 2.4, $\| f_a^* \cdot f_b^* \|^2_{Sob} = \| f_a^* \|^2_{Sob} \cdot \| f_b^* \|^2_{Sob} \leq \| f_a \|^2_{Sob} \cdot \| f_b \|^2_{Sob} = \| f_a \cdot f_b \|^2_{Sob} \leq L$. Finally, observe that the following two problems yield identical minima:

$$\min_{f_a^*, f_b^*} \frac{1}{T} \sum \left[y_t - f_a^*(x_{at}) \cdot f_b^*(x_{bt}) \right]^2 \quad s.t. \quad \| f_a^* \cdot f_b^* \|^2_{Sob} \leq L, \quad f_a^* \in M_a, \quad f_b^* \in M_b$$

$$\min_{c_a, c_b} \frac{1}{T} \sum_t \left[y_t - [R_a c_a]_t \cdot [R_b c_b]_t \right]^2 \quad s.t. \quad c_a' R_a c_a \cdot c_b' R_b c_b \leq L \quad \blacksquare$$

PROOF OF PROPOSITION 5.4.2: use $sup_{x_a, x_b} | \hat{f}_a(x_a) \cdot \hat{f}_b(x_b) - f_{ao}(x_a) \cdot f_{bo}(x_b) | \overset{a.s.}{\to} 0$ and the condition $\hat{f}_a(0) = f_{ao}(0) = 1$. ■

PROOF OF PROPOSITION 5.5.2: Suppose f_o is strictly monotone increasing so that $\exists$ B>0 s.t. $f' >$B on Q^1. With m >2 second derivatives are bounded hence $sup_x |\hat{f}' - f_o'| \overset{a.s.}{\to}$ 0. Thus, with probability arbitrarily close to 1, $\hat{f}'$ is eventually bounded away from 0. ■

PROOF OF PROPOSITION 5.6.1 AND THEOREMS 5.6.2, 5.6.3: The proof of Proposition 5.6.1 is a straightforward generalization of the Proof of Th. 5.1.1. See also Proofs of Th.'s 3.1.2 and 3.2.2. Proofs of Th. 5.6.2 and 5.6.3 are similar to the Proofs of Th. 5.2.1 and 5.2.2 respectively. ■

APPENDIX 2 -- CALCULATION OF REPRESENTORS

Bracketed superscripts will denote derivatives. Letting $<f,g>_{Sob} = \int_0^1 \sum_{k=0}^{m} f^{(k)}(x)g^{(k)}(x)dx$ we construct an $r_a \in \mathcal{H}^m[0,1]$ such that $<f,r_a>_{Sob} = f(a)$ for all $f \in \mathcal{H}^m[0,1]$. This representor r_a will be of the form:

$$r_a(x) = \begin{cases} L_a(x) & 0 \le x \le a \\ R_a(x) & a \le x \le 1 \end{cases}$$

where L_a and R_a are both analytic functions. For r_a of this form to be an element of $\mathcal{H}^m[0,1]$, it suffices that $L_a^{(k)}(a) = R_a^{(k)}(a)$, $0 \le k \le m-1$. Now write:

$$f(a) = <r_a, f>_{Sob} = \int_0^a \sum_{k=0}^{m} L_a^{(k)}(x) f^{(k)}(x)dx + \int_a^1 \sum_{k=0}^{m} R_a^{(k)}(x) f^{(k)}(x)dx$$

We ask that this be true for all $f \in \mathcal{H}^m[0,1]$ but by density it suffices to demonstrate the result for all $f \in C^\infty[0,1]$. Hence assume that $f \in C^\infty[0,1]$. Thus, integrating by parts, we have:

$$\sum_{k=0}^{m} \int_0^a L_a^{(k)}(x)\, f^{(k)}(x)dx = \sum_{k=0}^{m} \left\{ \sum_{j=0}^{k-1} (-1)^j\, L_a^{(k+j)}(x)\, f^{(k-j-1)}(x) \Bigg|_0^a \right.$$

$$\left. + (-1)^k \int_0^a L_a^{(2k)}(x)\, f(x)dx \right\}$$

$$= \sum_{k=0}^{m} \sum_{j=0}^{k-1} (-1)^j\, L_a^{(k+j)}(x)\, f^{(k-j-1)}(x) \Bigg|_0^a + \int_0^a \left\{ \sum_{k=0}^{m} (-1)^k\, L_a^{(2k)}(x) \right\} f(x)dx$$

letting $i = k - j - 1$ in the first sum, this may be written as

$$\sum_{k=0}^{m} \int_0^a L_a^{(k)}(x)\, f^{(k)}(x)dx = \sum_{k=1}^{m} \sum_{i=0}^{k-1} (-1)^{k-i-1}\, L_a^{(2k-1-i)}(x)\, f^{(i)}(x) \Bigg|_0^a$$

$$+ \int_0^a \left\{ \sum_{k=0}^{m} (-1)^k\, L_a^{(2k)}(x) \right\} f(x)dx$$

$$= \sum_{i=0}^{m-1} \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-1-i)}(x)\, f^{(i)}(x) \Big|_0^a + \int_0^a \left\{ \sum_{k=0}^{m} (-1)^k L_a^{(2k)}(x) \right\} f(x)dx$$

$$= \sum_{i=0}^{m-1} f^{(i)}(a) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-1-i)}(a) \right\} - \sum_{i=0}^{m-1} f^{(i)}(0) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-i-i)}(0) \right\}$$

$$+ \int_0^a \left\{ \sum_{k=0}^{m} (-1)^k L_a^{(2k)}(x) \right\} f(x)dx$$

Similarly, $\int_a^1 \sum_{k=0}^{m} R_a^{(k)}(x)\, f^{(k)}(x)dx$ may be written as

$$- \sum_{i=0}^{m-1} f^{(i)}(a) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-1-i)}(a) \right\} + \sum_{i=0}^{m-1} f^{(i)}(1) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-1-i)}(1) \right\}$$

$$+ \int_a^1 \left\{ \sum_{k=0}^{m} (-1)^k R_a^{(2k)}(x) \right\} f(x)dx .$$

Thus we require that both L_a and R_a are solutions of the constant coefficient differential equation

(A2.1) $$\sum_{k=0}^{m} (-1)^k u^{(2k)}(x) = 0 .$$

Boundary conditions are obtained by setting the coefficient of $f^{(i)}(a)$, $1 \le i \le m-1$, $f^{(i)}(0)$, $0 \le i \le m-1$ and $f^{(i)}(1)$, $0 \le i \le m-1$ to zero and the coefficient of $f(a)$ to 1. That is,

(A2.2)
$$\begin{aligned}
&\sum_{k=i+1}^{m} (-1)^{k-i-1} \left\{ L_a^{(2k-1-i)}(a) - R_a^{(2k-1-i)}(a) \right\} = 0 && 1 \le i \le m-1 \\
&\sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-1-i)}(0) = 0 && 0 \le i \le m-1 \\
&\sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-1-i)}(1) = 0 && 0 \le i \le m-1 \\
&\sum_{k=1}^{m} (-1)^{k-1} \left\{ L_a^{(2k-1)}(a) - R_a^{(2k-1)}(a) \right\} = 1
\end{aligned}$$

Furthermore, for $r_a \in \mathcal{H}^m[0,1]$, we require, $L_a^{(k)}(a) = R_a^{(k)}(a)$, $0 \le k \le m-1$. This results in

$(m-1)+m+m+1+m = 4m$ boundary conditions. The general solution of the differential equation (A2.1) is obtained by finding the roots of its characteristic polynomial $P_m(\lambda) = \sum_{k=0}^{m} (-1)^k \lambda^{2k}$. This is easily done by noting that $(1+\lambda^2)\, P_m(\lambda) = 1 + (-1)^m \lambda^{2m+2}$ and thus the characteristic roots are given by $\lambda_k = e^{i\theta_k}$, $\lambda_k \neq \pm i$, where

$$\theta_k = \begin{cases} \dfrac{(2k+1)\pi}{2m+2} & m \text{ even} \\[2ex] \dfrac{2k\,\pi}{2m+2} & m \text{ odd} \end{cases}$$

The general solution is given by the linear combination $\sum_k a_k\, e^{(Re(\lambda_k))x} \sin(Im(\lambda_k))x$ where the sum is taken over $2m$ linearly independent real solutions of (A2.1).

Let $L_a(x) = \sum_{k=1}^{2m} a_k u_k(x)$ and $R_a(x) = \sum_{k=1}^{2m} b_k u_k(x)$ where the u_k, $1 \leq k \leq 2m$ are $2m$ basis functions of the solution space of (A2.1). To show that r_a exists and is unique, we need only show that the boundary conditions (A2.2) uniquely determine the a_k and b_k. Since we have $4k$ unknowns *(2m a_k's and 2m b_k's)* and $4m$ boundary conditions, (A2.2) is in fact a square $4m$, x $4m$ linear system in the a_k's and b_k's. Thus it suffices to show that the only solution of the associated homogenous system is the zero vector. Now suppose that $L_a^h(x)$ and $R_a^h(x)$ are the functions corresponding to the solutions of the homogeneous system (i.e. with the coefficient of $f(a)$ in (A2.2) set to 0 instead of 1). Then, by exactly the same integration by parts, it follows that $< r_a^h, f >_{Sob} = 0$ for all $f \in C^\infty [0,1]$. Hence r_a^h , $L_a^h(x)$ and $R_a^h(x)$ are all identically zero and thus by the linear independence of the $u_k(x)$, so are the a_a and b_k. ■