Consistent Estimation of Shape-Restricted Functions and Their Derivatives

Pok Man Chak* Neal Madras† J. Barry Smith‡

November 15, 2001§

Abstract

We examine the estimation problem for shape-restricted functions that are continuous, non-negative, monotone non-decreasing, and strictly concave. A sieve estimator based on bivariate Bernstein polynomials is proposed. This estimator is drawn from a sieve, a set of shape-restricted Bernstein polynomials, which grows with the sample size in such a way that it becomes dense in the set of shape-restricted functions. Under some mild conditions, we show that this sieve estimator of the true function and the estimators of its first and second derivatives are uniformly consistent. The estimators of elasticities of substitution are uniformly consistent as well.

Keywords and Phrases: shape-restricted functions, bivariate Bernstein polynomials, flexible functional forms, sieve estimator, uniform consistency.

JEL Classification: C13, C14, C15, C51.

---

*Department of Economics, York University, Toronto, Ontario, Canada, M3J 1P3. E-mail address: pmc@dept.econ.yorku.ca.

†Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada, M3J 1P3. E-mail address: madras@mathstat.yorku.ca. The author would like to acknowledge the support by the Natural Sciences and Engineering Research Council of Canada of this work.

‡Department of Economics, York University, Toronto, Ontario, Canada, M3J 1P3. E-mail address: jbsmith@rsrch.econ.yorku.ca.

§The authors would like to thank P. Rilstone, T. Stengos and A. Wong for their helpful comments.

1 Introduction

Considerable effort continues to be directed towards obtaining estimates and estimators of the cost, production and demand models arising in economic theory. Many important advances have been made in both understanding and solving the statistical problems that arise in trying to estimate functions which are known only up to a set of properties which they must possess in order to be regular or consistent with economic theory. Notwithstanding the considerable strides that have been made, many issues remain unresolved.

The first generation of parametric models includes, amongst others, Cobb-Douglas and CES cost and production models. With standard assumptions on errors, these finite parameter models yield consistent and asymptotically normal estimates of a true underlying model only if the underlying model is Cobb-Douglas or CES.

The second generation of finite parameter models includes, amongst others, Generalized Leontief [Diewert (1971)], Translog [Christensen, Jorgenson and Lau (1971)] and Generalized Square Root Quadratic [Diewert (1974)] models. As with the first generation models (FGM's), these second generation models (SGM's) are capable of providing consistent and asymptotically normal estimates of the true underlying model only if the true model falls in the given finite parameter class. SGM's dominate FGM's to the extent that a given SGM can typically embed several FGM's. Moreover, in an approximation sense, the SGM's often dominate FGM's in terms of flexibility. In particular, SGM's can provide a second order approximation of an arbitrary function at a specified point. As White (1980) has carefully shown, though, flexibility in approximation does not imply flexibility in estimation. Estimates of function values and the first two (orders of) derivatives of a second order flexible finite parameter model are typically not consistent for a general true model. Attempts to rank SGM's in terms of quality of fit have shown that the "best" SGM may change depending upon the value of the unknown parameters of the true model [see, for example, Guilkey, Lovell and Sickles (1983)].

A third generation of parametric models (TGM's) has arisen from the pioneering work of Gallant (1981, 1982), El Badawi, Gallant and Souza (1983), and Gallant and Souza (1991). This literature considers sequences of parametric models which evolve towards an infinite parameter model as the sample size increases. Each element in the sequence of parametric models is drawn from a family of functions. For example, Gallant's work involves Fourier series augmented by a quadratic term. It is important to choose a family of functions (e.g. Fourier) such that the sequence of models becomes dense in the set of functions of interest (e.g. cost or production functions with given properties). The statistical problem involves showing that a growing set of estimated models will converge to (be consistent for) the true underlying function. El Badawi, Gallant and Souza (1983) have shown that the Fourier sequence can provide consistent estimates of elasticities of unknown models. Moreover, estimators of the elasticities can be asymptotically normally distributed [see Gallant and Souza (1991)]. An important issue regarding consistency and asymptotic normality is the rate at which the number of parameters increases for models in the sequence. This is the classical problem of estimation for infinite parameter models. El Badawi, Gallant and Souza (1983) use the term "semi-nonparametric regression" to describe their model. The fact that the parametric model "grows" or "adapts" with the sample size provides a nonparametric component to the analysis.

To this point, the only valid TGM is that provided by Gallant et. al.. As discussed in Diewert and Wales (1993), other attempts such as the Asymptotically Ideal Model (AIM) of Barnett and Jonas (1983) and the Globally Flexible Almost Ideal Demand System (GFAIDS) of Chalfant (1987) require restrictions which prevent them from being fully flexible and thus consistency cannot be achieved except in restricted cases.

Consistent with the directions for future research identified by Gallant (1984), this paper introduces a new third generation parametric model based on Bernstein polynomials. For each sample size we define a sieve or set of shape-restricted Bernstein polynomials. The best fit of the data is obtained by choosing one element from the relevant sieve. The sieve grows with the sample size in such a way that it becomes dense in the set of all functions of interest. Under mild and reasonable conditions, we show that the estimator is uniformly consistent. Moreover, increasing orders of derivatives of the true functions are estimated in a uniformly consistent fashion. As with all of the literature, the method of proof of results to some extent depends upon the specific family of functions from which the third generation sequence is drawn. One convenient feature of working with Bernstein polynomials is that proofs of flexibility, approximation and consistency can be presented in a straightforward and compact fashion. Our results imply uniform Sobolev convergence to the true underlying data generation process. As such we obtain consistent estimates of elasticities just as in the Fourier case. Numerically our point estimates arise as the solution of a classical quadratic programming problem.

The rest of the paper is organized as follows. In section 2, we introduce our model and the Bernstein sieve estimator. The technical results of our sieve estimator and proofs are presented in Section 3. In Section 4, we present some simulation results. Some conclusions are offered in Section 5.

2 The Model and The Estimator

Our model[1] in this paper is

$$Y_i \,=\, h(X_{1i}, X_{2i}) \,+\, \epsilon_i\,,$$

where (X_{1i}, X_{2i}, Y_i) are *i.i.d.* on $[0,1]\times[0,1]\times[0,\bar{d}]$ with probability measure P and $E_P[\epsilon_i|X_{1i}, X_{2i}] = 0$, and h is an arbitrary continuous, non-negative, monotone non-decreasing and strictly concave function with continuous third derivatives on $(0,1)\times(0,1)$. We assume that $Pr[X_i = 0] = 0$, $i = 1, 2$. h here can be regarded as a strictly concave function that has been transformed from a homogeneous of degree one, concave and strictly quasi-concave unit cost function by taking one of the variables as numeraire.[2] Thus, X_1 and X_2 are factor ratios. By doing this, we can avoid any problems arising from Hessians with zero determinants in approximation [see Chak (2001)]. Moreover, consistency of the estimator of the homogeneous function can be deduced from the consistency result of h. The regression of Y on X_1 and X_2 is

$$E_P[Y_i|X_{1i} = x_{1i}, X_{2i} = x_{2i}] \,=\, h(x_{1i}, x_{2i})\,,$$

[1]In this paper, we focus on the bivariate case. Note that the results for the bivariate case can be easily extended to the univariate case and higher dimensions

[2]Let $f = f(w_1, w_2, w_3)$ be a homogeneous of degree one, concave and strictly quasi-concave function. It can be shown that $h(x_1, x_2) = f\left(\frac{w_1}{w_3}, \frac{w_2}{w_3}, 1\right)$ is strictly concave [see Chak (2001)].

which is just the unknown function. The data available for the estimation are realization triples (x_{1i}, x_{2i}, y_i) satisfying

$$y_i \;=\; h(x_{1i}, x_{2i}) + \mathrm{e}_i \,,$$

where e_i's are the unobserved realizations of ϵ_i's.

Let Ψ_0 be the set of all continuous, nonnegative functions h from $[0,1] \times [0,1]$ into $[0, \bar{d}]$ with the property that h is monotone non-decreasing, strictly concave in C^3 on $(0,1) \times (0,1)$. To estimate h consistently, one needs to construct a sequence of estimators $\{\hat{h}_n\}$ that lies in Ψ_0 and converges to h in the $L_2(P)$ metric space. Consistent estimation for the derivatives of h leads to consistent estimates of the elasticities of substitution. In this paper, we propose a sieve estimator based on Bernstein polynomials which has this desirable property. Before introducing this estimator, we first define the polynomials and then present their shape-preserving properties in approximation.

The two-variable Bernstein polynomial approximation to a function $h(x_1, x_2)$ takes the form:

$$\begin{aligned} B^h_{n_1,n_2}(x_1, x_2) \;&=\; \sum_{v_1=0}^{n_1} \sum_{v_2=0}^{n_2} h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) \binom{n_1}{v_1} \binom{n_2}{v_2} \\ &\qquad {x_1}^{v_1} {x_2}^{v_2} (1-x_1)^{n_1-v_1} (1-x_2)^{n_2-v_2} \\ &=\; \sum_{v_1=0}^{n_1} \sum_{v_2=0}^{n_2} h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) P_{v_1,n_1}(x_1) P_{v_2,n_2}(x_2) \,, \end{aligned}$$

where

$$P_{v,n}(x) = \binom{n}{v} x^v (1-x)^{n-v} \,.$$

$$(x_1, x_2) \in [0,1] \times [0,1] \,, \;\; \text{and} \;\; n_1, \; n_2 \in \mathbb{N} \,.$$

Note that $B^h_{n_1,n_2}(x_1, x_2)$ is the mean of $h(x_1, x_2)$ given Binomial distributions for

x_1 and x_2. Bernstein polynomials have some nice properties and one of them is described in the following theorem:

Theorem 1: [Butzer (1953)] *If h is continuous on the closed unit square with continuous m^{th} partial derivatives on the open unit square: $0 < x_1 < 1,\ 0 < x_2 < 1$, then we have the following result:*

$$\lim_{n_1 \to \infty, n_2 \to \infty} \frac{\partial^m B^h_{n_1,n_2}}{\partial x_1^q \partial x_2^{m-q}}(x_1, x_2) \ = \ \frac{\partial^m h}{\partial x_1^q \partial x_2^{m-q}}(x_1, x_2), \qquad m = 0, 1, \cdots,$$

provided that, for any two finite positive numbers r and t,

$$0 < r \leq \frac{n_1 + 1}{n_2 + 1} \leq t < \infty\,.$$

If the partial derivatives of h are continuous on the closed unit square, then the convergence is uniform without restriction on n_1 and n_2.[3]

Proof: See Butzer (1953) and Kingsley (1951).

The above results show that, given the restriction on the growth of n_i, $i = 1, 2$, the sequences of derivative functions of the bivariate Bernstein polynomials converge at least pointwise to those of their respective true functions as $n_1, n_2 \to \infty$. This means that two-dimensional Bernstein polynomials are shape-preserving functions for large n_1 and n_2, and hence global flexibility is guaranteed. If the true function has continuous derivatives, then we have uniform convergence and *uniform flexibility.*

Making use of this elegant result in approximation, we construct our estimators

[3]This result is based on a theorem in Butzer (1953). Butzer's theorem is more general than this, allowing h to be merely bounded on $[0, 1] \times [0, 1]$, and asserting convergence at any point (x_1, x_2) for which the total differential of h exists.

based on Bernstein polynomials which are defined as follows:

$$\begin{aligned} B_{n_1,n_2}(x_1,x_2) &= \sum_{v_1=0}^{n_1}\sum_{v_2=0}^{n_2} c_{v_1,v_2}\binom{n_1}{v_1}\binom{n_2}{v_2} \\ &\qquad {x_1}^{v_1}{x_2}^{v_2}(1-x_1)^{n_1-v_1}(1-x_2)^{n_2-v_2} \\ &= \sum_{v_1=0}^{n_1}\sum_{v_2=0}^{n_2} c_{v_1,v_2}P_{v_1,n_1}(x_1)P_{v_2,n_2}(x_2)\,, \end{aligned}$$

where c_{v_1,v_2}'s are coefficients that are restricted in such a way that B_{n_1,n_2} is non-negative, monotone non-decreasing and strictly concave.[4] To carry out our estimation, we use the sieve method introduced by Grenander in 1981 as described in Grenander (1981), Geman (1981), and Geman and Hwang (1982). In this method, the best fitting restricted function is chosen from a subset of the parameter space, and this subset is allowed to grow with the sample size. This sequence of subsets is called a *sieve*. We define a stage "n" sieve as follows:

$$S_n^{[A]} = \left\{ h_n \equiv B_{n_1,n_2} \;\middle|\; B_{n_1,n_2} \in \Psi_0,\ \left|\frac{\partial^3 B_{n_1,n_2}}{\partial x_1^q\,\partial x_2^{3-q}}\right| \leq A,\ q=0,1,2,3 \right\},$$

where n_1, n_2 are functions of n satisfying the condition in Theorem 1. Here, n is the sample size chosen before estimation is performed, and hence may not be optimal. The constant A can be finite or infinite. Also let $\Psi_0^{[A]}$ be the set of functions h in Ψ_0 that satisfy the third-derivative bounds

$$\left|\frac{\partial^3 h}{\partial x_1^q \partial x_2^{3-q}}\right| \leq A, \qquad q=0,1,2,3.$$

Note that $S_n^{[A]}$ is a subset of $\Psi_0^{[A]}$. Moreover, $S_{n-1}^{[A]} \subset S_n^{[A]}$, $\forall\, n$, and thus the sieve grows with the sample size n. For each n, the optimal estimate $\hat{h}_n$ is selected from

[4]One principal difference between Bernstein approximation and Bernstein estimation is the restrictions on coefficients. In approximation, without coefficient restrictions, the bivariate Bernstein polynomials of non-negative, monotone non-decreasing and strictly concave functions are always non-negative, monotone non-decreasing, but not necessarily concave for small indexes. However, this concavity violation does not occur to the univariate approximating Bernstein polynomials of concave functions for any index.

the Bernstein polynomials in $S_n^{[A]}$, and is defined as follows:

$$\hat{h}_n = \underset{h_n \in S_n^{[A]}}{arg\,min} \; \frac{1}{n}\sum_{i=1}^{n}\Big\{y_i - h_n(x_{1i}, x_{2i})\Big\}^2 .$$

Choosing $\hat{h}_n$ is equivalent to choosing a vector of $(n_1+1)(n_2+1)$ optimal, restricted coefficients $\{\hat{c}_{0,0}, \hat{c}_{1,0}, \ldots, \hat{c}_{n_1,n_2}\}$. This optimization problem can be solved by quadratic programming because the constraints are all linear in the parameters.

This constrained least squares sieve estimator inherits the shape-preserving property of Bernstein polynomials in approximation which enables us to estimate the true function as well as its derivatives consistently. Moreover, under some mild assumptions, uniform consistency results can also be obtained. All these results will be discussed in the next section.

3 Consistency of the Estimator and its First and Second Derivatives

In this section, we present the consistency theorems of the estimator and its first and second derivatives. We begin with a consistency result for the function. In the following theorem, we prove the almost sure convergence of the sequence of constrained least squares sieve estimators $\{\hat{h}_n\}$ to the true function h in the $L_2(P)$ metric space. By definition, the convergence can be expressed as the following:

$$\int_{[0,1]\times[0,1]} \Big|\hat{h}_n(x_1, x_2) - h(x_1, x_2)\Big|^2 dP(x_1, x_2) \longrightarrow 0 \;\; a.s. \;\text{ as } n \to \infty.$$

Since our estimating and true functions lie in the $L_2(P)$ metric space, we can also write the convergence in the following way:

$$E_P\Big[\hat{h}_n(X_1, X_2) - h(X_1, X_2)\Big]^2 \longrightarrow 0 \;\; a.s. \;\text{ as } n \to \infty,$$

where E_P is the expectation taken with respect to the probability measure P on (X_1, X_2, Y).

Theorem 2: *Let $h \in \Psi_0^{[\infty]}$. Also let $\hat{h}_n \in S_n^{[\infty]}$, where $\hat{h}_n$ is the constrained least squares sieve estimator. Then the sequence $\{\hat{h}_n\}$ converges to h on $[0,1] \times [0,1]$ in the $L_2(P)$ metric space almost surely. That is,*

$$E_P\Big[\hat{h}_n(X_1, X_2) - h(X_1, X_2)\Big]^2 \longrightarrow 0 \quad a.s. \quad as\ n \to \infty.$$

The result also holds if we assume that $h \in \Psi_0^{[D]}$ and $\hat{h}_n \in S_n^{[A]}$, $D < A < \infty$.

Proof: See the Appendix.

Notice that by letting $\epsilon_0 = 0$ if $A < \infty$, and letting $\epsilon_0 > 0$ if $A = \infty$, it can be shown that the set $\Psi_0^{[A]}$ is equicontinuous on $[\epsilon_0, 1] \times [\epsilon_0, 1]$.[5] See Lemma 2 in the Appendix. Also, pointwise convergence can be obtained by making use of the monotone non-decreasing property of $\hat{h}_n$ and the $L_2(P)$ convergence. Thus, with equicontinuity and pointwise convergence, uniform convergence in probability on a compact set[6] is established. This is described in the next theorem.

Theorem 3: *Let $\hat{h}_n \in S_n^{[A]}$, where $\hat{h}_n$ is the constrained least squares sieve estimator, and suppose that $\{\hat{h}_n\}$ converges to h on $[0,1] \times [0,1]$ in the $L_2(P)$ metric space. Let $\epsilon_0 = 0$ if $A < \infty$, and let $\epsilon_0 > 0$ if $A = \infty$. Let supp(P) denote the support of (X_1, X_2) with respect to P. Then with probability 1, $\{\hat{h}_n\}$ converges to h uniformly on the closure of int(supp$(P)) \cap [\epsilon_0, 1] \times [\epsilon_0, 1]$, where int(supp$(P)$) denotes the interior of (supp(P)).*

[5] It turns out that this equicontinuity condition plays a very important role in establishing uniform convergence of $\{\hat{h}_n\}$ and its associated first and second derivatives.

[6] Here we prove the "strong" or "deterministic" equicontinuity of our sequence of estimators. A stochastic equicontinuity condition is developed in Newey (1991). In Newey's paper, stochastic equicontinuity and pointwise convergence are used to establish uniform convergence in probability to equicontinuous functions on a compact set.

Proof: See the Appendix.

The above theorem takes into consideration the cases where the support may have irregular shape on the unit square. The result provided in the following remark below assumes that the support is regular.

Remark: If *supp*(P) contains $[\epsilon_0, 1] \times [\epsilon_0, 1]$, $\epsilon_0 \geq 0$, then with probability 1 $\{\hat{h}_n\}$ converges to h uniformly on $[\epsilon_0, 1] \times [\epsilon_0, 1]$.

Next we have a consistency theorem for the first derivatives of $\hat{h}_n$.

Theorem 4: *Let $\hat{h}_n \in S_n^{[A]}$, where $\hat{h}_n$ is the constrained least squares sieve estimator. Also assume $Pr[int(supp(P))] = 1$. Let $\epsilon_0 = 0$ if $A < \infty$, and let $\epsilon_0 > 0$ if $A = \infty$. Then for $i = 1, 2$, $\left\{\frac{\partial \hat{h}_n}{\partial x_i}\right\}$ converges to $\frac{\partial h}{\partial x_i}$ on $[\epsilon_0, 1] \times [\epsilon_0, 1]$ pointwise on int(supp(P)) and in the $L_2(P)$ metric space.*

Proof: See the Appendix.

Also, Theorem 4 is obtained with the assumption that the support may have irregular shape, and holes may even occur. The result given in the next remark excludes all the irregularities of the support and adds the condition that all the second derivatives of the constrained least squares sieve estimator are bounded.

Remark: If $A < \infty$, then for some constant K we have $\left|\frac{\partial^2 \hat{h}_n}{\partial x_1^q \, \partial x_2^{2-q}}\right| \leq K$, $q = 0, 1, 2$. Thus $\left\{\frac{\partial \hat{h}_n}{\partial x_1}\right\}$ and $\left\{\frac{\partial \hat{h}_n}{\partial x_2}\right\}$ are equicontinuous sequences on $[0, 1] \times [0, 1]$. Also, if *supp*(P) contains $[\epsilon_0, 1] \times [\epsilon_0, 1]$, $\epsilon_0 \geq 0$, then by Lemma 6 in the Appendix, $\left\{\frac{\partial \hat{h}_n}{\partial x_i}\right\}$ converges uniformly to $\frac{\partial h}{\partial x_i}$, $i = 1, 2$, on $[\epsilon_0, 1] \times [\epsilon_0, 1]$ with probability 1.

Again, equicontinuity is used to establish uniform convergence.

We proceed to the consistency theorem for the second derivatives of $\hat{h}_n$.

Theorem 5: *Let $D < A < \infty$. Assume that $h \in \Psi_0^{[D]}$. Let $\hat{h}_n \in S_n^{[A]}$, where $\hat{h}_n$ is the constrained least squares sieve estimator. Also assume $Pr[int(supp(P))] = 1$. Then for $q = 0, 1, 2$, $\left\{\frac{\partial^2 \hat{h}_n}{\partial x_1^q \partial x_2^{2-q}}\right\}$ converges to $\left\{\frac{\partial^2 h}{\partial x_1^q \partial x_2^{2-q}}\right\}$ on $[0,1] \times [0,1]$ pointwise on $int(supp(P))$ and in the $L_2(P)$ metric space.*

Proof: See the Appendix.

Remark: Since $A < \infty$, then for $q = 0, 1, 2$, $\left\{\frac{\partial^2 \hat{h}_n}{\partial x_1^q \partial x_2^{2-q}}\right\}$ is equicontinuous sequence on $[0,1] \times [0,1]$. If $supp(P)$ contains $[0,1] \times [0,1]$,[7] then by Lemma 6, with $q = 0, 1, 2$, $\left\{\frac{\partial^2 \hat{h}_n}{\partial x_1^q \partial x_2^{2-q}}\right\}$ converges to $\left\{\frac{\partial^2 h}{\partial x_1^q \partial x_2^{2-q}}\right\}$ uniformly on $[0,1] \times [0,1]$ with probability 1.

Remark: Since uniform convergence of the sequences of estimators and their first and second partial derivatives is established, uniform convergence of the sequence of estimators of elasticities of substitution follows by standard arguments.

In addition to these strong results, we also find that if $\{\hat{h}_n\} \in S_n^{[A]}$ converges to h uniformly on $[\epsilon_0, 1] \times [\epsilon_0, 1]$ with probability 1, $\epsilon_0 \geq 0$, then the sequence of coefficients corresponding to $\{\hat{h}_n\}$ also converges to the sequence of coefficients of $\{B^h_{n_1,n_2}\}$. This result is described by the theorem below.

Theorem 6: *Let $\hat{h}_n \in S_n^{[A]}$, where $A < \infty$ or $A = \infty$, and $\hat{h}_n$ is the constrained least squares sieve estimator. Also let*

$$\hat{h}_n(x_1, x_2) = \sum_{v_1=0}^{n_1} \sum_{v_2=0}^{n_2} \hat{c}_{v_1,v_2} P_{v_1,n_1}(x_1) P_{v_2,n_2}(x_2) .$$

If $\{\hat{h}_n\}$ converges to h uniformly on $[\epsilon_0, 1] \times [\epsilon_0, 1]$ with probability 1, $\epsilon_0 \geq 0$, then

[7] The support is defined to be the smallest closed set whose complement has probability zero.

for any $\epsilon_1 > \epsilon_0$*, with probability 1 we have*

$$\max_{v_1,v_2:\frac{v_1}{n_1}\geq\epsilon_1,\frac{v_2}{n_2}\geq\epsilon_1}\left|\hat{c}_{v_1,v_2} - h\left(\frac{v_1}{n_1},\frac{v_2}{n_2}\right)\right| \longrightarrow 0 \ \ a.s. \ \ as \ \ n_1,n_2\to\infty\,.$$

Proof: See the Appendix.

Note that the Bernstein coefficients $h\left(\frac{v_1}{n_1},\frac{v_2}{n_2}\right)$ are the points on the surface of h. So the convergence of $\{\hat{c}_{v_1,v_2}\}$ to $\left\{h\left(\frac{v_1}{n_1},\frac{v_2}{n_2}\right)\right\}$ provides a means of visualization of the convergence of the estimating surfaces $\{\hat{h}_n\}$ to the true surface h.

4 Monte Carlo Simulations

4.1 Background

In this section we use the results of several Monte Carlo experiments to deduce further properties of the proposed estimator. In large measure our approach follows that of Chalfant and Gallant (1985). Our approach differs in two important ways. First, we adopt a traditional specification of error terms and leave parts of the errors in variables analysis of Chalfant and Gallant to later work. We allow for varying quality of sample information in an alternative fashion. Second, because our approach provides a consistent estimator of elasticities of substitution as well as the cost function and all of its derivatives, we can provide more information about how the model fits a "representative" data set. These additional results are useful, in part, because they point to how share equation information is crucial in obtaining accurate estimates of higher order derivatives (and elasticities) in small samples. Overall, the Monte Carlo results suggest that the estimator provides extremely accurate estimates of an unknown technology and its properties even in very small samples. Moreover, the quality of the estimator is not diminished as the model, to some extent, becomes highly overparameterized.

4.2 Description of Experiments

Following Chalfant and Gallant (Table 2, p.214), we adopt as our true unknown cost function the Translog technology defined by:

$$\ln C = \alpha_0 + \sum_{i=1}^{3} \alpha_i \ln P_i + \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \gamma_{ij} \ln P_i \ln P_j \tag{1}$$

with the parameter values: $\alpha_0 = 0.756928$, $\alpha_1 = 0.191614$, $\alpha_2 = 0.208386$, $\alpha_3 = 0.6$, $\gamma_{11} = 0.08$, $\gamma_{12} = -0.08$, $\gamma_{22} = 0.08$, and $\gamma_{13} = \gamma_{23} = \gamma_{33} = 0$. The data are normalized so that $\ln P_i = 0$, $i = 1, 2, 3$, at the sample mean. The second row of Table 1 provides the "true" values of the cost function, its derivatives and elasticities of substitution at the mean of the data.

As described in Chalfant and Gallant, the time series (P_{it}) are chosen so that they are representative of the colinear series one typically encounters in practice. This is accomplished as follows: First, the aggregate US data provided in Berndt and Wood (1975) and Berndt and Khaled (1979) are used to create time series for the prices of labour (P_{3t}), capital (P_{2t}) and an aggregate of energy and materials (P_{1t}). As mentioned in Chalfant and Gallant, autoregressive processes are estimated and the fitted data are used as the "true" price data. In principle, data series of any length can be constructed in this way. We restricted attention to the case of 25 observations.

The true series for $\ln C$ is obtained by evaluating equation (1) at each data point. The series for C required for the Bernstein sieve estimator is obtained by taking the exponential of the logarithmic cost series. In the numeric work we use

$\frac{C}{P_2}$ as the true cost variable. As well, the variables P_1 and P_3 are transformed to P_1^* and P_3^* where:

$$P_i^* = \frac{\frac{P_i}{P_2} - \min\left(\frac{P_i}{P_2}\right)}{\max\left(\frac{P_i}{P_2}\right) - \min\left(\frac{P_i}{P_2}\right) + 0.01} \qquad i = 1, 3.$$

This transformation guarantees that P_1^* and P_3^* both lie in the (half) open interval $[0, 1)$. Chalfant and Gallant also scale their data.

At this point our experimental design begins to differ somewhat from that of Chalfant and Gallant. In particular, we investigate the behaviour of an estimator for both single and multi-equation models. One of our design parameters is the strength of the signal in the data. As such, we obtain an indication of the properties of our estimator with increasingly noisy data in small samples.

In the case of single equation models, we proceed as follows. We begin with a series for $\frac{C}{P_2}$ obtained by transforming equation (1). We then create an "observed" dependent variable series as

$$\left(\frac{C}{P_2}\right)_t^* = \left(\frac{C}{P_2}\right)_t + s(r^2)\epsilon_t\,, \tag{2}$$

where ϵ is *i.i.d.* $N(0, 1)$ and $s(r^2)$ is a scaling factor which depends upon the parameter r^2. If r^2=1, then $s(r^2)$=0. Similarly, for values of r^2=0.9, 0.7 and 0.5, $s(r^2)$ is set so that $R^2 = 1 - \frac{ESS}{TSS}$ for the equation equals the corresponding r^2 on average. Thus, as r^2 decreases we consider increasingly noisy models. It is possible to fit the r^2=1 model because the Bernstein model does not nest the true model. Hence, the r^2=1 case measures the effect of specification error alone. For smaller values of r^2 the error will have both specification and noise components. We choose a range of Bernstein models ranging from n_1=n_2=2 to n_1=n_2=5. The number of estimated parameters is $(n_1 + 1)(n_2 + 1)$ and ranges from 9 to 36.

The case of the multi-equation model is similar.[8] We estimate the cost function as well as demand functions for the aggregate input and labour. Each share (demand equation) is constructed as in equation (2). In particular, the true demand is obtained from the true cost model by differentiation with respect to the factor price. These derivative series become the conditional mean series for the stochastic demand equations. Errors in share and cost equations are independently Normally distributed but in general differ with respect to variance. For example, when all equations are assumed to have $r^2 = 0.7$ (say), the scaling term in each equation will differ because of differences in the variance of the (true) conditional mean series for the factor. The system of equations is estimated with a SURE estimator for a variety of r^2 values and differing numbers of parameters as in the single equation case.

For all of the simple and multi-equation models and for all values of r^2, n_1 and n_2 we choose the sample size n=25 and estimate each model with 5000 (sets of) error terms. The averages over these 5000 repetitions are computed at the mean of the data and reported, along with the true values, in Table 1. Although we do not present the results in detail, we also collected information on the observed variances of the point estimates.

We used a quadratic programming algorithm to estimate the parameters of the models. The estimated cost functions were constrained to be non-negative, monotone non-decreasing and concave. The absolute value of all third derivatives

[8] Strictly speaking, our proofs of consistency apply to the single equation case. Our estimator leads to estimators of first and second derivatives and these estimators were shown to converge in Sobolev norm even when no cost share information explicitly enters the estimation process. We fully expect that consistency properties will continue to hold when even more share information is introduced. We are now in the process of proving this result.

was constrained to be less than a fixed value. The results at the sample mean were not sensitive to this restriction perhaps because the true function does not have highly variable second derivatives throughout the sample. As the third derivative bound is set to smaller values, the algorithm takes a longer time to converge.

4.3 Discussion of the Results

On balance the Monte Carlo results are extremely encouraging. As the results in Table 1 show, the system of equations estimator provides virtually unbiased estimates of the properties of an unknown technology. The results do not seem to be sensitive to increases in the number of parameters entering the model. As well, noisy data do not seem to degrade the performance of the estimator. The curvature restrictions appear to sharpen the focus upon the signal component on the data. Third derivative constraints operate in an interesting fashion. As discussed in Chak (2001), combined with the curvature restrictions, they serve to temper oscillatory behaviour in the estimated second derivative functions as the number of parameters increases. The curvature restrictions keep the estimated function from interpolating the data as the number of parameters (and hence, the order of the Bernstein polynomials) increases. In the absence of third derivative bounds, overfitting still manifests itself but now it appears as rapid changes in the size of (but not sign of) the second derivatives. This unconstrained behaviour in the changes in second derivatives provides a partial explanation of why third derivative bounds are needed to prove consistency of the Bernstein estimator of second derivatives.

One result that stands out in small samples is that the single equation model was not able to provide accurate estimates of second derivatives and the elasticities of substitution. While this stands in contrast to the multi-equation results, it should not be considered surprising. Equally poor results arise when one attempts to estimate higher order derivatives and elasticities using the same data but with the "true" translog cost function replacing the Bernstein polynomials. Colinear data combined with a growing amount of noise quickly degrade the performance of even the true single equation model.

We plotted the distributions of the estimated parameters and computed the standard errors. In all cases the distributions were unimodal. In the multi-equation case, the shape was Normal. As expected, estimated standard deviations increased as the data became more noisy.

5 Conclusion

In this paper we have presented a new estimator of shape-restricted technologies based upon Bernstein polynomials. The results are quite promising. The estimator provides uniformly consistent estimates of the unknown technology and its derivatives. Monte Carlo results further suggest that bias in the estimates is small even in very small samples. It appears that overparameterization is not an issue but that share equation information is important in small samples.

We are presently extending the research in several directions. We are attempting to derive the limiting distribution of the estimator. One promising result in this direction comes form the Monte Carlo results where it appears that the parameter estimates follow a Normal distribution even in small samples. We are considering the application of the estimator to a variety of applied problems and we will compare the performance of the estimator with others including the Fourier series model of Gallant et. al. We are examining the use of the estimator in a variety of other settings where the restrictions include quasi-convexity and quasi-concavity. Finally, we are examining whether bootstrap techniques can be usefully employed in small samples.

Appendix

Lemma 1: *If $A < \infty$, then for every $h \in \Psi_0^{[A]}$, we have*

$$\left|\frac{\partial h}{\partial x_i}\right| \leq 4\bar{d} + \frac{A}{2} \quad \textit{and} \quad \left|\frac{\partial^2 h}{\partial x_1^q \partial x_2^{2-q}}\right| \leq 4\bar{d} + A$$

at every point of $[0,1] \times [0,1]$, $i = 1, 2$, $q = 0, 1, 2$.

Proof: Let $x_2 = a$, an arbitrary point in $[0, 1]$. Since $A < \infty$, then for $y, z \in [0, 1]$ we have

$$\left|\frac{\partial^2 h}{\partial x_1^2}(y,a) - \frac{\partial^2 h}{\partial x_1^2}(z,a)\right| \leq A|y-z| \ . \tag{A.1}$$

Also by concavity

$$\frac{\partial h}{\partial x_1}(y,a) \ \leq \ \frac{h(y,a) - h(0,a)}{y - 0} \ \leq \ \frac{\bar{d} - 0}{y} \ = \ \frac{\bar{d}}{y} \ .$$

Now let $y < z$. Then

$$\begin{aligned} & \frac{\partial h}{\partial x_1}(y,a) - \frac{\partial h}{\partial x_1}(z,a) \\ = \ & \int_y^z \left[-\frac{\partial^2 h}{\partial x_1^2}(t,a)\right] dt \\ \geq \ & \inf_t \left|\frac{\partial^2 h}{\partial x_1^2}(t,a)\right| |y-z| \ . \end{aligned}$$

Thus

$$\frac{\bar{d}}{\frac{1}{2}} \geq \frac{\partial h}{\partial x_1}(\frac{1}{2}, a) \geq \frac{\partial h}{\partial x_1}(\frac{1}{2}, a) - \frac{\partial h}{\partial x_1}(1, a) \geq \inf_t \left|\frac{\partial^2 h}{\partial x_1^2}(t,a)\right| \left|1 - \frac{1}{2}\right| \ .$$

So

$$\inf_t \left| \frac{\partial^2 h}{\partial x_1^2}(t,a) \right| \leq 4\bar{d},$$

and hence by $(A.1)$

$$\sup_t \left| \frac{\partial^2 h}{\partial x_1^2}(t,a) \right| \leq 4\bar{d} + A.$$

Since $a \in [0,1]$ is arbitrary, this proves the Lemma's second derivative bound for $q = 2$. Similar arguments work for $q = 0$. For $q = 1$, it follows from concavity that

$$\frac{\partial^2 h}{\partial x_1^2} \frac{\partial^2 h}{\partial x_2^2} - \left(\frac{\partial^2 h}{\partial x_1 \partial x_2} \right)^2 \geq 0\,,$$

which implies that

$$\left| \frac{\partial^2 h}{\partial x_1 \partial x_2} \right| \leq \sqrt{\frac{\partial^2 h}{\partial x_1^2} \frac{\partial^2 h}{\partial x_2^2}}\,.$$

Therefore

$$\sup_t \left| \frac{\partial^2 h}{\partial x_1 \partial x_2}(t,a) \right| \leq 4\bar{d} + A,$$

and this completes the proof of the Lemma's second derivative bounds.

Note next that

$$\frac{\partial h}{\partial x_1}(0,a) - \frac{\partial h}{\partial x_1}(\frac{1}{2},a) \leq \sup_t \left| \frac{\partial^2 h}{\partial x_1^2}(t,a) \right| \left| \frac{1}{2} - 0 \right|.$$

This comes to

$$\begin{aligned} \frac{\partial h}{\partial x_1}(0,a) &\leq \frac{1}{2} \sup_t \left| \frac{\partial^2 h}{\partial x_1^2}(t,a) \right| + \frac{\partial h}{\partial x_1}(\frac{1}{2},a) \\ &\leq \frac{1}{2}(4\bar{d} + A) + 2\bar{d} \\ &= 4\bar{d} + \frac{A}{2}\,. \end{aligned}$$

Since $\frac{\partial h}{\partial x_1}(0,a) \geq \frac{\partial h}{\partial x_1}(x_1,a),\ \forall\, x_1 \in [0,1]$, this proves the Lemma's first derivative bound for $i = 1$. The $i = 2$ case is identical. □

Lemma 2: *Let* $\epsilon_0 > 0$ *and* $A = \infty$*. Then for every* $h \in \Psi_0^{[A]}$ *and* $i = 1, 2$*,*

$$\frac{\partial h}{\partial x_i} \le \frac{\bar{d}}{\epsilon_0}$$

at every point of $[\epsilon_0, 1] \times [\epsilon_0, 1]$*, and* $\Psi_0^{[A]}$ *is equicontinuous on* $[\epsilon_0, 1] \times [\epsilon_0, 1]$*. If* $A < \infty$*, then* $\Psi_0^{[A]}$ *is equicontinuous on* $[0, 1] \times [0, 1]$*.*

Proof: Let $(x_1, x_2) \in [\epsilon_0, 1] \times [\epsilon_0, 1]$, $\epsilon_0 > 0$, and $h \in \Psi_0^{[A]}$, $A = \infty$. Then $h(x_1, x_2) \le \bar{d}$ and $h(0, x_2) \ge 0$. By concavity,

$$\frac{\partial h}{\partial x_1}(x_1, x_2) \le \frac{h(x_1, x_2) - h(0, x_2)}{x_1 - 0} \le \frac{\bar{d} - 0}{\epsilon_0 - 0} = \frac{\bar{d}}{\epsilon_0} .$$

Let $\frac{\bar{d}}{\epsilon_0} = M$. Then

$$\frac{\partial h}{\partial x_1}(x_1, x_2) \le M .$$

Similarly

$$\frac{\partial h}{\partial x_2}(x_1, x_2) \le M .$$

Thus, equicontinuity of $\Psi_0^{[A]}$ follows by using the standard arguments. If $A < \infty$, the result can be established similarly using Lemma 1. □

For a function f on a metric space Ω with metric ρ, define the modulus of continuity of f to be the function $\omega(f; \cdot)$ on $(0, \infty)$ given by

$$\omega(f; \delta) = \sup \{ |f(z') - f(z'')| : z', z'' \in \Omega, \ \rho(z', z'') < \delta \} .$$

Lemma 3: *Let* f *be a continuous function on* $[0, 1] \times [0, 1]$*, and let* $\omega(f; \cdot)$ *denote its modulus of continuity with respect to the Euclidean metric. Then the Bernstein polynomials* $B^f_{r,r}$ *of* f *satisfy*

$$\left| f(x_1, x_2) - B^f_{r,r}(x_1, x_2) \right| \le \frac{5}{4} \omega \left(f; \sqrt{\frac{2}{r}} \right)$$

for every $(x_1, x_2) \in [0, 1] \times [0, 1]$*.*

Proof: See Chak (2001). □

Lemma 4: *Let h be a strictly concave function in $\Psi_0^{[D]}$. Let $A = \infty$ if $D = \infty$, and let $A > D$ if $D < \infty$. Then there exists a sequence of Bernstein polynomials $\{q_n\}$ such that each $q_n \in S_n^{[A]}$ and $\{q_n\}$ converges to h uniformly on $[0,1] \times [0,1]$.*

Proof: First assume that $D < \infty$. Then by Theorem 1, there exists an $N_1 \geq 1$ and an $N_2 \geq 1$ such that $B^h_{n_1,n_2} \in S_n^{[A]}$, $\forall\, n_1 \geq N_1$ and $\forall\, n_2 \geq N_2$. So, take $q_n = B^h_{n_1,n_2}$ if $n_1 \geq N_1$ and $n_2 \geq N_2$, and q_n to be an arbitrary function in $S_n^{[A]}$ if $n_1 < N_1$ or $n_2 < N_2$. Then it follows that $\{q_n\}$ converges to h uniformly on $[0,1] \times [0,1]$.

Now assume that $D = \infty$. For $0 < \theta < 1$, define the function h_θ on $[0,1] \times [0,1]$ by

$$h_\theta(x_1, x_2) \;=\; h((1-\theta)x_1 + \theta, (1-\theta)x_2 + \theta)\,.$$

As θ decreases to 0, observe that $(1-\theta)x_i + \theta$ decreases to x_i, and hence h_θ decreases to h. Since these functions are all continuous, we deduce that h_θ converges uniformly to h on $[0,1] \times [0,1]$ [by Theorem 7.13 of Rudin (1976)]. Hence for each integer $N \geq 1$ we can choose a $\theta = \theta(N) \in (0,1)$ such that

$$||h_\theta - h||_\infty \;<\; \frac{1}{2N}\,.$$

Since the function h has finite and continuous third derivatives on $[\theta(N), 1] \times [\theta(N), 1]$, it follows that h_θ has bounded third derivatives on $[0,1] \times [0,1]$ (where this bound can depend on N). That is, $h_\theta \in \Psi_0^{[D']}$ for some finite D'. By the first part of this lemma, there exists an integer $n = n(N)$ and a Bernstein polynomial q_n in $S_n^{[\infty]}$ such that

$$||h_\theta - q_n||_\infty \;<\; \frac{1}{2N}\,.$$

Therefore q_n converges uniformly to h. $\square$

Now we consider functions g of (x_1, x_2, y) of the following form. Define

$$\tilde{\Psi}_0 = \{g : g(x_1, x_2, y) = h_1(x_1, x_2)h_2(x_1, x_2) - 2h_3(x_1, x_2)y \, , \, h_i \in \Psi_0 \}.$$

Also, for a sequence of random variables $\{(X_{1i}, X_{2i}, Y_i)\}_{i \geq 1}$, we define the empirical distribution operators P_n as follows:

$$P_n g = \frac{1}{n} \sum_{i=1}^{n} g(X_{1i}, X_{2i}, Y_i) \, .$$

Lemma 5:

$$\sup_{g \in \tilde{\Psi}_0} |P_n g - E_P(g)| \to 0 \quad a.s. \quad as \quad n \to \infty.$$

Proof: We shall rely on Theorem 24 of Chapter 2 of Pollard (1984). For $\epsilon > 0$ and for a suitable bounded family of functions $\mathcal{F}$ (such as $\tilde{\Psi}_0$), we define $N_1(\epsilon, P_n, \mathcal{F})$ to be the smallest number m such that there exist m functions $u_1, \ldots, u_m$ (not necessarily in $\mathcal{F}$) such that

$$\min_j P_n |g - u_j| \leq \epsilon \quad \text{for every } g \in \mathcal{F}.$$

Pollard's theorem says that if $\log N_1(\epsilon, P_n, \mathcal{F})/n$ converges to 0 in probability for every $\epsilon > 0$, then $\sup_{g \in \mathcal{F}} |P_n g - E_P(g)|$ converges to 0 almost surely.

First, fix $\epsilon > 0$.

For a function $h : [0,1] \times [0,1] \to \mathbf{R}$ and a number $t \in (0,1]$, define the new function $h^{\vee t}$ as follows:

$$h^{\vee t}(x_1, x_2) = h(\max\{x_1, t\}, \max\{x_2, t\}) \, .$$

Now assume that $h \in \Psi_0^{[\infty]}$ and $t \in (0,1]$. By Lemma 2, we deduce that

$$z', z'' \in [0,1] \times [0,1] \;\Rightarrow\; |h^{\vee t}(z') - h^{\vee t}(z'')| \leq \sqrt{2}\frac{\bar{d}}{t}||z' - z''||_2$$

(where $||\cdot||_2$ denotes Euclidean norm). Therefore, the modulus of continuity of $h^{\vee t}$ satisfies

$$\omega(h^{\vee t}; \delta) \leq \frac{\sqrt{2}\bar{d}\delta}{t} \quad \text{for all } \delta > 0 \tag{A.2}$$

for $h \in \Psi_0^{[A]}$.

For $t \in (0,1)$, let $L(t)$ denote the L-shaped region $[0,1] \times [0,1] \setminus [t,1] \times [t,1]$, and let $\sharp_n(L(t))$ denote the number of $i \in \{1, \ldots, n\}$ such that $(X_{1i}, X_{2i}) \in L(t)$. Then we have

$$P_n|h - h^{\vee t}| \;\leq\; 2||h||_\infty \sharp_n(L(t))/n. \tag{A.3}$$

For positive integer r and positive real s, let $\mathcal{B}_{r,s}$ be the following discrete set of Bernstein-type polynomials:

$$\mathcal{B}_{r,s} \;=\; \{\sum_{v_1=0}^{r}\sum_{v_2=0}^{r} c_{v_1,v_2} P_{v_1,r}(x_1) P_{v_2,r}(x_2) \;:\; 0 \leq c_{v_1,v_2} \leq \bar{d}, c_{v_1,v_2}/s \in \mathbf{Z}\}\,.$$

Then $\mathcal{B}_{r,s}$ contains at most $(1 + \bar{d}/s)^{(r+1)^2}$ functions.

Now consider a function $g = h_1 h_2 - 2h_3 y$ in $\tilde{\Psi}_0$, where each h_i is in Ψ_0. Let $g^{\vee t} = h_1^{\vee t} f_h^{\vee t} - 2h_3^{\vee t} y$. As in Equation $(A.3)$, we see that

$$P_n|g - g^{\vee t}| \;\leq\; 8\bar{d}^2\, \sharp_n(L(t))/n. \tag{A.4}$$

Now, $\sharp_n(L(t))/n$ converges to $P\{(X_{1i}, X_{2i}) \in L(t)\}$ with probability 1. So we need to choose $t > 0$ small enough so that

$$8\bar{d}^2 P\{(X_{1i}, X_{2i}) \in L(t)\} \;<\; \frac{\epsilon}{4}\,.$$

For each $i = 1, 2, 3$, we obtain from Lemma 3 and Equation $(A.2)$ that

$$||h_i^{\vee t} - B_{r,r}^{h_i^{\vee t}}||_\infty \le \frac{5\bar{d}}{2t\sqrt{r}}.$$

Next, with t chosen as above, choose r large enough so that $5\bar{d}/(2t\sqrt{r}) < \epsilon/(16\bar{d})$. Finally, choose $s = \frac{\epsilon}{16\bar{d}}$. For each $i = 1, 2, 3$, let

$$u_i(x_1, x_2) = \sum_{v_1=0}^{r} \sum_{v_2=0}^{r} s\lfloor h_i^{\vee t}\left(\frac{v_1}{r}, \frac{v_2}{r}\right)/s\rfloor P_{v_1,r}(x_1) P_{v_2,r}(x_2)$$

Then $u_i \in \mathcal{B}_{r,s}$ and $||u_i - B_{r,r}^{h_i^{\vee t}}||_\infty \le s$, so we have

$$||h_i^{\vee t} - u_i||_\infty \le ||h_i^{\vee t} - B_{r,r}^{h_i^{\vee t}}||_\infty + ||u_i - B_{r,r}^{h_i^{\vee t}}||_\infty \le \frac{\epsilon}{16\bar{d}} + \frac{\epsilon}{16\bar{d}}$$

for $i = 1, 2, 3$. Now define the function

$$q(x_1, x_2, y) = u_1(x_1, x_2) u_2(x_1, x_2) - 2u_3(x_1, x_2) y \,. \tag{A.5}$$

Then we have

$$\begin{aligned} ||g^{\vee t} - q||_\infty &\le ||h_1^{\vee t}(h_2^{\vee t} - u_2)||_\infty + ||(h_1^{\vee t} - u_1)u_2||_\infty + 2||(h_3^{\vee t} - u_1)y||_\infty \\ &\le 4\bar{d} \max_i ||h_i^{\vee t} - u_i||_\infty \\ &\le \frac{\epsilon}{2}\,. \end{aligned}$$

Thus, for each $g \in \tilde{\Psi}_0$, there is a q of the form $(A.5)$ such that

$$\begin{aligned} P_n|g - q| &\le P_n|g - g^{\vee t}| + P_n|g^{\vee t} - q| \\ &\le 8\bar{d}^2 \sharp_n(L(t))/n + \frac{\epsilon}{2}. \end{aligned}$$

If $8\bar{d}^2 \sharp_n(L(t))/n$ is less than $\frac{\epsilon}{2}$, then an upper bound for $N_1(\epsilon, P_n, \tilde{\Psi}_0)$ is the number of all functions of the form $(A.5)$ where each u_i is in $\mathcal{B}_{r,s}$, i.e.

$$N_1(\epsilon, P_n, \tilde{\Psi}_0) \le |\mathcal{B}_{r,s}|^3 \le \left(1 + \frac{\bar{d}}{s}\right)^{3(r+1)^2}.$$

(Observe that the upper bound does not depend on n.) Therefore

$$\Pr\left\{\log N_1(\epsilon, P_n, \tilde{\Psi}_0) \;>\; \log\left(1+\frac{\bar{d}}{s}\right)^{3(r+1)^2}\right\} \;\leq\; \Pr\left\{8\bar{d}^2\,\sharp_n(L(t))/n \geq \frac{\epsilon}{2}\right\},$$

and the right-hand side converges to 0 as $n \to \infty$ (since our choice of t ensured that $\limsup_n 8\bar{d}^2\,\sharp_n(L(t))/n \leq \frac{\epsilon}{4}$ a.s.) This proves the condition for Pollard's theorem, and the Lemma follows. □

Proof of Theorem 2: For any $\xi > 0$ and take q_n from Lemma 4, with probability 1 there exists an N such that $\forall\, n \geq N$

$$\begin{aligned}
& E_P\Big[\hat{h}_n^2(X_1,X_2) - 2\hat{h}_n(X_1,X_2)Y\Big] \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n}\Big[\hat{h}_n^2(X_{1i},X_{2i}) - 2\hat{h}_n(X_{1i},X_{2i})Y_i\Big] + \xi \quad \text{(Lemma 5)} \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n}\Big[q_n^2(X_{1i},X_{2i}) - 2q_n(X_{1i},X_{2i})Y_i\Big] + \xi \quad \text{(definition of } \hat{h}_n) \\
\leq\; & E_P\Big[q_n^2(X_1,X_2) - 2q_n(X_1,X_2)Y\Big] + 2\xi \quad \text{(Lemma 5)} \\
\leq\; & E_P\Big[h^2(X_1,X_2) - 2h(X_1,X_2)Y\Big] + 3\xi \quad \text{(Dominated Convergence)}\,.
\end{aligned}$$

Therefore, for sufficiently large n,

$$E_P\Big[\hat{h}_n^2(X_1,X_2) - 2\hat{h}_n(X_1,X_2)Y\Big] - E_P\Big[h^2(X_1,X_2) - 2h(X_1,X_2)Y\Big] \leq 3\xi\,.$$

Substituting $Y = h(X_1,X_2) + \epsilon$ into the expression above:

$$\begin{aligned}
& E_P\Big[\hat{h}_n^2(X_1,X_2) - 2\hat{h}_n(X_1,X_2)h(X_1,X_2) - 2\hat{h}_n(X_1,X_2)\epsilon \\
& -h^2(X_1,X_2) + 2h^2(X_1,X_2) + 2h(X_1,X_2)\epsilon\Big] \leq 3\xi\,.
\end{aligned}$$

Recalling that $E_P[\epsilon|X_1,X_2] = 0$ we have

$$E_P\Big[h(X_1,X_2)\epsilon\Big]$$

$$
\begin{aligned}
&= E_P\Big[E_P\Big(h(X_1, X_2)\epsilon | X_1, X_2\Big)\Big] \\
&= E_P\Big[h(X_1, X_2) E_P\Big(\epsilon | X_1, X_2\Big)\Big] \\
&= 0\,.
\end{aligned}
$$

Similarly,

$$
E_P\Big[\hat{h}_n(X_1, X_2)\epsilon\Big] = 0\,.
$$

Therefore

$$
\begin{aligned}
& E_P\Big[\hat{h}_n^2(X_1, X_2) - 2\hat{h}_n(X_1, X_2)h(X_1, X_2) + h^2(X_1, X_2)\Big] \\
= \;& E_P\Big[\hat{h}_n(X_1, X_2) - h(X_1, X_2)\Big]^2 \\
\le \;& 3\xi \quad \text{as} \quad n \to \infty\,,
\end{aligned}
$$

and this completes the proof. □

Lemma 6: *Let $\{\psi_n\}$ be an equicontinuous set of functions on $[\epsilon_0, 1] \times [\epsilon_0, 1]$ and ϕ be a continuous function on $[\epsilon_0, 1] \times [\epsilon_0, 1]$, $\epsilon_0 \ge 0$. Also let Θ be a subset of $[\epsilon_0, 1] \times [\epsilon_0, 1]$. If $\{\psi_n\}$ converges to ϕ pointwise on Θ, then $\{\psi_n\}$ converges to ϕ uniformly on $\bar{\Theta}$, where $\bar{\Theta}$ is the closure of Θ.*

Proof: See Chak (2001). □

Proof of Theorem 3: We will first show pointwise convergence of $\{\hat{h}_n\}$ to h. That is

$$
\lim_{n\to\infty} \hat{h}_n(a, b) \;=\; h(a, b)\,, \;\; \forall\,(a, b) \in int(supp(P))\,,
$$

Suppose it is not true. That is, suppose

$$
\lim_{n\to\infty} \hat{h}_n(a, b) \;\ne\; h(a, b)\,,
$$

for some $(a, b) \in int(supp(P))$. Let us assume that

$$
\limsup_{n\to\infty} \hat{h}_n(a, b) \;>\; h(a, b)\,.
$$

[Similar result can be obtained if we assume that

$$\liminf_{n\to\infty} \hat{h}_n(a,b) \;<\; h(a,b)\,.]$$

Then there exists an $\eta > 0$ and a subsequence $\{\hat{h}_{n_k}\} \in S_n^{[A]}$ such that

$$\hat{h}_{n_k}(a,b) \;>\; h(a,b) \;+\; \eta\,.$$

Now choose a $\delta > 0$ such that

$$h(x_1,x_2) - h(a,b) \;<\; \frac{\eta}{2}\,,$$

$$\forall\,(x_1,x_2) \in [a,a+\delta]\times[b,b+\delta] \subset int(supp(P)).$$

Since $\hat{h}_{n_k} \in S_n^{[A]}$, then $\hat{h}_{n_k}$ is a monotone non-decreasing function. Hence for every $(x_1,x_2) \in [a,a+\delta]\times[b,b+\delta]$ we have

$$\hat{h}_{n_k}(x_1,x_2) \;\geq\; \hat{h}_{n_k}(a,b)\,,$$

which implies that

$$\left|\hat{h}_{n_k}(x_1,x_2) - h(x_1,x_2)\right| \;\geq\; \frac{\eta}{2}\,.$$

Therefore

$$\int\limits_{[a,a+\delta]\times[b,b+\delta]} \left|\hat{h}_{n_k}(x_1,x_2) - h(x_1,x_2)\right|^2 dP \;\geq\; \frac{\eta^2}{4} Pr\Big([a,a+\delta]\times[b,b+\delta]\Big)\,.$$

It follows that

$$\int\limits_{supp(P)} \left|\hat{h}_{n_k}(x_1,x_2) - h(x_1,x_2)\right|^2 dP \;\;\not\longrightarrow\;\; 0\,,$$

which contradicts the convergence of $\{\hat{h}_n\}$ to h on $[0,1]\times[0,1]$ in the $L_2(P)$ metric space. Thus with probability 1

$$\lim_{n\to\infty} \hat{h}_n(a,b) \;=\; h(a,b)\,,\;\; \forall\,(a,b) \in int(supp(P))\,.$$

Using Lemma 2 and Lemma 6, $\{\hat{h}_n\}$ converges to h uniformly on the closure of $int(supp(P)) \cap [\epsilon_0, 1] \times [\epsilon_0, 1]$ with probability 1. □

Proof of Theorem 4: Let $(x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1]$. Since $\lim_{n\to\infty} \hat{h}_n = h$, $\forall (x_1, x_2) \in int(Supp(P))$, then by Theorem 25.7 of Rockafellar (1970) and for $i = 1, 2$ we have

$$\lim_{n\to\infty} \frac{\partial \hat{h}_n}{\partial x_i} = \frac{\partial h}{\partial x_i}\,, \quad \forall\, (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1]\,.$$

Thus for $i = 1, 2$

$$\lim_{n\to\infty} \left| \frac{\partial \hat{h}_n}{\partial x_i} - \frac{\partial h}{\partial x_i} \right|^2 = 0\,, \quad \forall\, (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1]\,.$$

Also by Lemma 1 and Lemma 2, taking $H = 4\bar{d} + \frac{A}{2}$ if $A < \infty$ and $H = \frac{\bar{d}}{\epsilon_0}$ if $A = \infty$, we have

$$\left| \frac{\partial \hat{h}_n}{\partial x_i} \right| \leq\, H\,, \quad \forall\, (x_1, x_2) \in [\epsilon_0, 1] \times [\epsilon_0, 1] \text{ and } \forall\, n\,,$$
$$\left| \frac{\partial h}{\partial x_i} \right| \leq\, H\,, \quad \forall\, (x_1, x_2) \in [\epsilon_0, 1] \times [\epsilon_0, 1]\,.$$

So by Lebesgue's dominated convergence theorem, for $i = 1, 2$

$$\lim_{n\to\infty} \int\limits_{[\epsilon_0,1]\times[\epsilon_0,1]} \left| \frac{\partial \hat{h}_n}{\partial x_i} - \frac{\partial h}{\partial x_i} \right|^2 dP \;=\; 0\,.$$ □

Proof of Theorem 5: For $q = 0, 1, 2, 3$ we have

$$\left| \frac{\partial^3 h}{\partial x_1^q\, \partial x_2^{3-q}} \right| \leq D \quad \text{and} \quad \left| \frac{\partial^3 \hat{h}_n}{\partial x_1^q\, \partial x_2^{3-q}} \right| \leq A\,,$$
$$\forall\, n \text{ and } \forall\, (x_1, x_2) \in [0,1] \times [0,1].$$

Now let $q = 3$. We obtain

$$\left| \frac{\partial^3 h}{\partial x_1^3} \right| - D \leq 0 \quad \text{and} \quad \left| \frac{\partial^3 \hat{h}_n}{\partial x_1^3} \right| - A \leq 0.$$

Define

$$\varphi_n(x_1, x_2) = \frac{\partial \hat{h}_n}{\partial x_1} - \frac{1}{2} A x_1^2 ,$$
$$\varphi(x_1, x_2) = \frac{\partial h}{\partial x_1} - \frac{1}{2} A x_1^2 .$$

Then $\{\varphi_n\}$ converges to φ pointwise on $int(supp(P)) \cap [0,1] \times [0,1]$. Note that

$$\frac{\partial^2 \varphi_n}{\partial x_1^2} = \frac{\partial^3 \hat{h}_n}{\partial x_1^3} - A \leq 0 ,$$
$$\frac{\partial^2 \varphi}{\partial x_1^2} = \frac{\partial^3 h}{\partial x_1^3} - A \leq 0 .$$

Hence $\varphi_n(\cdot, x_2)$ and $\varphi(\cdot, x_2)$ are concave in x_1 on $[0,1]$. For a given x_2, let

$$A_{x_2} = \Big\{ x_1 \ : \ (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1] \Big\}.$$

By Theorem 25.7 of Rockafellar (1970) we get

$$\lim_{n \to \infty} \frac{\partial \varphi_n}{\partial x_1}(\cdot, x_2) = \frac{\partial \varphi}{\partial x_1}(\cdot, x_2) , \quad \forall \ x_1 \in A_{x_2} .$$

So

$$\lim_{n \to \infty} \frac{\partial \varphi_n}{\partial x_1} = \frac{\partial \varphi}{\partial x_1} , \quad \forall \ (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1] ,$$

and thus

$$\lim_{n \to \infty} \frac{\partial^2 \hat{h}_n}{\partial x_1^2} = \frac{\partial^2 h}{\partial x_1^2} , \quad \forall \ (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1] .$$

Hence

$$\lim_{n \to \infty} \left| \frac{\partial^2 \hat{h}_n}{\partial x_1^2} - \frac{\partial^2 h}{\partial x_1^2} \right|^2 = 0 , \quad \forall \ (x_1, x_2) \in int(supp(P)) \cap [0,1] \times [0,1] .$$

Since $A < \infty$, then by Lemma 1

$$\left| \frac{\partial^2 \hat{h}_n}{\partial x_1^2} \right| \leq 4\bar{d} + A , \quad \forall \ (x_1, x_2) \in [0,1] \times [0,1] \text{ and } \forall \ n ,$$
$$\left| \frac{\partial^2 h}{\partial x_1^2} \right| \leq 4\bar{d} + A , \quad \forall \ (x_1, x_2) \in [0,1] \times [0,1] .$$

By Lebesgue's dominated convergence theorem

$$\lim_{n\to\infty} \int\limits_{[0,1]\times[0,1]} \left| \frac{\partial^2 \hat{h}_n}{\partial x_1^2} - \frac{\partial^2 h}{\partial x_1^2} \right|^2 dP \;=\; 0\,.$$

Using a similar argument

$$\lim_{n\to\infty} \int\limits_{[0,1]\times[0,1]} \left| \frac{\partial^2 \hat{h}_n}{\partial x_2^2} - \frac{\partial^2 h}{\partial x_2^2} \right|^2 dP \;=\; 0\,.$$

We can also prove that

$$\lim_{n\to\infty} \int\limits_{[0,1]\times[0,1]} \left| \frac{\partial^2 \hat{h}_n}{\partial x_1 \partial x_2} - \frac{\partial^2 h}{\partial x_1 \partial x_2} \right|^2 dP \;=\; 0$$

if we define

$$\varphi_n = \frac{\partial \hat{h}_n}{\partial x_1} - \frac{1}{2} A x_2^2$$

and then take the subsequent partial derivatives with respect to x_2. □

Proof of Theorem 6:

$$\begin{aligned} & \left| \hat{c}_{v_1,v_2} - h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) \right| \\ \leq \; & \left| \hat{c}_{v_1,v_2} - \hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) \right| + \left| \hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) - h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) \right|. \end{aligned}$$

Write

$$\hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) = \sum_{\tilde{v}_1=0}^{n_1} \sum_{\tilde{v}_2=0}^{n_2} \hat{c}_{\tilde{v}_1,\tilde{v}_2} P_{\tilde{v}_1,n_1}\left(\frac{v_1}{n_1}\right) P_{\tilde{v}_2,n_2}\left(\frac{v_2}{n_2}\right).$$

Then

$$\begin{aligned} & \left| \hat{c}_{v_1,v_2} - \hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) \right| \\ \leq \; & \sum_{\tilde{v}_1=0}^{n_1} \sum_{\tilde{v}_2=0}^{n_2} \Big(\left| \hat{c}_{v_1,v_2} - \hat{c}_{\tilde{v}_1,v_2} \right| + \left| \hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2} \right| \Big) P_{\tilde{v}_1,n_1}\left(\frac{v_1}{n_1}\right) P_{\tilde{v}_2,n_2}\left(\frac{v_2}{n_2}\right). \end{aligned}$$

Since $\hat{h}_n$ is monotone non-decreasing and strictly concave, we have

$$\Delta_{v_1} = \hat{c}_{v_1+1,v_2} - \hat{c}_{v_1,v_2} \geq 0\,, \quad v_1 = 0, 1, \cdots, n_1 - 1,$$

and

$$\Delta_0 \geq \Delta_1 \geq \cdots \geq \Delta_{n_1-2} \geq \Delta_{n_1-1} \,.$$

Thus

$$(v_1+1)\Delta_{v_1} \leq \sum_{k_1=0}^{v_1} \Delta_{k_1} = \hat{c}_{v_1+1,v_2} - \hat{c}_{0,v_2} \leq \bar{d}\,,$$

which implies that

$$\Delta_{v_1} \leq \frac{\bar{d}}{v_1+1} \leq \frac{\bar{d}}{v_1}\,, \quad v_1 > 0\,.$$

Now choose $\frac{v_1}{n_1} \geq \epsilon_1 > \epsilon_0$. We obtain

$$\Delta_{v_1} = \hat{c}_{v_1+1,v_2} - \hat{c}_{v_1,v_2} \leq \frac{\bar{d}}{\epsilon_1 n_1}\,.$$

Hence, by also choosing $\frac{\tilde{v}_1}{n_1} > \epsilon_0$, it follows that

$$|\hat{c}_{v_1,v_2} - \hat{c}_{\tilde{v}_1,v_2}| \leq |v_1 - \tilde{v}_1| \frac{\bar{d}}{\epsilon_0 n_1} = \left|\frac{v_1}{n_1} - \frac{\tilde{v}_1}{n_1}\right| \frac{\bar{d}}{\epsilon_0}\,.$$

Using a similar argument we get

$$|\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}| \leq |v_2 - \tilde{v}_2| \frac{\bar{d}}{\epsilon_0 n_2} = \left|\frac{v_2}{n_2} - \frac{\tilde{v}_2}{n_2}\right| \frac{\bar{d}}{\epsilon_0}\,.$$

Given $\epsilon > 0$, there exists a $\delta = \min\{\frac{\epsilon\epsilon_0}{16\bar{d}}, \epsilon_1\text{-}\epsilon_0\}$ such that for $\left|\frac{v_1}{n_1} - \frac{\tilde{v}_1}{n_1}\right| < \delta$ and $\left|\frac{v_2}{n_2} - \frac{\tilde{v}_2}{n_2}\right| < \delta$ we have $|\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{v_1,v_2}| < \frac{\epsilon}{16}$ and $|\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}| < \frac{\epsilon}{16}$. So

$$\begin{aligned}
&\left| \hat{c}_{v_1,v_2} - \hat{h}_n\Big(\frac{v_1}{n_1}, \frac{v_2}{n_2}\Big) \right| \\
\leq\ & \sum_{\left|\frac{v_1}{n_1} - \frac{\tilde{v}_1}{n_1}\right| < \delta} \ \sum_{\left|\frac{v_2}{n_2} - \frac{\tilde{v}_2}{n_2}\right| < \delta} \Big(|\hat{c}_{v_1,v_2} - c_{\tilde{v}_1,v_2}| + |\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}|\Big) P_{\tilde{v}_1,n_1}\Big(\frac{v_1}{n_1}\Big) P_{\tilde{v}_2,n_2}\Big(\frac{v_2}{n_2}\Big) \\
+\ & \sum_{\left|\frac{v_1}{n_1} - \frac{\tilde{v}_1}{n_1}\right| < \delta} \ \sum_{\left|\frac{v_2}{n_2} - \frac{\tilde{v}_2}{n_2}\right| \geq \delta} \Big(|\hat{c}_{v_1,v_2} - \hat{c}_{\tilde{v}_1,v_2}| + |\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}|\Big) P_{\tilde{v}_1,n_1}\Big(\frac{v_1}{n_1}\Big) P_{\tilde{v}_2,n_2}\Big(\frac{v_2}{n_2}\Big) \\
+\ & \sum_{\left|\frac{v_1}{n_1} - \frac{\tilde{v}_1}{n_1}\right| \geq \delta} \ \sum_{\left|\frac{v_2}{n_2} - \frac{\tilde{v}_2}{n_2}\right| < \delta} \Big(|\hat{c}_{v_1,v_2} - \hat{c}_{\tilde{v}_1,v_2}| + |\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}|\Big) P_{\tilde{v}_1,n_1}\Big(\frac{v_1}{n_1}\Big) P_{\tilde{v}_2,n_2}\Big(\frac{v_2}{n_2}\Big)
\end{aligned}$$

$$+ \sum_{\left|\frac{v_1}{n_1}-\frac{\tilde{v}_1}{n_1}\right|\geq\delta} \sum_{\left|\frac{v_2}{n_2}-\frac{\tilde{v}_2}{n_2}\right|\geq\delta} \Big(|\hat{c}_{v_1,v_2} - \hat{c}_{\tilde{v}_1,v_2}| + |\hat{c}_{\tilde{v}_1,v_2} - \hat{c}_{\tilde{v}_1,\tilde{v}_2}|\Big) P_{\tilde{v}_1,n_1}\left(\frac{v_1}{n_1}\right) P_{\tilde{v}_2,n_2}\left(\frac{v_2}{n_2}\right)$$

$$< \frac{4\epsilon}{16} + \frac{\bar{d}}{4n_1\delta^2} + \frac{\bar{d}}{4n_2\delta^2} + \frac{2\bar{d}}{4n_1\delta^2 4n_2\delta^2} .$$

The last expression is obtained using Chebychev's inequality [see 1.1.7 of Lorentz (1953)]. Thus, for sufficiently large n_1, n_2 it follows that

$$\left|\hat{c}_{v_1,v_2} - \hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right)\right| < \frac{8\epsilon}{16} = \frac{\epsilon}{2} ,$$

$$\left|\hat{h}_n\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right) - h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right)\right| < \frac{\epsilon}{2} .$$

Hence

$$\left|\hat{c}_{v_1,v_2} - h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right)\right| < \epsilon .$$

Observe that the above result is independent of the choice of v_1 and v_2. Therefore

$$\max_{v_1,v_2:\frac{v_1}{n_1}\geq\epsilon_1,\frac{v_2}{n_2}\geq\epsilon_1} \left|\hat{c}_{v_1,v_2} - h\left(\frac{v_1}{n_1}, \frac{v_2}{n_2}\right)\right| \longrightarrow 0 \text{ a.s. as } n_1, n_2 \to \infty . \qquad \square$$

References

Barnett, William A. and Jonas, Andrew B. (1983), "The Müntz-Szatz Demand System: An Application of A Globally Well Behaved Series Expansion," *Economic Letters*, 11, 337-342.

Berndt, E. R. and Khaled, M. S. (1979), "Parametric Productivity Measurement and Choice Among Flexible Functional Forms," *Journal of Political Economy*, 87, 1220-1245.

Berndt, E. R. and Wood, D. O. (1975), "Technology, Prices, and the Derived Demand for Energy," *The Review of Economics and Statistics*, 57, 259-268.

Butzer, Paul L. (1953), "On Two-Dimensional Bernstein Polynomials," *Canadian Journal of Mathematics*, 5, 107-113.

Chak, Pok Man (2001), "Approximation and Consistent Estimation of Shape-Restricted Functions and Their Derivatives," Ph.D. Thesis, Department of Economics, York University.

Chalfant, James A. (1987), "A Globally Flexible, Almost Ideal Demand System," *Journal of Business and Economic Statistics*, 5, 233-242.

Chalfant, James A. and Gallant A. Ronald (1985), "Estimating Elasticities with the Fourier Cost Function," *Journal of Econometrics*, 28, 205-222.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1971), "Conjugate Duality and the Transcendental Logarithmic Production Function," *Econometrica*, 39, 255-256.

Diewert, W. E. (1971), "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy*, 79, 481-507.

—– (1974), "Applications of Duality Theory," In: Intriligator, M. D. and Kendrick, D. A. (eds.), *Frontiers of Quantitative Economics*, Vol. II, North-Holland, Amsterdam, 176-199.

Diewert, W. E. and Wales, T. J. (1993), "Linear and Quadratic Spline Models for Consumer Demand Functions," *Canadian Journal Economics*, 26, 77-106.

El Badawi, Ibrahim, Gallant, A. Ronald and Souza, Geraldo (1983), "An Elasticity Can Be Estimated Consistently Without A Priori Knowledge of Functional Form," *Econometrica*, 51, 1731-1751.

Gallant, A. Ronald (1981), "On the Bias in Flexible Functional Forms and An Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211-245.

—– (1982), "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20, 285-323.

—– (1984), "The Fourier Flexible Form," *American Journal of Agricultural Economics*, 66, 204-208.

Gallant, A. Ronald and Souza, Geraldo (1991), "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329-353.

Geman, Stuart (1981), "Sieves for Nonparametric Estimation of Densities and Regressions," Reports in Pattern Analysis, No. 99, Division of Applied Mathematics, Brown University.

Geman, Stuart and Hwang, Chii-Ruey (1982), "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *The Annals of Statistics*, 10, No. 2, 401-414.

Grenander, Ulf (1981), *Abstract Inference*, Wiley, New York.

Guilkey, David K., Knox Lovell, C. A. and Sickles, Robin C. (1983), "A Comparison of the Performance of Three Flexible Functional Forms," *International Economic Review*, 24, 591-616.

Kingsley, E. H. (1951), "Bernstein Polynomials for Functions of Two Veriables of Class $C^{(k)}$," *Proceedings American Mathematical Society*, 64-71.

Lorentz, George G. (1953), *Bernstein Polynomials*, University of Toronto Press, Toronto.

Newey, Whitney K. (1991), "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161-1167.

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

Rudin, Walter (1976), *Principles of Mathematical Analysis*, McGraw-Hill, New York.

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

Table 1 Estimated Cost Function, Derivatives and Substitution Elasticities at Sample Mean

		n*	C	C_1	C_2	C_3	C_{11}	C_{22}	C_{33}	σ_{11}	σ_{12}	σ_{13}	σ_{22}	σ_{23}	σ_{33}
	True Values		2.132	0.409	0.444	1.279	-0.160	-0.181	-0.512	-2.000	-1.000	1.000	-2.000	1.000	-0.667
r^2=1.0	Without Share Equations	9	2.132	0.408	0.444	1.279	-0.163	-0.187	-0.519	-2.078	-0.997	1.009	-2.025	1.021	-0.677
		16	2.132	0.408	0.444	1.279	-0.160	-0.181	-0.512	-2.043	-1.005	1.002	-1.957	1.001	-0.677
		25	2.132	0.408	0.444	1.279	-0.160	-0.181	-0.511	-2.041	-1.003	1.000	-1.995	0.999	-0.677
		36	2.132	0.409	0.444	1.279	-0.155	-0.182	-0.512	-1.982	-1.022	0.988	-1.970	1.011	-0.677
	With Share Equations	9	2.132	0.408	0.444	1.279	-0.156	-0.182	-0.515	-2.001	-1.084	0.998	-1.969	1.013	-0.670
		16	2.132	0.408	0.444	1.279	-0.160	-0.181	-0.513	-2.045	-1.005	1.002	-1.959	1.001	-0.677
		25	2.132	0.408	0.444	1.279	-0.160	-0.181	-0.511	-2.040	-1.004	1.000	-1.956	0.999	-0.677
		36	2.132	0.409	0.444	1.279	-0.155	-0.181	-0.512	-2.040	-1.004	1.000	-1.956	1.000	-0.677
r^2=0.9	Without Share Equations	9	2.138	0.488	0.433	1.217	-1.156	-1.700	-2.405	-4.687E9	20.090	2.152	-57.159	6.799	-4.044
		16	2.140	0.509	0.432	1.198	-2.245	-1.653	-3.341	1.530E9	-7.259	7.971	-155.480	5.982	-6.049
		25	2.140	0.512	0.432	1.195	-2.941	-1.486	-3.644	1.130E9	2.681	10.513	-22475.200	3.361	-7.156
		36	2.140	0.503	0.433	1.205	-3.247	-1.511	-3.918	2.844E8	10.383	11.099	-8960.639	4.713	-7.681
	With Share Equations	9	2.132	0.408	0.444	1.280	-0.158	-0.184	-0.514	-2.027	-1.014	0.988	-1.992	1.013	-0.679
		16	2.132	0.409	0.444	1.279	-0.163	-0.184	-0.517	-2.085	-1.000	1.013	-1.990	1.010	-0.674
		25	2.132	0.409	0.444	1.279	-0.159	-0.180	-0.508	-2.083	-1.990	1.223	-1.950	0.993	-0.671
		36	2.131	0.408	0.444	1.279	-0.154	-0.170	-0.496	-1.968	-1.010	0.978	-1.840	0.961	-0.646
r^2=0.7	Without Share Equations	9	2.142	0.575	0.424	1.143	-1.399	-2.952	-2.646	-1.854E10	1957.289	-5.485	-4.304	8.765	-3.285E9
		16	2.145	0.602	0.438	1.105	-2.461	-3.065	-3.706	7.611E9	289.474	7.595	-1.943E9	4.209	5.762E8
		25	2.145	0.612	0.430	1.102	-3.208	-2.828	-3.964	4.735E9	-57.858	11.019	-8967.907	16.052	-1.351E8
		36	2.140	0.503	0.433	1.205	-3.247	-1.511	-3.918	1.017E9	21.633	9.226	-46186.550	14.830	6.398E7
	With Share Equations	9	2.132	0.408	0.444	1.280	-0.159	-0.168	-0.515	-2.030	-1.008	0.996	-2.021	1.020	-0.671
		16	2.132	0.409	0.444	1.279	-0.168	-0.189	-0.525	-2.144	-0.995	1.029	-2.050	1.027	-0.684
		25	2.132	0.408	0.444	1.279	-0.162	-0.173	-0.507	-2.068	-1.014	1.011	-1.875	0.973	-0.660
		36	2.132	0.408	0.444	1.279	-0.155	-0.161	-0.490	-1.979	-1.025	0.987	-1.749	0.932	-0.638
r^2=0.5	Without Share Equations	9	2.147	0.654	0.425	1.067	-1.677	-4.035	-2.735	-2.885E10	524.285	-13.939	-16358.150	73.938	-1.959E10
		16	2.148	0.667	0.427	1.053	-2.612	-4.151	-3.909	1.266E10	-46.896	7.130	12539.960	37.380	3.237E9
		25	2.149	0.683	0.429	1.037	-3.323	-4.168	-4.347	6.675E9	-43.779	10.426	-86241.920	-6.637	-1.035E9
		36	2.150	0.678	0.434	1.038	-3.526	-3.982	-4.543	1.536E9	40.082	6.285	-43966.210	31.106	-1.141E8
	With Share Equations	9	2.131	0.408	0.443	1.280	-0.163	-0.195	-0.524	-2.085	0.985	1.004	-2.138	1.048	-0.682
		16	2.131	0.408	0.443	1.280	-0.173	-0.194	-0.536	-2.217	-0.995	1.050	-2.127	1.049	-0.697
		25	2.132	0.408	0.444	1.280	-0.166	-0.172	-0.513	-2.122	-1.034	1.034	-1.879	0.977	-0.668
		36	2.132	0.408	0.444	1.280	-0.160	-0.158	-0.491	-2.054	-1.021	1.007	-1.727	0.920	-0.639

Notes: Factor demands: $C_i = \partial C/\partial P_i$; cross partial derivatives: $C_{ij} = \partial^2 C/\partial P_i \partial P_j$; Allen-Uzawa elasticities of substitution: $\sigma_{ij} = CC_{ij}/C_iC_j$;
r^2 = approximate R^2 for each equation; n* = number of estimated parameters; order of Bernstein Polynomial = $\sqrt{n^*} - 1$.