Nonparametric Estimation of Additive Models with Homogeneous Components*

Wolfgang Härdle[†], Woocheol Kim[†], and Gautam Tripathi[‡]

[†]Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin

[‡]Department of Economics, University of Wisconsin-Madison

June 2000

Abstract

The importance of homogeneity as a restriction on functional forms has been well recognized in economic theory. Imposing additive separability is also quite popular since many economics models become easier to interpret and estimate when the explanatory variables are additively separable. In this paper we combine the two restrictions and propose a two step nonparametric procedure for estimating additive models whose unknown component functions may be homogeneous of known degree. In the first step we obtain preliminary estimates of the components by imposing homogeneity on local linear fits. In the second step these pilot estimates are marginally integrated to produce estimators of the additive components which possess optimal rates of convergence. We derive the asymptotic theory of these two step estimators and illustrate their use on data collected from livestock farms in Wisconsin.

1 Introduction

Nonparametric methods play a useful role in exploratory data analysis by producing consistent estimates of models without relying upon any particular parameterization of the underlying functional forms. Such methods are particularly useful in economics, since in most cases economic theory does not reveal the exact functional relationship between variables. What can be deduced from the theory is usually limited to qualitative or shape properties of the underlying functional forms. For example, although we do not know the exact functional form of a Marshallian demand function, economic theory shows that these functions are homogeneous of degree zero in prices and income (see, for instance, Hildenbrand (1994)). While associating particular functional forms with economic models can create a potential source of misspecification, it does seem reasonable to incorporate restrictions imposed by the theory into estimation procedures. Imposing a valid shape restriction on an estimator usually enhances its performance and leads to better inference.

*We thank Stefan Sperlich for providing us with the cleaned up data for the application in section 4, and Xenia Matschke for some useful comments. The authors also acknowledge support by the Deutsche Forschungsgemeinschaft via SFB 373 at Humboldt University, Berlin.

The shape property we study in this paper is homogeneity. Homogeneous functions seem to be pervasive in economic analysis. For instance, apart from the demand function example described earlier, in microeconomic theory the cost minimizing behavior of competitive firms implies linear homogeneity of the cost function in input prices. The homogeneity restriction is strong and from an econometric point of view it delivers useful information for statistical inference. In particular, homogeneity can be exploited to increase the statistical accuracy of estimates. In a recent paper Tripathi and Kim (2000) provide some interesting results on estimating a nonparametric regression model when the regression surface is homogeneous of known degree. Compared with the usual nonparametric estimate of a conditional mean function which is obtained without imposing any prior restrictions, homogeneity of the regression surface permits a reduction in the dimension of the surface and so leads to a faster rate of convergence of the surface estimates. Despite this dimension reduction, Tripathi and Kim's approach suffers from the usual curse of dimensionality when the number of regressors is large. Moreover, the standard nonparametric regression model they study does not allow for the possibility of imposing homogeneity on a subset of the regressors. To overcome these limitations, in this paper we assume additive separability among the homogeneous and non homogeneous components and extend the estimation of homogeneous functions to nonparametric additive models. Additive separability is frequently used to simplify structure and is basic to many economic models. See, for example, Deaton and Muellbauer (1980). In nonparametric regression the assumption of additive separability alleviates the deterioration in the attainable convergence rates in high dimensions — the well known "curse of dimensionality" problem (Stone, 1985).

Our objective is to estimate a model with additively separable components when the functional forms of the components are unknown, although we do know that some of them are homogeneous of known degree. Specifically, we let

$$y_i = f_1(X_{1i},\ldots,X_{di}) + f_2(Z_{1i},\ldots,Z_{si}) + \varepsilon_i, \qquad (1.1)$$

where at least one component function, say f_1, is homogeneous of known degree α_1; i.e.

$$f_1(\lambda x_1,\ldots,\lambda x_d) = \lambda^{\alpha_1} f_1(x_1,\ldots,x_d);\ \forall(\lambda, x) \in R_{++} \times S_1;$$

where S_1 denotes the support of $(X_1,\ldots,X_d)$ and α_1 is known. We follow a two step procedure to estimate (1.1): In the first step we use a local linear approach, see Fan (1992), to impose homogeneity on estimates of f_1 and f_2. These preliminary estimates are consistent, although not rate-optimal. Therefore, in the second step we marginally integrate the preliminary estimates to obtain rate optimal estimators of f_1 and f_2.

As an example of (1.1) consider a case where the observed cost of production for a competitive firm is the sum of the variable and fixed costs. While standard microeconomic theory restricts the variable costs to be linearly homogeneous in input prices, we can assume the fixed cost to be an unknown function of some other covariates. Note that (1.1) allows f_2 to be homogeneous functions of known degree α_2, where α_2 may or may not be the same as α_1. This comes in handy if we wish to estimate the cost function of a firm producing two distinct products using different inputs. Although we do not pursue it in this paper, the approach used for estimating (1.1) can also be extended to handle a model with multiplicatively separable components and dependent data: According to the theory of option pricing, the option price π_t is homogeneous in the price of the underlying asset S_t, and the exercise price K. See Ghysels et al (1998). Under multiplicative separability, a nonparametric model for the option price can be constructed as $\pi_t = f_1(S_t; K) f_2(T - t; X_t) + \varepsilon_t$, where f_1 is linearly homogeneous, $T - t$ indexes "time to expiration," and X_t denotes some other variables such as S_{t-1} or volatility.

The paper is organized as follows. In section 2 we introduce the numeraire approach as a convenient tool for imposing homogeneity in local linear estimation, and describe our two step estimation procedure. Section 3 has the main statistical results. In section 4 we apply our method to estimate an additively separable aggregate level production function for livestock production in Wisconsin. All proofs have been confined to the Appendix.

2 Estimation

2.1 Numeraire Approach and Marginal Integration

For convenience of exposition we begin by assuming that both components in (1.1) are homogeneous. Of course, throughout the paper we maintain the assumption that the degree of homogeneity is known. To impose homogeneity on our estimator we reparameterize (1.1) using a numeraire argument. So let

$$U_i \equiv (X_{1i}/X_{di}; \ldots; X_{(d-1)i}/X_{di}); \quad V_i \equiv (Z_{1i}/Z_{si}; \ldots; Z_{(s-1)i}/Z_{si});$$

$\beta_1(U_i) = f_1(U_i; 1)$, and $\beta_2(V_i) = f_2(V_i; 1)$. From the homogeneity of f_1 it follows that

$$f_1(X_{1i}; \ldots; X_{di}) = X_{di}^{\alpha_1} f_1\left(X_{1i}/X_{di}; \ldots; X_{(d-1)i}/X_{di}; 1\right) = X_{di}^{\alpha_1} \beta_1(U_i):$$

Hence we can rewrite (1.1) as

$$Y_i = X_{di}^{\alpha_1} \beta_1(U_i) + Z_{si}^{\alpha_2} \beta_2(V_i) + \varepsilon_i: \qquad (2.2)$$

Observe that additive models with only one homogeneous component can be treated as a special case of (2.2) by letting $Z_{si} = 1$ and $V_i = Z_i$. Therefore, w.l.o.g we can work with (2.2) as far as obtaining the statistical theory for estimating (1.1) is concerned. More importantly, when $\alpha_1 = \alpha_2 = 1$, (2.2) generalizes the varying-coefficient model of Hastie and Tibshrani (1993). In Chen and Tsay (1993), and Chen (1997), the coefficient functions were assumed to depend on a common covariate. However, the β_i's in (2.2) are allowed to be functions of different covariates. Thus studying (2.2) is of independent interest and allows us to deal with a larger class of varying-coefficient models. Henceforth, we focus on estimating (2.2).

The benefits from reparameterization are obvious. To obtain consistent estimates of f_1, one has only to estimate β_1 via $\hat{f}_1(x) = x_d^{\alpha_1} \hat{\beta}_1(u)$. The resulting estimator $x \mapsto \hat{f}_1(x)$ is homogeneous of degree α_1 by construction. Furthermore, since β_1 has only $(d-1)$ arguments, homogeneity helps in enhancing the optimal rate of convergence to $n^{-2/(4+(d-1))}$ under twice differentiability. However, the reader should note that an optimal-rate estimator for β_1 cannot be obtained by a simple application of local linear regression to (2.2). To get a faster rate we use the "marginal integration" (MI) technique, which has been developed by Newey (1994), Tjøstheim and Auestadt (1994), and Linton and Nielsen (1995). To get some intuition behind MI consider (1.1) but without any homogeneity restriction on the components. Let $\hat{m}(x; z)$ be a pilot nonparametric smoother for the conditional mean function — for instance, this could be the Nadaraya-Watson kernel smoother. An estimate of $f_1(x)$ can be obtained by integrating out z in $\widehat{f}(x;z)$. In other words, we can employ $\widehat{f}_1(x) = \int \widehat{f}(x;z)dP_Z$ or $\widehat{f}_1(x) = \frac{1}{n}\sum_{k=1}^{n}\widehat{f}(x;Z_k)$ to estimate $f_1(x)$ consistently up to some constant, where P_Z is a probability measure w.r.t Z. When f_1 is twice continuously differentiable, $\widehat{f}_1(x)$ converges at the optimal rate $n^{-2/(4+d)}$. See Linton and Nielsen (1995) or Linton and Härdle (1997). When applied to (2.2) this argument also shows that we can consistently estimate $f_1(x)$ at rate $n^{-2/(4+d)}$. However, the resulting estimate cannot be homogeneous. Moreover, the rate $n^{-2/(4+d)}$ is not optimal. It is slower than $n^{-2/(4+(d-1))}$, the attainable rate for (2.2). To obtain an estimator with the better rate we note that the first-step estimate $\widehat{\beta}_1(u)$ is effectively a function of $(u;V_i)$. Hence, if we marginally integrate $\widehat{\beta}_1(u;V_i)$ w.r.t. V_i we will get the rate optimal estimate of β_1. Notice that here we do not require an identifying assumption such as $E(f_1(X)) = 0$ usually imposed in other additive models. We now make the preceding discussion rigorous.

2.2 Two-Step Procedure

Let $\widehat{\beta}_{j0}$ and $\widehat{\beta}_{jk}$ denote estimates of the level β_j and the partial derivative of β_j w.r.t. the kth component, respectively. Also let $w = (u;v)$, and $\widehat{\beta}(w) = (\widehat{\beta}_{10}(w), \widehat{\beta}_{11}(w), \ldots, \widehat{\beta}_{1(d-1)}(w), \widehat{\beta}_{20}(w), \widehat{\beta}_{21}(w), \ldots, \widehat{\beta}_{2(s-1)}(w))$. If $\beta(w) = (\beta_1(u);\beta_2(v))$ is differentiable at w, we can locally approximate β_j by a linear function. In the first step we estimate local linear fits for β_j and its partial derivatives as:

$$\begin{aligned}\widehat{\beta}(w) = \arg\min_b \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\Big\{y_i &- \Big[b_{10} + \sum_{k=1}^{d-1} h_1 b_{1k}\Big(\frac{U_{ki}-u_k}{h_1}\Big)\Big]X_{di}^{\alpha_1} \\ &- \Big[b_{20} + \sum_{k=1}^{s-1} h_2 b_{2k}\Big(\frac{V_{ki}-v_k}{h_2}\Big)\Big]Z_{si}^{\alpha_2}\Big\}^2 .\end{aligned} \qquad (2.3)$$

Here $K_h(w) = K_{h_1}(u)K_{h_2}(v) = \frac{1}{h_1^{d-1}}\frac{1}{h_2^{s-1}}\prod_{l=1}^{d-1}K\left(\frac{u_l}{h_1}\right)\prod_{l=1}^{s-1}K\left(\frac{v_l}{h_2}\right)$, and K is a real valued function on $\mathbb{R}$ with compact support. Minimizing (2.3) w.r.t. the b_1's and b_2's yields

$$\widehat{\beta}(w) = Q^{-1}\left(\mathcal{M}(w)^T\mathcal{W}(w)\mathcal{M}(w)\right)^{-1}\mathcal{M}(w)^T\mathcal{W}(w)y; \qquad (2.4)$$

where $Q = \mathrm{diag}_{d+s}(1;h_1;\ldots;h_1;1;h_2;\ldots;h_2)$, $y = (y_1;\ldots;y_n)^T$, and $\mathcal{W}(w) = \mathrm{diag}_{1\le i\le n}K_h(W_i - w)$. The matrix $\mathcal{M}(w)$ is given by $\mathcal{M}(w) = [\mathcal{M}_1(w)\ \mathcal{M}_2(w)]$, where

$$\mathcal{M}_1(w) = \begin{bmatrix} X_{d1}^{\alpha_1} & X_{d1}^{\alpha_1}(U_{11}-u_1)/h_1 & \cdots & X_{d1}^{\alpha_1}\left(U_{(d-1)1}-u_{(d-1)}\right)/h_1 \\ \vdots & \vdots & \vdots & \vdots \\ X_{dn}^{\alpha_1}; & X_{dn}^{\alpha_1}(U_{1n}-u_1)/h_1 & \cdots & X_{dn}^{\alpha_1}\left(U_{(d-1)n}-u_{(d-1)}\right)/h_1 \end{bmatrix};$$

and

$$\mathcal{M}_2(w) = \begin{bmatrix} Z_{s1}^{\alpha_2} & Z_{s1}^{\alpha_2}(V_{11}-v_1)/h_2 & \cdots & Z_{s1}^{\alpha_2}\left(V_{(s-1)1}-v_{(s-1)}\right)/h_2 \\ \vdots & \vdots & \vdots & \vdots \\ Z_{sn}^{\alpha_2} & Z_{sn}^{\alpha_2}(V_{1n}-v_1)/h_2 & \cdots & Z_{s1}^{\alpha_2}\left(V_{(s-1)n}-v_{(s-1)}\right)/h_2 \end{bmatrix}:$$

Note that when only one of the component functions, say f_1, is homogeneous the expressions given above can be simplified by plugging in $Z_{si} = 1$ and $V_{ki} = Z_{ki}$ for

$k = 1,\ldots,s$. Furthermore, the notation $\widehat{\beta}_{1k}(w)$ emphasizes the fact that the estimated value depends on both u and v, although $\widehat{\beta}_{1k}(w)$ itself is a consistent estimate of $\beta_{1k}(u)$. In particular, this shows that the rate of convergence of $\widehat{\beta}_{10}(w)$ is the same as that of a multivariate smoothing regression with covariates W. In the second step, to achieve the optimal rate of convergence for the first additive component, we marginally integrate the pilot estimates $\widehat{\beta}_{10}(u;V_i)$ over V_i; i.e. we obtain

$$\begin{aligned}\widehat{\beta}^{*}_{10}(u) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\beta}_{10}(u;V_i) \qquad (2.5)\\ &= \frac{1}{n}\sum_{i=1}^{n} e_1^T\left\{\mathcal{W}(u;V_i)^T\mathcal{K}(u;V_i)\mathcal{W}(u;V_i)\right\}^{-1}\left\{\mathcal{W}(u;V_i)^T\mathcal{K}(u;V_i)y\right\};\end{aligned}$$

where e_k is a $(d+s)\times 1$ column vector with the kth element being one and others zero. Similarly, $\widehat{\beta}^{*}_{20}(v) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\beta}_{20}(U_i;v)$, and an estimate of the regression surface can be written as $\widehat{m}^{*}(x;z) = x_d^{\alpha_1}\widehat{\beta}^{*}_1(u) + z_s^{\alpha_2}\widehat{\beta}^{*}_2(v)$.

3 Asymptotic Theory

The results in this section are derived under the following technical conditions:

A1. $\{y_i;X_i;Z_i\}_{i=1}^{n}$ is iid with $E(\varepsilon|X;Z)=0$ and $E\left\{\varepsilon^2|X;Z\right\}=\sigma_0^2(X;Z)<\infty$.

A2. The functions $f_1; f_2; \beta_1; \beta_2; \sigma$ and the densities (marginal and joint) $p_X; p_Z; p_{X;Z}$ are twice continuously differentiable with bounded partial derivatives.

A3. $p_X; p_Z$ and $p_{X;Z}$ are bounded away from zero on their compact supports.

A4. The matrix $E\left\{W^TW|X_d=x_d; Z_s=z_s\right\}$ is of full rank and its inverse is element by element bounded in a neighborhood of $(x_d;z_s)$.

A5. The kernel K is a compactly supported density such that $\int uK(u)du = 0$. Also; $|K(x_1)-K(x_2)|<c|x_1-x_2|$ for all x_1 and x_2 in its support.

A6. $h_i\downarrow 0$ and $nh_i^{d+s-2}\to\infty$ for $i\in\{1;2\}$.

A7. $h_2^{(s-1)}/h_1^2\to\infty$; $nh_1^{(d-1)}h_2^{2(s-1)}/\ln^2 n\to\infty$, $h_1\downarrow 0$, and $nh_1\to\infty$.

Most of the assumptions above are standard in the kernel estimation literature. The additional bandwidth condition in A7 allows us to use approximation Lemmas 4.2 and 4.3 of Yang, Härdle, and Nielsen (1999) in the proofs. Our first result is the asymptotic normality of the local linear fit given in (2.4). It is easy to see that (2.4) can be interpreted as a WLS estimate for a linear model. So let $W_{1i}=U_i$, $W_{2i}=V_i$, $U_i=(U_{1i},\ldots,U_{in})^T$, and $V_i=(V_{1i},\ldots,V_{in})^T$. Observe that we can write $\widehat{\beta}(w)=Q^{-1}S_n^{-1}t_n$, where

$$S_n = \begin{bmatrix} S_{11n}(w) & S'_{21n}(w)\\ S_{21n}(w) & S_{22n}(w)\end{bmatrix}; \quad S_{ij} = \begin{bmatrix} s^{00}_{ijn}(w) & s^{01T}_{ijn}(w)\\ s^{01}_{ijn}(w) & s^{11}_{ijn}(w)\end{bmatrix} \quad i,j\in\{1;2\};$$

and for $k, l \in \{0, 1\}$

$$s_{jkn}^{lm}(w) = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left[\left(\frac{W_{ji}-w_j}{h_j}\right)^{l}\cdot\left(\frac{W_{ki}-w_k}{h_k}\right)^{m}\right]^{T} \times X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)} X_{di}^{\alpha_1(2-k)} Z_{si}^{\alpha_2(k-1)}.$$

Furthermore, $t_n = \begin{bmatrix} t_{1n}(w) \\ t_{2n}(w) \end{bmatrix}$ and $t_{jn} = \begin{bmatrix} t^0_{jn}(w) \\ t^1_{jn}(w) \end{bmatrix}$, where

$$t^l_{jn}(w) = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left(\frac{W_{ji}-w_j}{h_j}\right)^{l} X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)} y_i.$$

Since we are only interested in the level estimates, we can separate them out by premultiplying with $E^T = [e_1, e_{d+1}]^T$, i.e.

$$\hat{b}_0(w) \equiv \left[\hat{b}_{10}(w), \hat{b}_{20}(w)\right]^T = E^T Q^{-1} S_n^{-1} t_n = E^T S_n^{-1} t_n.$$

Under A2 we can use Lemma 1 in the appendix to decompose the estimation errors as

$$\hat{b}_0(w) - \beta_0(w) = BIAS_n + \hat{e}_n + o_p(h_1^2) + o_p(h_2^2); \tag{3.6}$$

where the bias term and the leading stochastic terms are given by

$$BIAS_n = E^T Q^{-1} S_n^{-1} B_n;\; B_n = [B_{1n}, B_{2n}]^T;\; B_{jn} = \left[B^0_{jn}, B^1_{jn}\right]^T;$$

$$B^l_{nj} = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left(\frac{W_{ji}-w_j}{h_j}\right)^{l} X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)} \times \left\{\sum_{k=1}^{2}\frac{h_j^2}{2} X_{di}^{\alpha_1(2-k)} Z_{si}^{\alpha_2(k-1)}\left[\left(\frac{W_{ki}-w_k}{h_j}\right)^{T} D^2\beta_k(w_k)\left(\frac{W_{ki}-w_k}{h_k}\right)\right]\right\};$$

$$\hat{e}_n = E^T Q^{-1} S_n^{-1}(w)\tau_n(w);$$

$$\tau_n = \begin{bmatrix} \tau_{1n}(w) \\ \tau_{2n}(w) \end{bmatrix};\; \tau_{jn}(w) = \begin{bmatrix} \tau^0_{jn}(w) \\ \tau^1_{jn}(w) \end{bmatrix};$$

$$\tau^l_{jn}(w) = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left(\frac{W_{ji}-w_j}{h_j}\right)^{l} X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)} \varepsilon_i.$$

Henceforth, let $\mu_K^2 = \int K(u)u^2\,du$ and $\|K\|_2^2 = \int K^2(r)\,dr$. Applying the result for the convergence of the bias term $BIAS_n$ (see Lemma 2 in appendix), and the asymptotic normality of the stochastic term $\hat{e}_n$ (see Lemma 3 in appendix), we obtain the following result. Let $\sigma_1^2(W, X_d, Z_s) = E(\varepsilon^2 \mid (W, X_d, Z_s))$.

Theorem 1. Under A1–A6, $\sqrt{nh^{d+s-2}}\left[\hat{b}_0(w) - \beta_0(w) - BIAS_\beta\right] \xrightarrow{L} N(0, \Sigma_\beta)$, where

$$BIAS_\beta = \frac{h^2}{2}\mu_K^2 \begin{bmatrix} tr\left(D^2\beta_1(w_1)\right) + \dfrac{E(X_d^{\alpha_1} Z_s^{\alpha_2} \mid W = w)}{E\left(X_d^{2\alpha_1} \mid W = w\right)} tr\left(D^2\beta_2(w_2)\right) \\ tr\left(D^2\beta_2(w_2)\right) + \dfrac{E(X_d^{\alpha_1} Z_s^{\alpha_2} \mid W = w)}{E\left(Z_s^{2\alpha_2} \mid W = w\right)} tr\left(D^2\beta_1(w_1)\right) \end{bmatrix}$$

and

$$\Sigma_{\beta}^{*} = \frac{\|K\|_2^2}{p_W(w)}\begin{bmatrix} \dfrac{E\left(X_d^{2\alpha_1}\sigma_1^2(W,X_d,Z_s)|W=w\right)}{E^2\left(X_d^{2\alpha_1}|W=w\right)} & \dfrac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}\sigma_1^2(W,X_d,Z_s)|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)} \\ \dfrac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}\sigma_1^2(W,X_d,Z_s)|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)} & \dfrac{E\left(Z_s^{2\alpha_2}\sigma_1^2(W,X_d,Z_s)|W=w\right)}{E^2\left(Z_s^{2\alpha_2}|W=w\right)} \end{bmatrix}.$$

Large sample behavior of the estimated regression surface $\hat f(x;z) = x_d^{\alpha_1}\hat\beta_1(u) + z_s^{\alpha_2}\hat\beta_2(v)$ is given by the following corollary.

Corollary 2. Under the conditions of Theorem 1,

$$\sqrt{nh^{d+s-2}}\left[\hat f(x;z) - f(x;z) - BIAS_f\right] \to N(0;\Sigma_f);$$

where $BIAS_f = \frac{h^2}{2}\left[x_d^{\alpha_1}; z_s^{\alpha_2}\right]' BIAS$, and

$\Sigma_f = p_W^{-1}(w)\|K\|_2^2\{x_d^{2\alpha_1}\frac{E\left(X_d^{2\alpha_1}\sigma_1^2(W;X_d;Z_s)|W=w\right)}{E^2\left(X_d^{2\alpha_1}|W=w\right)} + 2x_d^{\alpha_1}z_s^{\alpha_2}\frac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}\sigma_1^2(W;X_d;Z_s)|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)}$
$+ z_s^{2\alpha_2}\frac{E\left(Z_s^{2\alpha_2}\sigma_1^2(W;X_d;Z_s)|W=w\right)}{E^2\left(Z_s^{2\alpha_2}|W=w\right)}\}.$

Note that the convergence rates in Theorem 1 and Corollary 2 are of the order $O_p(\sqrt{nh^{d+s-2}})$, which is to be expected by looking at the dimension of the kernel function employed in the smoothing. As in Fan (1992) the asymptotic bias takes a simple form. Namely, it depends only on the second derivatives of the functional form being estimated. However, the bias of $\hat\beta_{10}(w)$ also depends on $D^2\beta_2(w_2)$. This is a natural extension of the result in Tripathi and Kim (1999) where the authors looked at the simpler model $Y_i = X_{di}^{\alpha_1}\beta_1(U_i) + \varepsilon_i$. If the disturbance term is homogeneous of degree zero in $(X;Z)$, the asymptotic variance simplifies to $\|K\|_2^{2(d+s-2)}\sigma_1^2(W;1;1)/p_W(w)$, which takes the usual asymptotic variance of kernel estimatros.

We now obtain the limiting distribution for the second step estimates in (25). Recalling the decomposition for the pilot estimates in (36), the estimation error $\hat\beta_1^*(u) - \beta_1(u)$ from the marginal integration procedure can be decomposed as

$$\frac{1}{n}\sum_{j=1}^{n}\left[\hat\beta_{10}(u;V_j) - \beta_1(u)\right] = BIAS_n^*(u) + \tilde{e}_n^*(u) + o_p(h_1^2) + o_p(h_2^2);$$

where the bias term $BIAS_n^*(u) = \frac{1}{n}\sum_{j=1}^{n} e_1^T S_n^{-1}(u;V_j)B_n(u;V_j)$, and the stochastic term $\tilde{e}_n^*(u) = \frac{1}{n}\sum_{j=1}^{n} e_1^T S_n^{-1}(u;V_j)\tau_n(u;V_j)$. The following result is proved in the appendix. Let $\sigma_2^2(W;X_d) = E(\varepsilon^2|W;X_d))$.

Theorem 3. Under A1-A5 and A7 it follows that:

(i) $\sqrt{nh_1^{d-1}}\left[\hat\beta_1^*(u) - \beta_1(u) - BIAS_{\beta_1}^*(u)\right] \to N\left(0;\Sigma_{\beta_1}^*\right)$, where

$$BIAS_{\beta_1}^*(u) = \mu_K^2\left[\frac{h_1^2}{2}tr\left(D^2\beta_1(u)\right) + \frac{h_2^2}{2}\int p_V(v)\frac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}|W=u;v\right)}{E\left(X_d^{2\alpha_1}|W=u;v\right)}tr\left(D^2\beta_2(v)\right)dv\right];$$

$$\Sigma_1^* = \|K\|_2^{2} \int \frac{p_V^2(s_2)}{p_W(u;s_2)} \frac{E\left(X_d^{2\alpha-1}\sigma^2(W;X_d)\,|\,W=(u;s_2)\right)}{E^2\left(X_d^{2\alpha-1}\,|\,W=(u;s_2)\right)}\, ds_2\,;$$

(ii) $\sqrt{nh_1^{d-1}}\left[\hat{f}_1^*(x) - f_1(x) - BIAS_{f_1}^*(x)\right] \xrightarrow{d} N\left(0;\Sigma_{f_1}^*\right)$; where $\Sigma_{f_1}^* = x_d^{2\alpha-1}\Sigma_1^*$ and $BIAS_{f_1}^*(x) = x_d^{\alpha-1}BIAS^*(u)$.

Theorem 3 shows that the marginally integrated pilot estimate achieves its optimal rate of convergence. A similar result holds for $\hat{f}_2^*(v)$, but with a different rate $\sqrt{nh_2^{s-1}}$. If we choose a smaller bandwidth in the nuisance direction so that $h_2^2/h_1^2 \to 0$, the bias term in Theorem 3 simplifies to $BIAS^*(u) = \frac{h_1^2}{2}\mu_K^2\, tr\left[D^2\beta_1(u)\right]$. Moreover, if the error term is homoscedastic, the asymptotic variance reduces to $\|K\|_2^{2(d-1)}\sigma_0^2 \int \frac{p_V^2(s_2)}{p_W(u;s_2)}ds_2$. Thus, for this special case, our results coincide with the results obtained from marginal integration for unrestricted additive models. We end this section with a corollary to Theorem 3. This shows the asymptotic normality of our two-step procedure for the special case $Z_s^{\alpha_2} = 1$ and $V_{ki} = Z_{ki}$ for $k = 1,\ldots,s$; i.e. when only the first component is homogeneous.

Corollary 4. Under the conditions of Theorem 3,

(i)

$$\sqrt{nh_1^{d-1}}\left[\hat{f}_1^*(x) - f_1(x) - BIAS_{f_1}^*(x)\right]$$
$$\xrightarrow{d} N\left(0;\ \|K\|_2^{2(d-1)}x_d^{2\alpha-1}\int \frac{p_Z^2(z)}{p_{U,Z}(u;z)} \frac{E\left(X_d^{2\alpha-1}\sigma^2(U;Z;X_d)\,|\,U=u;Z=z\right)}{E^2\left(X_d^{2\alpha-1}\,|\,U=u;Z=z\right)}dz\right);$$

where $BIAS_{f_1}^*(u)$ is used to denote

$$\mu_K^2 x_d^{\alpha-1}\left[\frac{h_1^2}{2}tr\left(D^2\beta_1(u)\right) + \frac{h_2^2}{2}\int p_Z(z)\frac{E\left[X_d^{\alpha-1}\,|\,U=u;Z=z\right]}{E\left(X_d^{2\alpha-1}\,|\,U=u;Z=z\right)}tr\left(D^2\beta_2(z)\right)dz\right];$$

(ii)

$$\sqrt{nh_2^s}\left[\hat{f}_2^*(z) - f_2(z) - BIAS_{f_2}^*(z)\right]$$
$$\xrightarrow{d} N\left(0;\ \|K\|_2^{2s}\int \frac{p_U^2(u)}{p_{U,Z}(u;z)}E\left[\sigma_\varepsilon^2(U;X_d;Z)\,|\,U=u;Z=z\right]du\right);$$

where

$$BIAS_{f_2}^*(u) = \mu_K^2\left\{\frac{h_2^2}{2}tr\left(D^2\beta_2(z)\right) + \frac{h_1^2}{2}\int p_U(u)E\left[X_d^{\alpha-1}\,|\,U=u;Z=z\right]tr\left(D^2\beta_1(u)\right)du\right\}:$$

4 Application

4.1 Setup

In this section we estimate an additively separable aggregate level production function for livestock production in Wisconsin. The data used was collected in 1987 by the Farm Credit Service of Saint Paul, Minnesota, by sampling more than 1000 farms in Wisconsin. However, we only use a subset of the original data after deleting: (i) farms with incomplete records, (ii) farms which were considered to be outliers, (iii) non livestock producing farms, and (iv) farms with at least one factor input set to zero. This leaves us with 250 observations. More details about this data set may be found in Chavas and Aliber (1993), Severance-Lossin (1994), and Severance-Lossin and Sperlich (1997). The output variable is livestock (y), while the inputs are family labor (X_1), miscellaneous inputs (e.g. repairs, rent, custom hiring, supplies, insurance, gas, oil and utilities) (X_2), intermediate (i.e. with useful life of one to ten years) assets (X_3), hired labor (X_4), and animal inputs (e.g. purchased feed, breeding, and veterinary services) (X_5). Since all variables are measured in dollars, what we actually have is cost and revenue data. However, assuming that the law of one price holds throughout Wisconsin, these dollar values are directly proportional to the quantities of the input factors and the output. In particular, under this assumption we can treat $(X_1, \ldots, X_5) = \mathbf{X}$ measured in dollars as legitimate factors of production.

We specify the production function $f(\mathbf{X})$ as being the conditional expectation of $y|\mathbf{X}$. Therefore, the following canonical regression model holds:

$$y = f(\mathbf{X}) + \varepsilon. \qquad (4.7)$$

We will refer to (4.7) as the "unrestricted" model. Before taking a nonparametric approach, let us carry out a purely parametric analysis by assuming that f has the popular Cobb-Douglas form; i.e. $f(\mathbf{X}) = A X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} X_4^{\beta_4} X_5^{\beta_5}$. Estimating the resulting linear model by OLS, we get

$$\widehat{\log y} = \underset{(0.289)}{1.886} + \underset{(0.020)}{0.063} \log X_1 + \underset{(0.025)}{0.289} \log X_2 + \underset{(0.034)}{0.305} \log X_3 + \underset{(0.007)}{0.031} \log X_4 + \underset{(0.023)}{0.277} \log X_5; \quad R^2 = 0.900.$$

At 1% level we failed to reject the hypothesis that $\sum_{i=1}^{5} \beta_i = 1$; i.e. we cannot reject the hypothesis of constant returns to scale under a Cobb-Douglas specification. Although this linear fit seems to be quite good, the results are susceptible to potential misspecification in the functional form for f. Additionally, since some factors of production may be "fixed", it seems reasonable to impose linear homogeneity only on those factors which are "variable."

We now relax the functional form assumption and add some structure to (4.7) by additively separating f into two components: f_1, which operates on the fixed factors of production; and f_2, which uses the variable factors of production. It still remains to decide which components of $\mathbf{X}$ should be treated as fixed (resp. variable). Since family labor supply is quite inelastic, it seems reasonable to think of X_1 as a fixed factor. On the other hand while it may also seem reasonable to treat miscellaneous inputs and intermediate assets as fixed factors at the farm level, recall that we are interested in estimating the aggregate level production function. At the industry level these factors may well be variable. For example, at the industry level, we would expect the larger farms to employ more of the intermediate assets than the smaller farms. A simple, albeit non rigorous, way to check whether factor l_i is fixed or variable is to plot l_i/y against y. If l_i is fixed then ceteris paribus we expect the ratio l_i/y to decline with increasing y. After scaling the data so that the sample variance is one, in Figure 1 we display the estimated Nadaraya-Watson regression curves of l_i/y against y for $i = 1, \dots, 5$. From this figure we can see that l_1/y (corresponding to family labor) is the only ratio that seems to decline with y. Therefore, we treat l_1 as the fixed (and $l_2, \dots, l_5$ as the variable) factors at the aggregate level. So let $\vec{l} = (\vec{x}, \vec{z})$, where $\vec{x} = (l_1)$ is the vector of fixed factors and $\vec{z} = (l_2, l_3, l_4, l_5)$ the vector of variable factors. Imposing additive separability on f in terms of $\vec{x}$ and $\vec{z}$, we can write

$$y = f_1(\vec{x}) + f_2(\vec{z}) + \varepsilon.$$

Finally, as we are treating $\vec{z}$ as the vector of variable inputs, let us also assume that f_2 is homogeneous of degree $r = 1$. Therefore, we have that

$$y = f_1(x_1) + z_4 f_2\left(\frac{z_1}{z_4}, \frac{z_2}{z_4}, \frac{z_3}{z_4}, 1\right) + \varepsilon. \tag{4.8}$$

Henceforth, we will call (4.8) the "restricted" model. The reader should keep in mind that the restrictions of additive separability and linear homogeneity, although seemingly reasonable in our context, are merely modeling assumptions which we take as given. Furthermore, although the specification in (4.8) is weaker than the componentwise additive separability used in Severance-Lossin and Sperlich (1997), it does impose a strong restriction on the marginal product of family labor; namely, that it is independent of other inputs. To ensure that these restrictions allow sensible inference, we have to statistically test whether the data support the specification given in (4.8). However, this testing is beyond the scope of the present paper and is left for future research. For the present we confine ourselves to estimating the production relationship described by (4.8).

4.2 Results

Although we can certainly estimate f_1, f_2, and f, due to the high dimensionality of the arguments of f_2 and f we cannot display their estimates as three dimensional surface plots. However, estimates of f_1 can certainly be displayed. In Figure 3(a) we plot the estimated f_1 versus the scaled x_1 (family labor). The same graph also displays the estimated marginal product of family labor ($df_1(x_1)/dx_1$) against x_1. Another useful feature of a production function that economists find interesting is the elasticity of scale $e(\vec{x}, \vec{z})$. As defined in Varian (1992, Page 16), the elasticity of scale measures the percent increase in output due to an increase in the scale of operations (i.e. a one percent increase in all inputs). Using Euler's theorem for homogeneous functions, it is straightforward to verify that:

$$e(\vec{x}, \vec{z}) = \begin{cases} \frac{\vec{x}' \nabla_{\vec{x}} f(\vec{x}, \vec{z}) + \vec{z}' \nabla_{\vec{z}} f(\vec{x}, \vec{z})}{f(\vec{x}, \vec{z})} & \text{for the unrestricted model,} \\ \frac{\vec{x}' \nabla_{\vec{x}} f_1(\vec{x}, \vec{z}) + r f_2(\vec{z})}{f_1(\vec{x}) + f_2(\vec{z})} & \text{for the restricted model.} \end{cases} \tag{4.9}$$

The presence of r in the above expression is merely for the sake of generality in case f_2 is homogeneous of degree r. For the present, since we are assuming that f_2 is linearly homogeneous, we can set $r = 1$. In Figures 2(a) and 2(b) we plot the estimated scale elasticities against the output y for the restricted and unrestricted models. Figures 2(c), 2(d), and 3(b) display the histograms of the output and family labor. These histograms show where the data is concentrated, which allows us to identify the range over which the obtained nonparametric estimates are valid.

The procedure described in the paper was implemented in GAUSS. The data was first scaled by the sample standard deviation to have unit sample variance. A Gaussian kernel was then used to estimate $f; f_1$, and f_2 at each of the 250 observations. Using these estimates we calculated $e(x_i; z_i)$ for $i = 1; \ldots; 250$. Forgoing a (possibly) complicated data driven approach, the bandwidths used in the estimation procedure were simply fixed at $h_1 = n^{-\frac{1}{4+d}}; h_2 = n^{-\frac{1}{3+s}}$ for the restricted model, and $h_3 = n^{-\frac{1}{4+d+s}}$ for the unrestricted model. The optimal bandwidths which minimize the mean squared errors from estimating $f_1; f_2$ and f are proportional to the values we have chosen. Table 1 provides a summary description of the estimated elasticity of scale[1].

Table 1: Estimated $e(x; z)$ for Wisconsin livestock data

Model	Estimated Scale Elasticity			
	Mean	Median	Mean	Median
	(full sample)		(excluding outliers)	
Restricted	1.067	1.018	1.060	1.016
Unrestricted	0.994	1.011	1.011	1.012

Notice that the average scale elasticity for the unrestricted model is slightly higher than $0.965 = \sum_{i=1}^{5} \hat{\beta}_i$, the value predicted by the Cobb-Douglas form. The median values of the estimated scale elasticity for the restricted and unrestricted models are very close. Moreover, a look at Figures 2(a) and 2(c) shows that, in the relevant range for Y, the point estimates of $e(x; z)$ for the restricted and unrestricted models track one another. Although not a statistical test by any means, this is perhaps a good indication that our additive specification is a valid restriction on (4.7). The estimated $e(x_i; z_i)$'s seem to fluctuate around 1 with the median value being slightly greater than 1. Although it may not be possible to rule out constant returns to scale at the aggregate level, increasing returns, if present, are not very substantial. On the average this seems to indicate that larger farms enjoy economies of scale (because they tend to use more of $I_2; \ldots; I_5$ than smaller farms). This perhaps is not very surprising.

However, a feature which is not as easy to explain is the form of the estimated f_1 and the marginal product of labor. By comparing Figures 3(a) and 3(b) observe that in the relevant range for x_1 the estimated f_1 is relatively flat and then suddenly jumps up. Correspondingly, the marginal product is very close to zero (occasionally taking on negative values) and then increases rapidly. This is hard to explain since the neoclassical theory of production postulates that the production function has positive, but diminishing, marginal product. There may be two possible explanations for this effect: (i) Specification error; i.e. that we may have misspecified the additivity restriction in (4.8). However, since the elasticity estimates for the restricted and unrestricted models are very similar, perhaps this may not be as severe a problem. (ii) A second (more plausible) explanation for this

[1] Columns 4 and 5 list the mean and median values of the estimated elasticities after discarding estimates which are more than four standard deviations away from the mean. In our case this left us with 244 observations.

effect may be the manner in which x_1 (family labor) enters the production process. So recall that in the data set family labor is reported in "dollar" units, and imagine a farm where some members of a family (say the children) are listed as being employed on the farm but in reality do not contribute substantially to the production process. If the price for employing family labor is calculated without taking into account the number of hours worked by each family member, more "wages" may get paid without more "work" being done. Loosely speaking, since family labor is measured in dollars, the "slack" in the number of family members employed is taken up by the price; i.e. in nominal terms the farm seems to be employing more family members which reduces the effect of adding a marginal worker. Clearly, this effect cannot persist. As the number of family members employed increases beyond a certain threshold, it will overwhelm the error caused by using the nominal wages and cause the marginal product to increase. In order to verify whether this is indeed the reason for the behavior of f_1, we would have to examine the manner in which the price for employing family labor was calculated in the data set. This, however, is beyond the scope of the present paper.

5 Conclusion

In this paper we show how to nonparametrically estimate an additive regression model when its components are homogeneous functions of known degree. The suggested procedure is easy to analyze and straightforward to implement. We use it to obtain some interesting results about livestock production in Wisconsin at the aggregate level: While there seem to be increasing returns to scale at the aggregate level, the economies of scale do not appear to be very large. The marginal product of family labor seems to be very close to zero up to a certain threshold and then increases (relatively) sharply. Possible explanations for such behavior are discussed above. Of course, a more careful empirical analysis needs to be performed if we want to claim any policy relevance for these results. ¤

Appendix

Lemma 1. (Decomposition of Estimation Errors) Assume that A2 holds. Then,

$$\hat{\beta}_0(w) - \beta_0(w) = \frac{h^2}{2} BIAS_n + e_n + o_p(h_1^2) + o_p(h_2^2)$$

Proof. First, observe that

$$\hat{\beta}_0(w) - \beta_0(w) = E^T Q^{-1} S_n^{-1}(w) \tau_n(w) + \left[E^T Q^{-1} S_n^{-1}(w) t_n^*(w) - \beta_0(w) \right];$$

where

$$\tau_n = \begin{bmatrix} \tau_{1n}(w) \\ \tau_{2n}(w) \end{bmatrix}; \ \tau_{jn}(w) = \begin{bmatrix} \tau^0_{jn}(w) \\ \tau^1_{jn}(w) \end{bmatrix};$$

$$t_n^* = \begin{bmatrix} t_1^*(w) \\ t_2^*(w) \end{bmatrix}; \ t_{jn}^*(w) = \begin{bmatrix} t^{*0}_{jn}(w) \\ t^{*1}_{jn}(w) \end{bmatrix};$$

$$\tau^l_{jn}(w) = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left(\frac{W_{ji}-w_j}{h_j}\right)^l X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)}\varepsilon_i;$$

$$t^{*l}_{jn}(w) = \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left(\frac{W_{ji}-w_j}{h_j}\right)^l X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)}\left[X_{di}^{\alpha_1}\beta_1(U_i) + Z_{si}^{\alpha_2}\beta_2(V_i)\right]:$$

Using Taylor Expansion together with the twice differentiability in A2,

$$\begin{aligned}\beta_j(W_{ji}) &= \beta_j(w_j) + h_j\left(\frac{W_{ji}-w_j}{h_j}\right)^T D\beta_j(w_j)\\ &\quad + \frac{h_j^2}{2}\left(\frac{W_{ji}-w_j}{h_j}\right)^T D^2\beta_j(w_j)\left(\frac{W_{ji}-w_j}{h_j}\right) + o(h_1^2) + o_p(h_2^2):\end{aligned}$$

If we plug in this in the expression for $t^{*l}_{jn}(w)$, it follows that

$$\begin{aligned}t^*_n &= \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left[X_{di}^{\alpha_1}; X_{di}^{\alpha_1}\left(\frac{W_{1i}-w_1}{h_j}\right)^T; Z_{di}^{\alpha_1}; Z_{si}^{\alpha_2}\left(\frac{W_{2i}-w_2}{h_j}\right)^T\right]^T \times\\ &\quad \left[X_{di}^{\alpha_1}\beta_1(U_i) + Z_{si}^{\alpha_2}\beta_2(V_i)\right]\\ &= S_n Q\beta(w) + B_n + o_p\left(h_1^2\right) + o_p(h_2^2);\end{aligned}$$

where

$$\begin{aligned}B_n &= \frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left[X_{di}^{\alpha_1}; X_{di}^{\alpha_1}\left(\frac{W_{1i}-w_1}{h_1}\right)^T; Z_{di}^{\alpha_1}; Z_{si}^{\alpha_2}\left(\frac{W_{2i}-w_2}{h_2}\right)^T\right]^T \times\\ &\quad \left\{\sum_{j=1}^{2}\frac{h_j^2}{2}X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)}\left[\left(\frac{W_{ji}-w_j}{h_j}\right)^T D^2\beta_j(w_j)\left(\frac{W_{ji}-w_j}{h_j}\right)\right]\right\}:\end{aligned}$$

The last equality comes from

$$\begin{aligned}&\sum_{j=1}^{2} X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)}\left[\beta_j(w_j) + h_j\left(\frac{W_{ji}-w_j}{h_j}\right)^T D\beta_j(w_j)\right]\\ &= \left[X_{di}^{\alpha_1}; X_{di}^{\alpha_1}\left(\frac{W_{1i}-w_1}{h_1}\right)^T; Z_{di}^{\alpha_1}; Z_{si}^{\alpha_2}\left(\frac{W_{2i}-w_2}{h_2}\right)^T\right] Q \times\\ &\quad [\beta_1(w_1); D\beta_1(w_1); \beta_2(w_2); D\beta_2(w_2)]^T;\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} K_h(W_i - w)\left[X_{di}^{\alpha_1}; X_{di}^{\alpha_1}\left(\frac{W_{1i}-w_1}{h_1}\right)^T; Z_{di}^{\alpha_1}; Z_{si}^{\alpha_2}\left(\frac{W_{2i}-w_2}{h_2}\right)^T\right]^T \times\\ &\sum_{j=1}^{2} X_{di}^{\alpha_1(2-j)} Z_{si}^{\alpha_2(j-1)}\left[\beta_j(w_j) + h_j\left(\frac{W_{ji}-w_j}{h_j}\right)^T D\beta_j(w_j)\right]\\ &= S_n Q[\beta_1(w_1); D\beta_1(w_1); \beta_2(w_2); D\beta_2(w_2)]^T = S_n Q\beta(w):\end{aligned}$$

Thus,

$$Q^{-1}S_n^{-1}(z)t_n^*(z) = \beta(w) + Q^{-1}S_n^{-1}(w)B_n + o_p(h_1^2) + o_p(h_2^2);$$

and finally,

$$\begin{aligned}&\hat\beta_0(w) - \beta_0(w)\\ = \;& E^TQ^{-1}S_n^{-1}(w)\epsilon_n(w) + E^T\left[Q^{-1}S_n^{-1}(w)t_n^*(w) - \beta(w)\right]\\ = \;& E^TQ^{-1}S_n^{-1}(w)\epsilon_n(w) + E^TQ^{-1}S_n^{-1}(w)B_n + o_p(h_1^2) + o_p(h_2^2).\end{aligned}$$

■

Lemma 2. (Convergence of S_n; B_n; Bias) Assume that A1–A6 hold. Then,

(i) $S_n(w) \to^p p_W(w)\begin{pmatrix}\mathcal{M}_{1+d}E\left(X_d^{2\alpha_1}|W=w\right) & 0_{(1+d)\times(1+s)}\\ 0_{(1+s)\times(1+d)} & \mathcal{M}_{1+s}E\left(Z_s^{2\alpha_2}|W=w\right)\end{pmatrix}$;

(ii) $B_n \to^p \frac{h^2}{2}p_W(w)\times$

$$\begin{bmatrix} E_{|W=w}\left(X_d^{2\alpha_1}\right)\mathrm{vech}'\left(\mu_K^2 I_d\right) & E_{|W=w}\left(X_d^{\alpha_1}Z_s^{\alpha_2}\right)\left[\mathrm{vech}'\left(\mu_K^2 I_s\right)\right] \\ E_{|W=w}\left(X_d^{2\alpha_1}\right)\left[\int K_1(s_1)s_1\mathrm{vech}'(s_1s_1')ds_1\right] & \\ E_{|W=w}\left(X_d^{\alpha_1}Z_s^{\alpha_2}\right)\left[\mathrm{vech}'\left(\mu_K^2 I_d\right)\right]' & E_{|W=w}\left(Z_s^{2\alpha_2}\right)\mathrm{vech}'\left(\mu_K^2 I_s\right)' \\ & E_{|W=w}\left(Z_s^{2\alpha_2}\right)\left[\int K_2(s_2)s_2\mathrm{vech}'(s_2s_2')ds_2\right] \end{bmatrix}\times$$

$$\begin{pmatrix}\mathrm{vech}\left(D^2\beta_1(w_1)\right)\\ \mathrm{vech}\left(D^2\beta_2(w_2)\right)\end{pmatrix};$$

where

$$\mathcal{M}_{1+d} = \begin{pmatrix}1 & 0_d\\ 0_d & \mu_K^2 I_d\end{pmatrix}.$$

(iii)

$$BIAS_n \to^p \frac{h^2}{2}\mu_K^2\begin{bmatrix}\mathrm{tr}\left(D^2\beta_1(w_1)\right) + E\left(X_d^{\alpha_1}Z_s^{\alpha_2}|W=w\right)/E\left(X_d^{2\alpha_1}|W=w\right)\mathrm{tr}\left(D^2\beta_2(w_2)\right)\\ \mathrm{tr}\left(D^2\beta_2(w_2)\right) + E\left(X_d^{\alpha_1}Z_s^{\alpha_2}|W=w\right)/E\left(Z_s^{2\alpha_2}|W=w\right)\mathrm{tr}\left(D^2\beta_1(w_1)\right)\end{bmatrix}$$

Proof.

(i) Under A1–A5, applying a WLLN for i.i.d observations shows that

$$s_{jk}^{lm}(w) = \frac{1}{n}\sum_{i=1}^{n}\varphi_i - E\{\varphi_i\} = o_p(1);$$

where

$$\varphi_i = K_h(W_i - w)\left(\left(\frac{W_{ji}-w_j}{h}\right)^l\left(\frac{W_{ki}-w_k}{h}\right)^m\right)^T X_{di}^{\alpha_1(2-j)}Z_{si}^{\alpha_2(j-1)}X_{di}^{\alpha_1(2-k)}Z_{si}^{\alpha_2(k-1)}.$$

Calculating $E(\varphi_i)$ is easy, just change variables and use the bounded convergence theorem; i.e.

$$
\begin{aligned}
& E(\varphi_i) \\
= & \int K(s) s_j^l (s_k^m)^T r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} r_1^{\alpha_1(2-k)} r_2^{\alpha_2(k-1)} p_{W,X_d,Z_s}(w_1+hs_1, w_2+hs_2, r_1, r_2)\, ds\, dr \\
= & \left(\int K(s) s_j^l (s_k^m)^T ds\right) p_W(w) E\left(X_d^{\alpha_1(2-j)} Z_s^{\alpha_2(j-1)} X_d^{\alpha_1(2-k)} Z_s^{\alpha_2(k-1)} \mid W = w\right)(1+o(1)).
\end{aligned}
$$

Thus,

$$
S_n(w) \xrightarrow{p} p_W(w) \begin{pmatrix} M_{1+d} E\left(X_d^{2\alpha_1} \mid W = w\right) & 0_{(1+d)\times(1+s)} \\ 0_{(1+s)\times(1+d)} & M_{1+s} E\left(Z_s^{2\alpha_2} \mid W = w\right) \end{pmatrix};
$$

where

$$
M_{1+d} = \begin{pmatrix} 1 & 0_d \\ 0_d & \mu_K^2 I_d \end{pmatrix}.
$$

(ii) Using the similar argument, when $h_1 = h_2 = h$,

$$
\begin{aligned}
B_n \xrightarrow{p} & \frac{h^2}{2} \int K(s) \left[r_1^{\alpha_1}, r_1^{\alpha_1} s_1^T, r_2^{\alpha_2}, r_2^{\alpha_2} s_2^T\right]^T \left\{\sum_{j=1}^{2} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \left[s_j^T D^2 \beta_j(w_j) s_j\right]\right\} \times \\
& p_{W,X_d,Z_s}(w_1+hs_1, w_2+hs_2, r_1, r_2)\, ds\, dr \\
= & \frac{h^2}{2} \int K(s) \left[r_1^{\alpha_1}, r_1^{\alpha_1} s_1^T, r_2^{\alpha_2}, r_2^{\alpha_2} s_2^T\right]^T \left\{\sum_{j=1}^{2} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \left[s_j^T D^2 \beta_j(w_j) s_j\right]\right\} ds \times \\
& p_{W,X_d,Z_s}(w, r_1, r_2)\, dr\, (1+o(1)).
\end{aligned}
$$

By straightforward algebra,

$$
\begin{aligned}
& \int K(s) r_k^{\alpha_k} \left\{\sum_{j=1}^{2} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \left[s_j^T D^2 \beta_j(w_j) s_j\right]\right\} ds \\
= & \; r_k^{\alpha_k} \sum_{j=1}^{2} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \mathrm{vech}^T\left(\mu_K^2 I_j\right) \mathrm{vech}\left(D^2 \beta_j(w_j)\right).
\end{aligned}
$$

Since

$$
\int K(s) r_k^{\alpha_k} s_k^T r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \left[s_j^T D^2 \beta_j(w_j) s_j\right] ds = 0, \text{ for } j \neq k;
$$

it follows that

$$
\begin{aligned}
& \int K(s) r_k^{\alpha_k} s_k^T \left\{\sum_{j=1}^{2} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} \left[s_j^T D^2 \beta_j(w_j) s_j\right]\right\} ds \\
= & \; r_k^{\alpha_k} r_1^{\alpha_1(2-k)} r_2^{\alpha_2(k-1)} \left(\int K_k(s_k) s_k^T \mathrm{vech}^T(s_k s_k^T)\, ds_k\right) \mathrm{vech}\left(D^2 \beta_k(w_k)\right).
\end{aligned}
$$

Therefore, for $k = 1,2$;

$$B_{kn} = \begin{pmatrix} B^0_{kn} \\ B^1_{kn} \end{pmatrix}$$

$$\to_p \frac{h^2}{2} \begin{pmatrix} \sum_{j=1}^2 \psi_{kj}\left[\text{vech}^T\left(\mu_K^2 I_{d_j}\right)\text{vech}\left(D^2\beta_j(w_j)\right)\right] \\ \psi_{kk}\int K(s_k)s_k\text{vech}^T(s_k s_k^T)ds_k \,\text{vech}\left(D^2\beta_k(w_k)\right) \end{pmatrix} + o(1);$$

where

$$\psi_{kj} = \int r_k^{\alpha_k} r_1^{\alpha_1(2-j)} r_2^{\alpha_2(j-1)} p_{W,X_d,Z_s}(w; r_1; r_2)\, dr.$$

That is,

$$\begin{aligned} B^0_{1n} &\to_p \frac{h^2}{2} p_W(w)\Big\{E\left(X_d^{2\alpha_1}|W = w\right)\left[\text{vech}^T\left(\mu_K^2 I_d\right)\text{vech}\left(D^2\beta_1(w_1)\right)\right] \\ &\quad + E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W = w\right)\left[\text{vech}^T\left(\mu_K^2 I_s\right)\text{vech}\left(D^2\beta_2(w_2)\right)\right]\Big\}; \\ B^1_{1n} &\to_p \frac{h^2}{2} p_W(w) E\left(X_d^{2\alpha_2}|W = w\right)\left(\int K_1(s_1)s_1\text{vech}^T(s_1 s_1^T)ds_1\right)\text{vech}\left(D^2\beta_1(w_1)\right); \end{aligned}$$

which proves (ii).

(iii) From (i) and (ii) it is obvious that

$$\begin{aligned} BIAS_n &= E^T Q^{-1} S_n^{-1}(w) B_n \to_p \\ &= \frac{h^2}{2}\begin{bmatrix} \text{vech}^T\left(\mu_K^2 I_d\right) & \frac{E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)}\left[\text{vech}^T\left(\mu_K^2 I_s\right)\right] \\ \frac{E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W=w\right)}{E\left(Z_s^{2\alpha_2}|W=w\right)}\left[\text{vech}^T\left(\mu_K^2 I_d\right)\right] & \text{vech}^T\left(\mu_K^2 I_s\right) \end{bmatrix}\cdot\begin{pmatrix} \text{vech}\left(D^2\beta_1(w_1)\right) \\ \text{vech}\left(D^2\beta_2(w_2)\right) \end{pmatrix} \\ &= \frac{h^2}{2}\mu_K^2\begin{bmatrix} \text{tr}\left(D^2\beta_1(w_1)\right) + E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W = w\right)/E\left(X_d^{2\alpha_1}|W = w\right)\text{tr}\left(D^2\beta_2(w_2)\right) \\ \text{tr}\left(D^2\beta_2(w_2)\right) + E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W = w\right)/E\left(Z_s^{2\alpha_2}|W = w\right)\text{tr}\left(D^2\beta_1(w_1)\right) \end{bmatrix}. \end{aligned}$$

∎

Lemma 3 Assume that A1–A6 hold. Then,

(i) (Asymptotic Normality of $\sqrt{nh^{d+s-2}}\left(\tau^0_{1n}(w), \tau^0_{2n}(w)\right)^T$)

$$\sqrt{nh^{d-1}}\,\tau_n(w) \to_L N\left(0;\, p_W(w)\|K\|_2^2\,\Sigma_\tau\right);$$

where

$$\Sigma_\tau = \begin{bmatrix} E\left(X_d^{2\alpha_1}\sigma_\varepsilon^2(W, X_d, Z_s)|W = w\right) & E\left(X_d^{\alpha_1} Z_s^{\alpha_2}\sigma_\varepsilon^2(W, X_d, Z_s)|W = w\right) \\ E\left(X_d^{\alpha_1} Z_s^{\alpha_2}\sigma_\varepsilon^2(W, X_d, Z_s)|W = w\right) & E\left(Z_s^{2\alpha_2}\sigma_\varepsilon^2(W, X_d, Z_s)|W = w\right) \end{bmatrix}.$$

(ii) (Asymptotic Normality of the main stochastic term $\tilde{e}_n$)

$$Q\sqrt{nh^{d+s-2}}\,\tilde{e}_n = E^T Q^{-1} S_n^{-1}(w)\sqrt{nh^{d+s-2}}\,\tau_n(w) \xrightarrow{d} N\left(0,\, p_W^{-1}(w)\|K\|_2^2\,\Sigma_{\tilde{e}}\right),$$

where

$$\Sigma_{\tilde{e}} = \begin{bmatrix} \dfrac{E\left(X_d^{2\alpha_1}\sigma_\varepsilon^2(W,X_d,Z_s)\,|W=w\right)}{E^2\left(X_d^{2\alpha_1}|W=w\right)} & \dfrac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s)\,|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)} \\ \dfrac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s)\,|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)} & \dfrac{E\left(Z_s^{2\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s)\,|W=w\right)}{E^2\left(Z_s^{2\alpha_2}|W=w\right)} \end{bmatrix}.$$

Proof. When deriving the asymptotic distribution of $E^TQ^{-1}S_n^{-1}(w)\tau_n(w)$ we only consider $\left[\tau_{1n}^0(w), \tau_{2n}^0(w)\right]$ from the result (Lemma 2) concerning the convergence of $E^TQ^{-1}S_n^{-1}(w)$; i.e.,

$$E^TQ^{-1}S_n^{-1}(w)\xrightarrow{p} p_W^{-1}(w)\begin{pmatrix} \left(E\left(X_d^{2\alpha_1}|W=w\right)\right)^{-1} & 0_{1\times d} & 0 & 0_{1\times s} \\ 0 & 0_{1\times d} & \left(E\left(Z_s^{2\alpha_2}|W=w\right)\right)^{-1} & 0_{1\times s} \end{pmatrix}.$$

(i) To calculate the variance of

$$\left[\tau_{1n}^0(w), \tau_{2n}^0(w)\right]' = \frac{1}{n}\sum_{i=1}^n K_h(W_i - w)\left[X_{di}^{\alpha_1}, Z_{si}^{\alpha_2}\right]'\varepsilon_i,$$

we use the iterative law of expectation with the argument of changing variables.

$$\begin{aligned}
&Var\left(\sqrt{nh^{d+s-2}}\left[\tau_{1n}^0(w), \tau_{2n}^0(w)\right]'\right) \\
&= \frac{1}{h^{d+s-2}}E\left(K^2[(W-w)/h]\begin{bmatrix} X_d^{2\alpha_1} & X_d^{\alpha_1}Z_{si}^{\alpha_2} \\ X_d^{\alpha_1}Z_s^{\alpha_2} & Z_{di}^{2\alpha_2}\end{bmatrix}\sigma_\varepsilon^2(W,X_d,Z_s)\right) \\
&= \int K^2(s)\begin{bmatrix} r_1^{2\alpha_1} & r_1^{\alpha_1}r_2^{\alpha_2} \\ r_1^{\alpha_1}r_2^{\alpha_2} & r_2^{2\alpha_2}\end{bmatrix}\sigma_\varepsilon^2(w,r_1,r_2)\,p_{W,X_d,Z_s}(w,r_1,r_2)\,ds\,dr_1\,dr_2\,(1+o(1)) \\
&= p_W(w)\int K^2(s)\,ds\int\begin{bmatrix} r_1^{2\alpha_1} & r_1^{\alpha_1}r_2^{\alpha_2} \\ r_1^{\alpha_1}r_2^{\alpha_2} & r_2^{2\alpha_2}\end{bmatrix}\sigma_\varepsilon^2(w,r_1,r_2)\frac{p_{W,X_d,Z_s}(w,r_1,r_2)}{p_W(w)}\,dr_1\,dr_2\,(1+o(1)) \\
&= p_W(w)\|K\|_2^2\,\Sigma_\tau\,(1+o(1)).
\end{aligned}$$

Now under A1–A5 the Lindeberg CLT holds (because the Lindeberg condition is satisfied when K has bounded support). Thus

$$\begin{aligned}
\sqrt{nh^{d+s-2}}\left[\tau_{1n}^0(w), \tau_{2n}^0(w)\right]' &= \frac{1}{\sqrt{nh^{d+s-2}}}\sum_{i=1}^n K_h(W_i - w)\left[X_{di}^{\alpha_1}, Z_{si}^{\alpha_2}\right]'\varepsilon_i \\
&\xrightarrow{d} N\left(0,\, p_W(w)\|K\|_2^2\,\Sigma_\tau\right).
\end{aligned}$$

(ii) From Lemma 2, $E^T Q^{-1} S_n^{-1}(w)\sqrt{nh^{d+s-2}}\,\epsilon_n(w) \to N\left\{0, p_W^{-1}(w)\|K\|_2^2\,\Sigma_\epsilon\right\}$. ■

Proof of Theorem 1 and Corollary 2. The proof of Theorem 1 is obvious from Lemma 3. Corollary 2 is also a consequence of Lemma 3 upon noting that

$$\sqrt{nh^{d+s-2}}\left[X_d^{\alpha_1}, Z_s^{\alpha_2}\right]\left[\hat\beta_0(w) - \beta_0(w) - BIAS\right] = \sqrt{nh^{d+s-2}}\left[\hat f(x) - f(x) - BIAS_f\right].$$

■

Proof of Theorem 3. If we use the uniform convergence result proved in Lemma 4.2 of Yang, Härdle, and Nielsen (1999), the first part of Lemma 2 holds uniformly in w with an approximation error of

$$O_p\left(\max(h_1, h_2) + \sqrt{\ln n / nh_1^{(d-1)}h_2^{(s-1)}}\right)$$

under assumption A7. Now, applying the same argument again in Lemma 4.3 of Yang, Härdle, and Nielsen (1999), we can show that

$$BIAS_n^*(u) = \frac{1}{n}\sum_{j=1}^n e_1^T S^{-1}(u,V_j)B_n(u,V_j)(1+o_p(1));$$

$$\hat\epsilon_n^*(u) = \frac{1}{n}\sum_{j=1}^n e_1^T S^{-1}(u,V_j)\epsilon_n(u,V_j)(1+o_p(1));$$

where S is given from Lemma 2 above.

(i) Bias term: Since

$$e_1^T S^{-1}(u,V_j) = p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)e_1^T;$$

we only consider the probability limit of

$$\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)e_1^T B_n(u,V_j)$$

as the asymptotic bias, $BIAS^*(u)$. Using Lemma 1, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)e_1^T B_n(u,V_j)\\
=\;& \frac{h_1^2}{2}\frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i-u)X_{di}^{\alpha_1}\left[\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)K_{h_2}(V_i-V_j)\right]\\
&\times X_{di}^{\alpha_1}\left(\frac{U_i-u}{h_1}\right)^T D^2\beta_1(u)\left(\frac{U_i-u}{h_1}\right)\\
&+\frac{h_2^2}{2}\frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i-u)X_{di}^{\alpha_1}Z_{si}^{\alpha_2}\Bigg\{\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)K_{h_2}(V_i-V_j)\\
&\times\left[\left(\frac{V_i-V_j}{h_2}\right)^T D^2\beta_2(V_j)\left(\frac{V_i-V_j}{h_2}\right)\right]\Bigg\}.
\end{aligned}$$

Since

$$\frac{1}{n}\sum_{j=1}^{n} p_W^{-1}(u;V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_j)\right)K_{h_2}(V_i-V_j)$$

$$= p_V(V_i)p_W^{-1}(u;V_i)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_i)\right)(1+o_p(1));$$

the first term of the bias converges in probability,

$$\frac{1}{n}\sum_{i=1}^{n} K_{h_1}(U_i-u)X_{di}^{2\alpha_1}p_V(V_i)p_W^{-1}(u;V_i)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_i)\right)$$
$$\times\left(\frac{U_i-u}{h_1}\right)^T D^2\beta_1(u)\left(\frac{U_i-u}{h_1}\right)$$
$$\to^p \int K_1(s_1)r_1^{2\alpha_1}p_V(s_2)p_W^{-1}(u;s_2)E^{-1}\left(X_d^{2\alpha_1}|W=(u;s_2)\right)\left[s_1^T D^2\beta_1(u)s_1\right]\times$$
$$p_{W,X_d}(u;s_2;r_1)\,ds\,dr_1;$$

which is simplified as

$$\int K_1(s_1)\left[s_1^T D^2\beta_1(u)s_1\right]ds_1\times$$
$$\int r_1^{2\alpha_1}\left(\int p_V(s_2)p_W^{-1}(u;s_2)E^{-1}\left(X_d^{2\alpha_1}|W=(u;s_2)\right)p_{W,X_d}(u;s_2;r_1)\,ds_2\right)dr_1$$
$$= \mathrm{vech}^T\left(\mu_K^2 I_d\right)\mathrm{vech}\left(D^2\beta_1(u)\right)\times$$
$$\int\frac{p_V(s_2)}{E\left(X_d^{2\alpha_1}|W=(u;s_2)\right)}\left(\int r_1^{2\alpha_1}p_{X_d|W}(r_1|u;s_2;)\,dr_1\right)ds_2$$
$$= \mathrm{vech}^T\left(\mu_K^2 I_d\right)\mathrm{vech}\left(D^2\beta_1(u)\right):$$

Similarly,

$$\frac{1}{n}\sum_{j=1}^{n} K_{h_2}(V_i-V_j)p_W^{-1}(u;V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_j)\right)\left(\frac{V_i-V_j}{h_2}\right)^T D^2\beta_2(V_j)\left(\frac{V_i-V_j}{h_2}\right)$$
$$\to p_W^{-1}(u;V_i)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_i)\right)p_V(V_i)\int K_2(s_2)s_2^T D^2\beta_2(V_i)s_2\,p_V(s_2)\,ds_2$$
$$= p_W^{-1}(u;V_i)E^{-1}\left(X_d^{2\alpha_1}|W=(u;V_i)\right)p_V(V_i)\mathrm{vech}^T\left(\mu_K^2 I_s\right)\mathrm{vech}\left(D^2\beta_2(V_i))\right);$$

and the second term converges in probability,

$$\frac{1}{n}\sum_{i=1}^{n} K_{h_1}(U_i-u)\frac{X_{di}^{\alpha_1}Z_{si}^{\alpha_2}p_V(V_i)}{p_W(u;V_i)E\left(X_d^{2\alpha_1}|W=(u;V_i)\right)}\mathrm{vech}^T\left(\mu_K^2 I_s\right)\mathrm{vech}\left(D^2\beta_2(V_i))\right)$$

$$\to^p \int K_1(s_1)\frac{r_1^{\alpha_1}r_2^{\alpha_2}p_V(s_2)}{p_W(u;s_2)E\left(X_d^{2\alpha_1}|W=(u;s_2)\right)}\mathrm{vech}^T\left(\mu_K^2 I_s\right)\mathrm{vech}\left(D^2\beta_2(s_2))\right)\times$$
$$p_{W,X_d,Z_s}(u;s_2;r_1;r_2)\,ds\,dr;$$

which is equivalent to

$$\int \frac{p_V(s_2)}{E\left(X_d^{2\alpha_1}|W=(u,s_2)\right)} \mathrm{vech}^T\left(\mu_K^2 I_s\right)\mathrm{vech}\left(D^2\beta_2(s_2)\right)\left[\int r_1^{\alpha_1} r_2^{\alpha_2} p_{X_d,Z_s|W}(r_1,r_2|u,s_2)\,dr\right] ds_2$$

$$= \int \frac{E\left(X_d^{\alpha_1} Z_s^{\alpha_2}|W=(u,s_2)\right)}{E\left(X_d^{2\alpha_1}|W=(u,s_2)\right)} \mathrm{vech}^T\left(\mu_K^2 I_s\right)\mathrm{vech}\left(D^2\beta_2(s_2)\right) p_V(s_2)\,ds_2 .$$

Therefore,

$$BIAS_n^*(u) \to \mu_K^2\left[\frac{h_1^2}{2}\mathrm{tr}\left(D^2\beta_1(u)\right)+\frac{h_2^2}{2}\int p_V(v)\frac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}|W=u,v\right)}{E\left(X_d^{2\alpha_1}|W=u,v\right)}\mathrm{tr}\left(D^2\beta_2(v)\right)dv\right].$$

(ii) Asymptotic Normality of the Stochastic term:

From

$$e_1^T S^{-1}(u,V_j) = p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)e_1^T;$$

we have only to consider the asymptotic distribution of

$$\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)e_1^T\tau_n(u,V_j).$$

Note that the above can be written as

$$\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)\frac{1}{n}\sum_{i=1}^n K_{1h}(U_i-u)K_{2h}(V_i-V_j)X_{di}^{\alpha_1}\varepsilon_i$$

$$= \frac{1}{n}\sum_{i=1}^n K_{1h}(U_i-u)X_{di}^{\alpha_1}\varepsilon_i\left[\frac{1}{n}\sum_{j=1}^n p_W^{-1}(u,V_j)E^{-1}\left(X_d^{2\alpha_1}|W=(u,V_j)\right)K_{2h}(V_i-V_j)\right]$$

$$= \frac{1}{n}\sum_{i=1}^n K_{1h}(U_i-u)X_{di}^{\alpha_1}\frac{p_V(V_i)}{p_W(u,V_i)E\left(X_d^{2\alpha_1}|W=(u,V_i)\right)}\varepsilon_i(1+o_p(1)).$$

The variance of this term is

$$
\begin{aligned}
&\mathrm{Var}\left(\sqrt{nh^{d-1}}\,\tau_n^*\right)\\
&= \frac{1}{h^{d-1}} E\left(K_1^2[(U-u)/h]\, \frac{X_d^{2\alpha-1} p_V^2(V)\,\sigma_2^2(W;X_d)}{p_W^2(u;V)\,E^2\left(X_d^{2\alpha-1}\mid W=(u;V)\right)}\right)\\
&= \int K^2(s_1)\, \frac{r_1^{2\alpha-1} p_V^2(s_2)\,\sigma_2^2(u;s_2;r_1)}{p_W^2(u;s_2)\,E^2\left(X_d^{2\alpha-1}\mid W=(u;s_2)\right)}\, p_{W;X_d}(u;s_2;r_1)\, ds\, dr_1\,(1+o(1))\\
&\approx \|K\|_2^{2(d-1)} \int \frac{p_V^2(s_2)}{p_W(u;s_2)\,E^2\left(X_d^{2\alpha-1}\mid W=(u;s_2)\right)} \left[\int r_1^{2\alpha-1}\sigma_2^2(u;s_2;r_1)\,\frac{p_{W;X_d}(u;s_2;r_1)}{p_W(u;s_2)}\,dr_1\right] ds_2\\
&= \|K\|_2^{2(d-1)} \int \frac{p_V^2(s_2)}{p_W(u;s_2)}\, \frac{E\left(X_d^{2\alpha-1}\sigma_2^2(W;X_d)\mid W=(u;s_2)\right)}{E^2\left(X_d^{2\alpha-1}\mid W=(u;s_2)\right)}\, ds_2 .
\end{aligned}
$$

Now, the standard argument of Lindeberg Central Limit Theorem gives

$$
\begin{aligned}
\sqrt{nh^{d-1}}\,\tilde{e}_{1n}^*(u) &\approx \frac{1}{\sqrt{nh^{d-1}}}\sum_{i=1}^n K_{1h}(U_i-u)\,X_{di}^{\alpha-1}\,\frac{p_V(V_i)}{p_W(u;V_i)\,E\left(X_d^{2\alpha-1}\mid W=(u;V_i)\right)}\,\varepsilon_i\\
&\xrightarrow{L} N\left(0;\ \|K\|_2^{2(d-1)} \int \frac{p_V^2(s_2)}{p_W(u;s_2)}\, \frac{E\left(X_d^{2\alpha-1}\sigma_2^2(W;X_d)\mid W=(u;s_2)\right)}{E^2\left(X_d^{2\alpha-1}\mid W=(u;s_2)\right)}\, ds_2\right) .
\end{aligned}
$$

■

References

[1] Buja, A., Hastie, T. and Tibshirani R. (1989). Linear smoothers and additive models (with discussion). Ann. Statist. 17, 453-555.

[2] Chavas, J.P., and M. Aliber (1993). An Analysis of economic Efficiency in Agriculture: A Nonparametric Approach, Journal of Agricultural and Resource Economics, 18

[3] Chen, R. (1997). Functional Coefficient Autoregressive Models: Estimation and Tests of Hypotheses, SFB 373, Humboldt University, Discussion Paper

[4] Chen, R., W. Härdle, O.B. Linton, and E. Severance-Lossin (1996). Estimation in additive nonparametric regression. Proceedings of the COMPSTAT conference Semmering. Eds. W. Härdle and M. Schimek. Physika Verlag, Heidelberg.

[5] Chen, R. and R. Tsay (1993). Functional-Coefficient Autoregressive Models, Journal of the American Statistical Association 88: 298-308

[6] Deaton, A. and J. Muellbauer (1980). Economics and Consumer Behavior. Cambridge University Press, Cambridge.

[7] Fan, J. (1992). Design-adaptive nonparametric regression. J. Am. Statist Soc. 82, 998-1004.

[8] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. Annals of Statistics. 21 196-216.

[9] Fan, J. and I. Gijbels (1992). Variable Bandwidth and Local Linear Regression Smoothers. Annals of Statistics 20, 2008-2036.

[10] Ghysels, E., E. Renault, O. Torres, and, V. Patilea (1998) Non-parametic Methods and Option Pricing, in Statistics in Finance edited by D. Hand and S. Jacka, Arnold, Great Britain

[11] Härdle, W. (1990). Applied Nonparametric Regression, Econometric Monograph Series 19. Cambridge University Press.

[12] Hastie, T. and Tibshirani, R. (1990). Generalized Additive Models. Chapman and Hall, London.

[13] Hastie, T., and R. Tibshirani (1993) Varying-Coefficient Models, J. R. Statist. Soc. B, 55, No. 4, pp. 757-796

[14] Hildenbrand, W. (1994). Market Demand: Theory and Empirical Evidence, Princeton University Press.

[15] Kim, W. and O. Linton, and N. Hengartner (2000). A Computationally Efficient Oracle Estimator of Additive Nonparametric Regression with Bootstrap Confidence Intervals. Forthcoming in Journal of Computational and Graphical Statistics.

[16] Linton, O. and W. Härdle (1996). Estimating additive regression models with known links. Biometrika 83, 529-540.

[17] Linton, O. and J. Nielsen (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82, 93-100.

[18] Newey, W. K. (1994). Kernel estimation of partial means. Econometric Theory. 10, 233-253.

[19] Severance-Lossin, E. (1994). Nonparametric Hypothesis Testing of Production Assertions Using data with Random Shocks, PhD. Thesis, Department of Agricultural Economics, University of Wisconsin-Madison.

[20] Severance-Lossin, E., and S. Sperlich (1997). Estimation of Derivatives for Additive Separable Models, Discussion paper #30. Quantification and Simulation Ökonomischer Prozesse, Humboldt-Universität zu Belrin.

[21] Stone, C. J. (1985). Additive regression and other nonparametric models. Ann. Statist. 13, 685-705.

[22] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. Ann. Statist. 14, 592-606.

[23] Tjøstheim, D., and Auestad, B. (1994). Nonparametric identification of nonlinear time series: projections. J. Am. Stat. Assoc. 89, 1398-1409.

[24] Tripathi G., and W. Kim (2000). Nonparametric Estimation of Homogeneous Functions, SFB 373, Humboldt University, Discussion Paper.

[25] Varian, H.R. (1992). Microeconomic Analysis W. Norton and Company, 3rd edn.

[26] Wand, M., and M. Jones (1995): Kernel Smoothing. Chapman and Hall, London.

[27] Yang, L., W. Härdle, and J. Nielsen (1999). Nonparametric Autoregression with Multiplicative Volatility and Additive Mean. Journal of Time Series Analysis 20(5), 579-604.

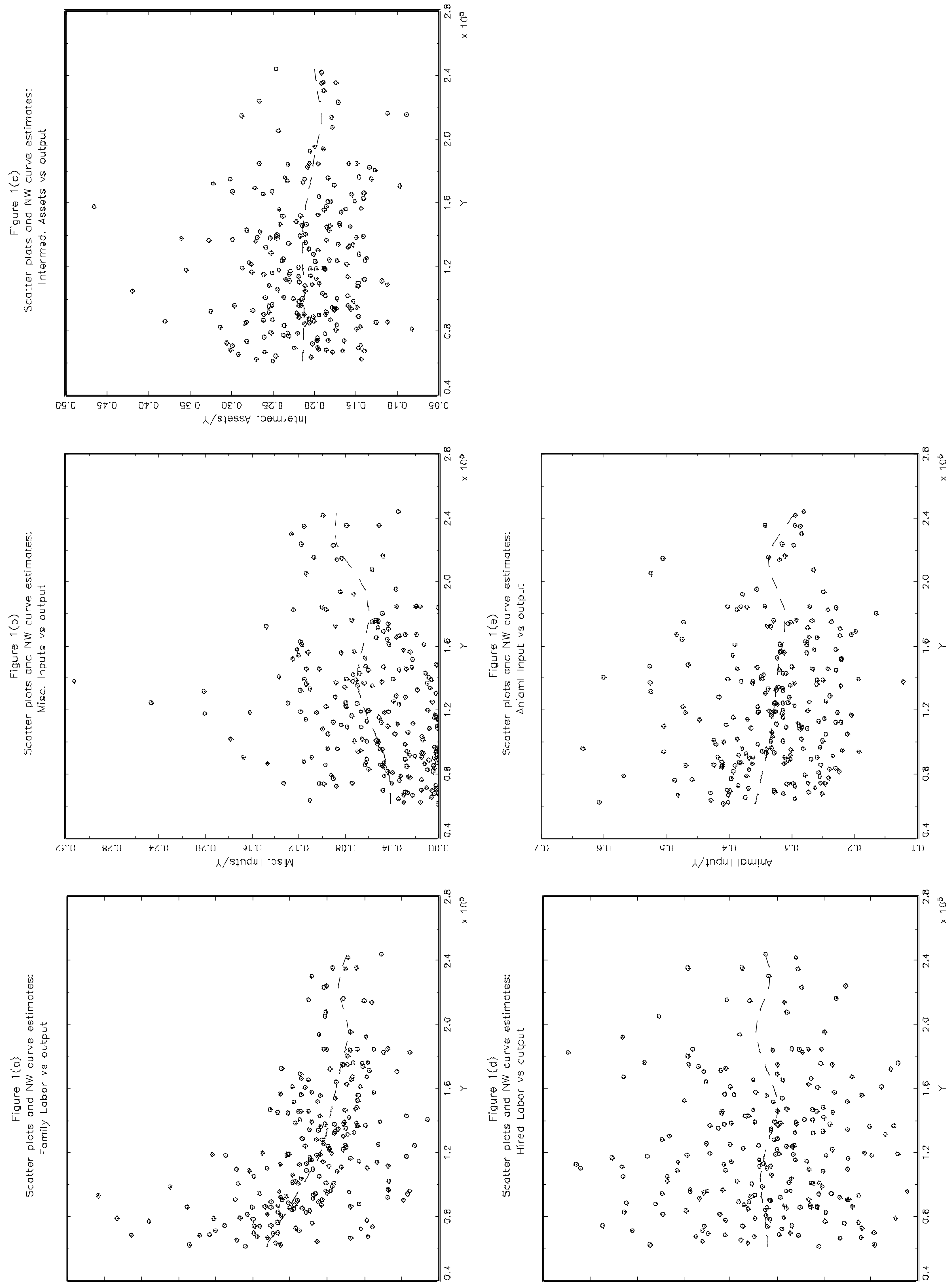


Figure 1(c)
Scatter plots and NW curve estimates:
Intermed. Assets vs output
Intermed. Assets/Y
Figure 1(b)
Scatter plots and NW curve estimates:
Misc. Inputs vs output
Misc. Inputs/Y
Figure 1(e)
Scatter plots and NW curve estimates:
Aniaml Input vs output
Animal Input/Y
Figure 1(a)
Scatter plots and NW curve estimates:
Family Labor vs output
Family Labor/Y
Figure 1(d)
Scatter plots and NW curve estimates:
Hired Labor vs output
Hired Labor/Y
Y
x 10^5

FIGURE 2: (a) Estimated Scale Elasticity from Restricted Model vs Y

(b) Estimated Scale Elasticity from Unrestricted Model vs Y

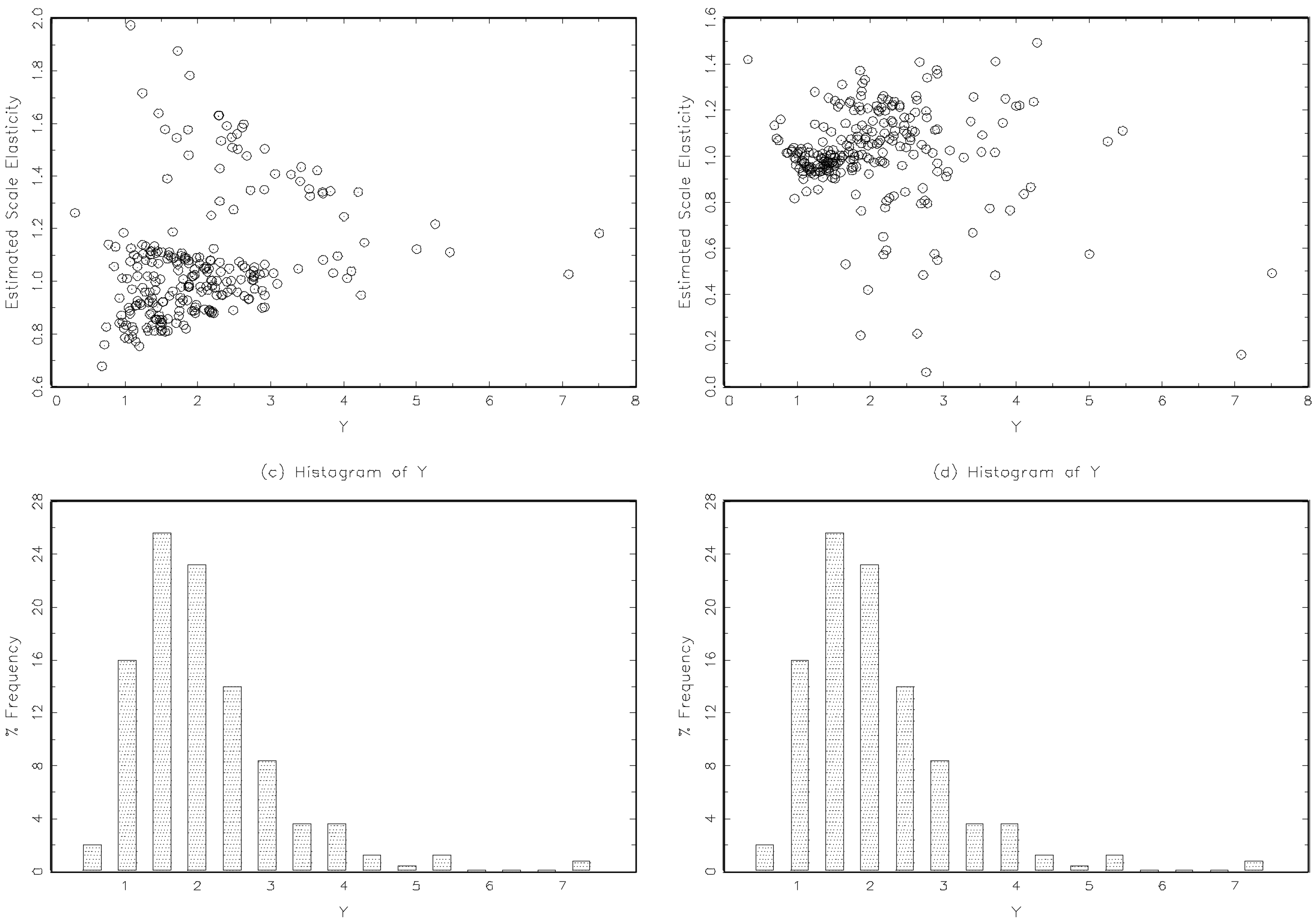
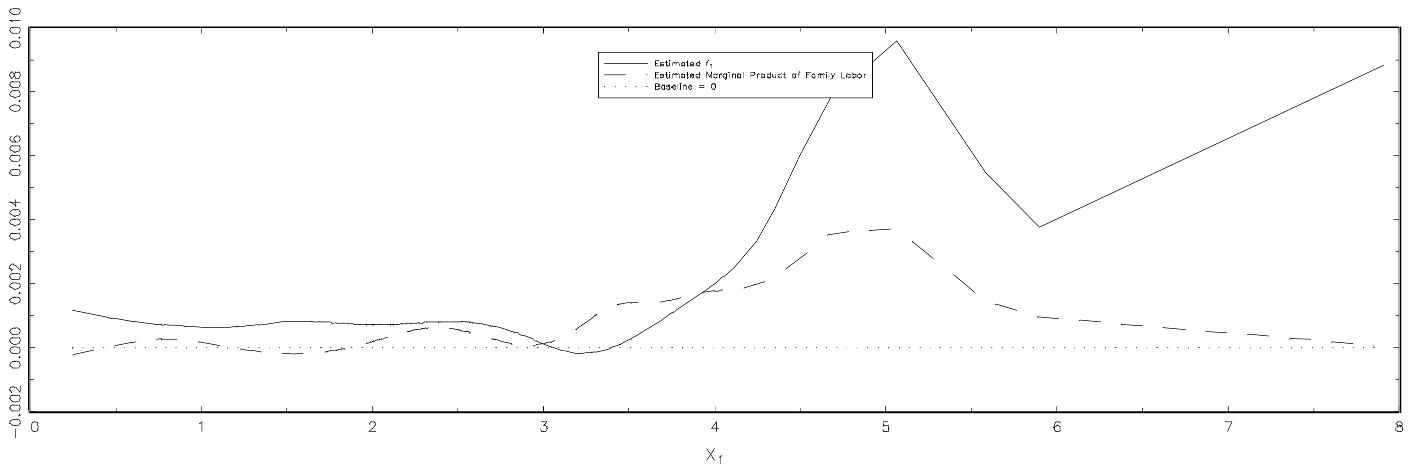


FIGURE 3: (a) Estimated f_1 and the marginal product of family labor



(b) Histogram of Family Labor