ESTIMATION OF DERIVATIVES FOR ADDITIVE SEPARABLE MODELS

E. SEVERANCE-LOSSIN AND S. SPERLICH
HUMBOLDT UNIVERSITÄT SFB 373, SPANDAUER STR. 1, 10178
BERLIN

Abstract. Additive regression models have a long history in nonparametric regression. It is well known that these models can be estimated at the one dimensional rate. Until recently, however, these models have been estimated by a backfitting procedure. Although the procedure converges quickly, its iterative nature makes analyzing its statistical properties difficult. Furthermore it is unclear how to estimate derivatives with this approach since it does not give a closed form for the estimator. Recently, an integration approach has been studied that allows for the derivation of a closed form for the estimator. This paper extends this approach to the simultaneous estimation of both the function and its derivatives by combining the integration procedure with a local polynomial approach. Finally the merits of this procedure with respect to the estimation of a production function subject to separability conditions are discussed. The procedure is applied to livestock production data from Wisconsin. It is shown that there is some evidence of increasing return to scale for larger farms. [1]

Keywords: Derivative Estimation, Nonparametric Regression, Additive Models, Production Function.

1. Introduction.

An additive nonparametric regression model has the form

$$m(x) = E(Y \mid X = x) = c + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha) \tag{1}$$

where Y is a scalar dependent variable, $X = (X_1, \cdots, X_d)$ is a vector of explanatory variables, c is a constant and $\{f_\alpha(\cdot)\}_{\alpha=1}^d$ is a set of unknown functions satisfying $E_{X_\alpha} f_\alpha(X_\alpha) = 0$, and $x = (x_1, \cdots, x_d)$. Models of this form naturally generalize the linear regression models and allow for independent interpretation of the effect of one variable on the mean function m. The linear model, however, assumes $\frac{\partial m}{\partial x_\alpha}$ is constant and so all higher order derivatives vanish. The model (1) allows for arbitrary derivatives in each variable. Models of this form are also interesting from a theoretical point of view since they combine flexible nonparametric modeling of many variables with statistical

[1] This paper is a complete revision of Discussion Paper 60 from (1995). The research was supported by Deutsche Forschungsgemeinschaft, SFB 373.

precision that is typical for just one explanatory variable. This paper is concerned with estimation of the functions $f_\alpha(\cdot)$ and $m(\cdot)$ and their derivatives.

In the statistical literature the additive regression model has been introduced in the early eighties and it has led to the development of a variety of theoretical and practical results. Buja, Hastie and Tibshirani (1989) and Hastie and Tibshirani (1990) give a good overview and analyze estimation techniques based on backfitting. The backfitting idea is to project the data onto the space of functions which are additive. This projection is done via least squares, where the least squares problem is solved with the Gauss-Seidel algorithm. Stone (1985, 1986) proves that model (1) can be estimated with a one-dimensional rate of convergence typical for estimating a single function f_α of one regressor only. Linton and Nielsen (1995) propose a method based on marginal integration of the mean function m for estimating f_α. Their analysis is restricted to the case of dimension $d = 2$. Chen, Härdle, Linton and Severance-Lossin (1995) extended this result to arbitrary d.

The present paper extends these earlier results in the following ways. First, a direct estimator based on the marginal integration idea of Linton and Nielsen is proposed not only for the function, but also for estimating its derivatives. Although there is a growing amount of literature concerned with the estimation of the model (1) and its generalizations, little attention has been given to the estimation of the derivatives of the $f_\alpha(\cdot)$'s. Second, by using a local polynomial approach the asymptotic bias of the estimator is independent of $\varphi(x)$, the density of X_i at x, and so is independent of design. We also give a practical method for selecting bandwidths.

The integration idea on which the estimator presented here is based comes from the following observation. If $m(x) = E(Y \mid X = x)$ is of the additive form (1), and the joint density of $\left(X_{i1}, \cdots, X_{i(\alpha-1)}, X_{i(\alpha+1)}, \cdots, X_{id}\right)$ is denoted as $\varphi_{-\alpha}$, then for a fixed $x \in I\!R$,

$$f_\alpha(x_\alpha) + c = \int m(x_1, \cdots, x_\alpha, \cdots, x_d)\varphi_{-\alpha}(x_1, \cdots, x_d) \prod_{\beta \neq \alpha} dx_\beta, \tag{2}$$

provided $E_{X_\beta} f_\beta(X_\beta) = 0$, $\beta = 1, \cdots, d$. In order to estimate the $f_\alpha(x_\alpha)$'s we first estimate the function $m(\cdot)$ with a multidimensional smoother and then integrate out the variables different from X_α. The smoother proposed in this paper is a local polynomial regression of degree p in the direction of interest and degree zero in the other directions. We establish the asymptotic normal distribution of the estimator for f_α explicitly deriving its bias and variance. In establishing this result we shall see that the rate of convergence for estimating the conditional mean function, $m(x)$, is the usual rate for regression smoothing with just one explanatory variable.

The rest of the paper is organized as follows. Section 2 presents a technique for estimating the functions and their derivatives for the additive model (1). A brief discussion of bandwidth selection is given in Section 3. A small simulation study in given in Section 4. The estimator is used to estimate a production function and its associated elasticities in Section 5. Section 6 concludes. Proofs of the asymptotic results for the estimator are given in the Appendix.

2. The Estimator. Let $(X_{i1},\cdots,X_{id},Y_i)$, $i=1,\cdots,n$ be an i.i.d. random sample related by

$$Y_i = c + \sum_{\alpha=1}^{d} f_\alpha(X_{i\alpha}) + \epsilon_i,$$

where ϵ_i has mean 0 with variance $\sigma^2(X_i)$ and is independent of all $X_{i\alpha}$'s. The functions $f_\alpha(\cdot)'s$ and the constant c are identified by the assumption $\int f_\alpha(t)\varphi_\alpha(t)dt = 0$, where $\varphi_\alpha(\cdot)$ is the marginal density of $X_{i\alpha}$. Let $K(\cdot)$ and $L(\cdot)$ be kernel functions with $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$ and $L_g(\cdot)$ defined similarly. For any kernel, $K(\cdot)$, define $\mu_q(K) = \int u^q K(u)\,du$ and $\|K\|_2^2 = \int K^2(u)\,du$. Define $\overline{X}_i$ to be the i^{th} observation with the α^{th} coordinate removed. Now consider the estimator of $f_\alpha^{(\nu)}(x_\alpha)$, the ν^{th} derivative of $f_\alpha(x_\alpha)$, given by

$$\widehat{f}_\alpha^{(\nu)}(x_\alpha) = (\nu!)\,\frac{1}{n}\sum_{l=1}^{n} E_\nu'\left(Z'W_lZ\right)^{-1} Z'W_lY,$$

where $Z_{ik} = \left((X_{i\alpha}-x_\alpha)^k\right)$, $Y=(Y_i)$, $i=1,\cdots,n$, $k=0,\cdots,p$, $E_\nu \in I\!R^{p+1}$ is a vector of zeros with the $\nu+1^{th}$ element equal to 1, and $W_l = diag\Big(\frac{1}{n}K_h(X_{i\alpha}-x_\alpha) \times L_g\left(\overline{X}_i-\overline{X}_l\right)\Big)_{i=1}^{n}$, compare Chen, Härdle, Linton, Severance-Lossin (1995) and Fan, Gasser, Gijbels, Brockmann and Engel (1993).

In order to derive the limiting distribution of this estimator we will make use of the idea of equivalent kernels as in the second paper of the above mentioned. We will show that the above estimator is asymptotically equivalent to a kernel estimator with a higher order kernel given by

$$K_\nu^*(u) = \sum_{t=0}^{p} s_{\nu t} u^t K(u)$$

with $S = (\int u^{t+s} K(u) du)_{0 \le t,s \le p}$ and $S^{-1} = (s_{st})_{0 \le t,s \le p}$. $K^*_{\nu h}$ is defined analogously to K_h. Derivatives are estimated by taking different rows of S^{-1}.

The following theorem gives the asymptotic behavior of the estimates. To simplify the notation we always write the α^{th} component of $\varphi(\cdot)$ first. Since the constant, c, can be estimated at rate $n^{-1/2}$ without loss of generality we may assume $c = 0$.

THEOREM 1. *Under conditions (A1)-(A5) given in the Appendix, $p - \nu$ odd and x in the interior of the support of $\varphi(\cdot)$ the asymptotic bias and variance of the estimator can be expressed as $n^{-\frac{p+1-\nu}{2p+3}} b_\alpha(x_\alpha)$ and $n^{\frac{-2(p+1-\nu)}{2p+3}} v_\alpha(x_\alpha)$, where*

$$b_\alpha(x_\alpha) = \frac{\nu! h_0^{p+1-\nu}}{(p+1)!} \mu_{p+1}(K^*_\nu) \left(f_\alpha^{(p+1)}(x_\alpha) \right)$$

and

$$v_\alpha(x_\alpha) = \frac{\nu!^2}{h_0^{2\nu+1}} \|K^*_\nu\|_2^2 \int \sigma^2(x_\alpha, \overline{x}) \frac{\varphi^2_{-\alpha}(\overline{x})}{\varphi(x_\alpha, \overline{x})} d\overline{x}.$$

furthermore,

$$n^{\frac{p+1-\nu}{2p+3}} \left\{ \widehat{f}_\alpha^{(\nu)}(x_\alpha) - f_\alpha^{(\nu)}(x_\alpha) \right\} \xrightarrow{L} N\left(b_\alpha(x_\alpha), v_\alpha(x_\alpha)\right).$$

The assumption that $p - \nu$ is odd assures that the lowest order term asymptotic bias does not include any terms involving φ or its derivatives. So the estimator is design adaptive in the sense of Fan (1992). Note that under the bandwidth assumption (A2) the rate of convergence is exactly that of the one-dimensional problem, so that the procedure avoids the curse of dimensionality and the bias is exactly the same as for the one dimensional case. Noting that the expression $\int \sigma^2(x_\alpha, \overline{x}) \frac{\varphi^2_{-\alpha}(\overline{x})}{\varphi(x_\alpha, \overline{x})} d\overline{x}$ is the expectation of the variance over the conditional density of x_α we can see that the variance is also what we would expect from the one-dimensional problem.

Unfortunately we can obtain the one dimensional rate given in Theorem 1 only by using higher order kernels for the nuisance directions with order q. Here q is dependent on the dimension d, the degree of the polynomials and the derivative (see Assumtion (A2) in the appendix). As an alternative Masry and Tjøstheim (1993) avoided using higher order kernels and prefered to renounce the optimal rate in their proof. We agree with them that the optimal one dimensional rate of convergence is probably also attainable with the common bandwidth restrictions but not trivial to proof. Further it is our view that at typical samples sizes other bias reduction techniques than using higher order kernels might be more useful. A method which might perform better in practice would be e. g. to use higher order local polynomials in the directions not of interest. In a typical derivative estimation problem one often chooses $p = \nu + 1$ as it is proposed and motivated in Fan, Gasser, Gijbels, Brockmann and Engel (1993).

The asymptotic distribution of the estimate of the entire function is given by the following theorem.

THEOREM 2. *Under the assumptions of Theorem 1,*

$$n^{\frac{p+1}{2p+3}} \left\{ \widehat{m}\left(x\right) - m\left(x\right) \right\} \xrightarrow{L} N\left(b(x), v(x)\right),$$

where $b(x) = \sum_{\alpha=1}^{d} b_\alpha\left(x_\alpha\right)$ *and* $v(x) = \sum_{\alpha=1}^{d} v_\alpha\left(x_\alpha\right)$.

It is worth noting that assumption (A3) is stronger than required to obtain the one dimensional rate of convergence. All that is needed is that is that the $f_\alpha(\cdot)$'s have p^{th} Lipschitz continuous derivatives. In this case, however, the bias can neither be explicitly calculated nor estimated. Since the bandwidth selection procedure described in the next section relies on a estimate of the bias we only state the result under the stronger smoothness assumption.

3. Bandwidth Selection. Choosing a bandwidth in practice is often a difficult problem. In this section we describe a 'plug-in' method for selecting a bandwidth. Our goal is not to find the optimum method for selecting a bandwidth, but rather to provide a method which is reasonable and can be applied easily. We make use of the fact that this estimation procedure allows for the estimation of the derivatives of the regression functions which are needed to determine the constant h_0, and that the expression for h_0 does not contain derivatives of $\varphi(\cdot)$.

The asymptotically optimal bandwidth constant, with respect to the integrated mean squared error (MISE) is given by

$$(3)\qquad h_0 = \left\{ \frac{(2\nu+1)\left\|K_\nu^{*2}\right\|_2^2 \int \sigma^2(x_\alpha,\overline{x}) \frac{\varphi_{-\alpha}^2(\overline{x})\varphi_\alpha(x_\alpha)}{\varphi(x_\alpha,\overline{x})} d\overline{x}dx_\alpha}{2(p+1-\nu)\left\{\frac{1}{(p+1)!}\mu_{p+1}(K_\nu^*)\right\}^2 \int \left\{f_\alpha^{(p+1)}(x_\alpha)\right\}^2 \varphi_\alpha(x_\alpha)dx_\alpha} \right\}^{\frac{1}{2p+3}}.$$

We suggest the following method for estimating the unknown quantities in (3).

The integral in the denominator is just the marginal expectation of $\left\{f_\alpha^{(p+1)}(x_\alpha)\right\}^2$ and can be estimated by

$$\int f_\alpha^{(p+1)}(x_\alpha)\varphi_\alpha(x_\alpha)dx_\alpha \approx \frac{1}{n}\sum_{i=1}^n \left\{\widehat{f}_\alpha^{(p+1)}(X_{i\alpha})\right\}^2.$$

Also,

$$\int \sigma^2(x_\alpha,\overline{x})\frac{\varphi_{-\alpha}^2(\overline{x})\varphi_\alpha(x_\alpha)}{\varphi(x_\alpha,\overline{x})}d\overline{x}dx_\alpha = E\left[\sigma^2(x_\alpha,\overline{x})\frac{\varphi_{-\alpha}^2(\overline{x})\varphi_\alpha(x_\alpha)}{\varphi^2(x_\alpha,\overline{x})}\right] \approx \frac{1}{n}\sum_{i=1}^n \widehat{\epsilon}_i^2\widehat{w}_i,$$

where $\widehat{\epsilon}_i$ is the residual of the regression at X_i and

$$\widehat{w}_i = \left[\frac{\sum_{t=1}^n L_g\left(\overline{X}_i - \overline{X}_t\right)}{\sum_{t=1}^n K_h(X_{i\alpha}-X_{t\alpha})L_g\left(\overline{X}_i-\overline{X}_t\right)}\right]^2 \left[\frac{1}{n}\sum_{t=1}^n K_h(X_{i\alpha}-X_{t\alpha})\right],$$

which is an estimation of the unknown density quantities. Since the estimates of the expectations are $n^{-1/2}$ consistent one should undersmooth the estimates to get a bias of this order.

The bandwidths in the directions not of interest, $g = (g_1,\cdots,g_{(\alpha-1)},g_{(\alpha+1)},\cdots,g_d)$, should be chosen so that the contribution to the bias from these directions is small compared to the direction of interest. Assume $L(u) = \prod_{\beta\neq\alpha} L_\beta(u_\beta)$ then, a careful examination of the proof of Theorem 1 shows that the contribution to the squared bias in the β^{th} direction is

$$b_\beta^2 = g_\beta^{2q}\mu_2^2(L_\beta)\left(E\left[\frac{1}{2}f_\beta''(X_{1\beta}) + \frac{f_\beta'(X_{1\beta})}{\varphi(X_1)}\frac{\partial\varphi(X_1)}{\partial x_\beta}\right]\right)^2.$$

We can again plug-in estimates and take a sample average. Since a good estimate is not needed only a rough idea of the size of the expression, we recommend plugging in a parametric estimate, at least for the quantities involving $\varphi(\cdot)$. For example we could approximate $\varphi(\cdot)$ by a normal distribution. Then choose g_β so that

$$b_\beta^2 << n^{\frac{-2(p+1-\nu)}{2p+3}}\left\{\nu! h_0^{p+1-\nu}\right\}^2\left\{\frac{\mu_{p+1}(K_\nu^*(u))}{(p+1)!}\right\}^2\frac{1}{n}\sum_{i=1}^n\left\{\widehat{f}_\alpha^{(p+1)}(X_{i\alpha})\right\}^2.$$

4. Simulation Results. In this section we do a small simulation study to evaluate the procedure's performance on data of typical sample size. The local polynomial based estimator is compared to the Nadaraya-Watson based estimator presented in Chen, Härdle, Linton and Severance-Lossin (1995). Since the Nadaraya-Watson based estimator has a closed form expression for $\widehat{f}(x)$ one can differentiate this expression to get an estimate of the derivative. We compare the two estimators for both the additive functions and their first derivatives.

We consider two different designs with $n = 200$ observations, $X \in I\!R^4$, and distribution $U[-3,3]^4$ and $N(0,\Sigma)$, with variance 1 and covariance 0.2. The regression model is $m(x) = \sum_{\beta=1}^{4} g_\beta(x_\beta)$, where

$$g_1(x) = 2x \qquad g_2(x) = x^2 - E[x^2]$$
$$g_3(x) = sin(-x) \qquad g_4(x) = e^x - E[e^x].$$

We used the optimal bandwidth minimizing the integrated mean squared error (MISE) of the estimated function on trimmed data.

In table 1 and 2 we give a survey of results of two different sets of designs. We present the averaged mean squared error of the estimators of the additive functions on trimmed data.

Since each estimator requires a different optimal bandwidths and the asymptotically optimal bandwidths are not necessarily the best for any given set of data, we compared the estimates using the optimal bandwidths for each estimator conditional on the data. Finding these bandwidths is computationally time consuming. A complete study comparing the two procedures with the backfitting algorithm is the subject of forthcoming work. We present the results of two typical replications for each design in Tables 1 and 2.

With one exception the local polynomial version of the estimator performed better in terms of average squared error. The local polynomial estimator seemed to have some trouble detecting the final upward sloping portion of the sin curve. Since both estimators performed poorly near the boundaries the average squared error was calculated over a trimmed region of data. To get a better idea of where these estimators performed well and where they performed poorly we graphed the bias and variance of the derivative estimates over a trimmed range of data for one replication. These results are shown in Figures 3-6.

These preliminary results show that the local polynomial version of the estimator provides gains in estimation accuracy. These gains while noticeable for the estimates of the functions are substantial in the case of derivative estimation.

5. Application to Production Function Estimation.

5.1. Parametric vs. Nonparametric Estimation. We consider the nonparametric estimation of a production function subject to strong separability conditions. Separable production functions have a long history dating back to Leontief (1947). These conditions yield many well known economic results (*e.g.* they allow one to aggregate inputs into indices). In addition, based on the results given in the previous

sections of this paper, we are able to estimate a production function of this form at the one dimensional rate. Since we avoid the curse of dimensionality which plagues multi-dimensional nonparametric regression we are able to get reasonable results with sample sizes which are typical in economic applications.

Estimating production technologies subject to separability constraints has been extensively studied and applied in a flexible functional forms (parametric) setting. In this setting separability conditions can be written as constraints on the parameters of the flexible functional form. The parametric estimation is then done subject to these constraints. The use of flexible functional forms has been rationalized by considering the functional forms to be the first terms of a Taylor series expansion of the true underlying technology. Unfortunately, there is evidence which suggests that this view is unreasonable and that least squared estimates of these parameters do not necessarily correspond well to the actual coefficients (Driscoll and Boisvert (1991), Chalfant and Gallant (1985)). There is also evidence that these flexible functional forms can perform quite poorly as a global approximation to a general function subject to the same constraints.

In addition to these estimated flexible functional forms not corresponding well to their Taylor series expansions for the low number of terms in the expansion typically employed in estimation, there is some question about how flexible these parametric functional forms really are. The work of Driscoll, McGuirk and Alwang (1992) shows that imposing separability conditions on the parameters of a parametric model can lead to a reduction of the model's flexibility beyond that implied by the imposed condition.

They define a model as flexible if the value of the function and its first two derivatives can all be independently estimated at a single point in the input space. This definition is motivated by the fact that the values typically considered in economic analyses are the level of production, the marginal productivity, and measures of elasticity, which are all determined by the function value and the values of its first two derivatives. They show that imposing separability conditions on commonly used parametric forms leaves less estimable parameters than those required for flexibility at a point.

We propose a nonparametric model which imposes strong separability of every input from every other input on the production function. Although it is not completely flexible at any single point in the support of the production inputs, it is nearly so and it maintains its level of flexibility globally. A parametric model may be more flexible at a single point, although commonly used forms such as the translog are not, no parametric model can maintain flexibility globally. In addition to estimating the function itself we also estimate its derivatives allowing for the estimation of various measures of elasticity.

5.2. The Model. Strong separability of every good from every other good requires a production function to be of the form,

$$y = G\left(\sum_{\alpha=1}^{d} g_{\alpha}\left(x_{\alpha}\right)\right), \tag{4}$$

where $G(x)$ is a monotonic function. In this paper we consider a slightly restricted form of (5). The model we estimate is of the form

$$\ln\left(y\right) = \sum_{\alpha=1}^{d} g_{\alpha}\left(x_{\alpha}\right) + c. \tag{5}$$

This model could be viewed as a nonparametric generalization of the Cobb-Douglas production technology. In the Cobb-Douglas model $g_{\alpha}\left(x_{\alpha}\right) = \beta_{\alpha} x_{\alpha}$, while we allow for arbitrary $g_{\alpha}\left(x_{\alpha}\right)$'s.

The model given by (5) is not completely flexible with respect to the definition given in Driscoll, McGuirk and Alwang (1992). They show that for a function to be flexible at a point and strongly separable the functional form needs $2d+2$ independent values for the function value and its first two derivatives. The model considered here allows for the independent estimation of the function value and d first derivatives and d second derivatives. These derivatives, whether they are estimated or not, are only constrained by smoothness conditions in the nonparametric regression. This gives $2d+1$ independent values for estimation at each point which is one less than that required for their definition of flexibility. However this independence holds globally. It is possible that a parametric model may have greater flexibility at a single point, but then the values for all other points are determined. The Cobb-Douglas model allows for the independent estimation of $d+1$ parameters and so is not as flexible as (5) at any point.

Model (5) remains unchanged if we rewrite it as

$$\ln(y) = \sum_{\alpha=1}^{d} f_\alpha(\ln(x_\alpha)) + c. \tag{6}$$

This form has the advantage of giving relatively simple expressions for measures of elasticity. The elasticity of output with respect to input x_α is simply $f'_\alpha(\ln(x_\alpha))$, so that scale elasticity can be expressed as

$$\theta = \sum_{\alpha=1}^{d} f'_\alpha(\ln(x_\alpha)). \tag{7}$$

Note that (6) and (7) can be expressed in terms of the functions $f_\alpha(\cdot)$ and their derivatives which can be estimated using the methods presented in the Section 3.

5.3. Estimation Results. We consider the estimation of a production function for livestock in Wisconsin. We use a subset (250 observations) of an original data set of over 1000 Wisconsin farms collected by the Farm Credit Service of St. Paul, Minnesota in 1987. The data were cleaned, removing outliers and incomplete records and selecting farms which only produce animal outputs. The data consists of farm level inputs and outputs measured in dollars. The output (Y) used in this analysis is livestock, and the input variables used are family labor, hired labor, miscellaneous inputs (repairs, rent, custom hiring, supplies, insurance, gas, oil, and utilities), animal inputs (purchased feed, breeding, and veterinary services), and intermediate run assets (assets with a useful life of one to 10 years) resulting in a five dimensional X variable.

We estimated the model using a normal kernel. The data were rescaled to have mean zero and standard deviation one. We used a bandwidth $h = .3$ in the direction of interest and $g = 1.2$ in the directions not of interest for estimating the functions and set $h = .4$ and $g = 1.3$ for the estimation of the derivatives. In the direction of interest the bandwidths are very close to the optimal bandwidths found by the procedure outlined in Section 3. We slightly oversmoothed here in order to present a less wiggly estimator. We found that in order to get reasonable results in areas of sparse data we had to choose a larger bandwidth in the directions not of interest. Because of this it is likely that we have a larger bias than one would expect based on Theorem 1. It is probable that using a design adaptive bandwidths would alleviate this finite sample problem, but investigating the behavior of such bandwidths in this setting is beyond the scope of this paper. The results of the estimation of the additive components and their derivatives are displayed in Figure 1 and Figure 2. The graphs in Figure 1 give some indication of nonlinearity. Figure 2 shows that for the two labor inputs and animal inputs this effect is real and systematic. The graphs seem to indicate that the elasticities for these inputs increases with their use. The sum of the derivatives (scale elasticity) is also shown in Figure 2. In order to get some idea of the variability of the estimates, confidence bands for the derivatives were constructed using the wild bootstrap method of Härdle and Marron (1991).

Although we can not reject the hypothesis that the scale elasticity is constant there seems to be a strong indication that scale elasticity increases with farm size. Our estimate of scale elasticity is greater than one, indicating increasing returns to scale, for larger farms, however, we can not reject constant or diminishing returns to scale everywhere based on our results. Eventhough our evidence is far from conclusive this study does give some indication that there are nonmarket forces constraining the amount of livestock produced on

Wisconson farms, since a farmer producing at an increasing returns to scale portion of the production frontier could increase profit by increasing production. There is stronger evidence of this effect for larger farms. It is worth noting that the elasticity estimates from a linear, Cobb-Douglas, model systematically underestimate the elasticities in regions of high data density. The estimates for the Cobb-Douglas model are also displayed in Figure 2.

Farms producing on regions of the production frontier which exhibit increasing returns to scale implies that these farms are not behaving as profit maximizers. Farms facing a production function which exhibits increasing returns to scale could increase profits by increasing the scale of their operation. It is very likely that there exist nonmarket pressures (*e.g.* liquidity constraints, imperfect land markets) which prevent them from increasing the size of their operation. It is interesting to note that this seems to be more the case for larger farms. This implies that alleviating these constraints on farm size might have little effect on smaller farms since they would receive no marginal benefits and would have to increase their size drastically to realize any gain in profit.

6. Conclusion. In this paper the integration idea of Linton and Nielson (1995) is applied to the estimation of the derivatives of the regression functions in an additive model. The results are obtained by averaging a local polynomial regression over the sample rather than by just averaging a kernel estimator. The derivatives are easily obtained from the local polynomial regression. Also, by using local polynomial regression instead of kernel regression the estimator is design adaptive since the bias is independent of $\varphi(\cdot)$.

In our presentation the one dimensional rate of convergence can only be realized for restricted values of d. One can weaken the restriction on d by exploiting extra smoothness in the directions not of interest. Although this can be done by choosing $L(\cdot)$ to be a higher order kernel, in practice a better idea might be to use a higher order local polynomial in the directions not of interest in the initial estimation.

The application in Section 5 demonstrates these methods in practice. Although the results are not conclusive there is some evidence that the regression is, in fact, nonlinear and that scale elasticity increases with farm size. If it is true that there are farms producing on regions of the production frontier which have increasing returns to scale it is likely that there are nonmarket forces at work which are constraining production to lower levels than the profit maximizing level. This study seems to indicate that this is more likely to be a problem faced by the larger farms in Wisconsin.

Acknowledgements: We would like to thank Oliver Linton, Joel Horowitz and Lijian Yang for helpful discussion and comments.

A. Appendix. This section establishes results characterizing the asymptotic behavior of the estimator. The following conditions are assumed to hold.

- A1: *The kernels* $K(\cdot)$ *and* $L(\cdot)$ are *positive, bounded, symmetric, compactly supported and Lipschitz continuous with* $\int K(u)\,du = 1$. $L(\cdot)$ is of order q.
- A2: *Bandwidths satisfy* $\frac{nhg^{2(d-1)}}{\ln^2(n)} \longrightarrow \infty$, $\frac{g^q}{h^{p+1-\nu}} \to 0$ *and* $h = h_0 n^{\frac{-1}{2p+3}}$.
- A3: *The functions* $f_s(\cdot)$*'s have bounded Lipschitz continuous* $(p+1)^{th}$ *derivatives..*
- A4: *The variance function,* $\sigma^2(\cdot)$, *is bounded and Lipschitz continuous.*
- A5: φ *and* $\varphi_{-\alpha}$ *are uniformly bounded away from zero and infinity and are Lipschitz continuous.*

The proof of Theorem 1 makes use of the following lemmas.

Lemma 1. *Let* $D_n = A + B_n$ *where* A^{-1} *exists and* $B_n = (b_{ij})_{1\le i,j\le p}$ *where* $b_{ij} = O_p(\delta_n)$ *then* $D_n^{-1} = A^{-1}(I + C_n)$ *where* $C_n = (c_{ij})_{1\le i,j\le p}$ *and* $c_{ij} = O_p(\delta_n)$, *where* δ_n *denotes a function of* n.

Proof: $D_n = (I + B_n A^{-1})A$ then D_n is invertible and has an inverse given by $D_n^{-1} = A^{-1}\left(I + \sum_{i=1}^{\infty}(B_n A^{-1})^i\right)$ if and only if the series on the right hand side converges with respect to the usual matrix norm. $\|B_n A^{-1}\| \le \|B_n\|\,\|A^{-1}\|$, and $\|B_n\| \le \sum(b_{ij}) = O_p(\delta_n)$. With probability one $\|B_n\| \le \frac{1}{2}$, so with probability tending to one D_n is invertible. Since $\|D_n^{-1} - A^{-1}\| \le \sum_{i=1}^{\infty}\|B_n A^{-1}\|^i\,\|A^{-1}\|$ and since $\max\{|c_{ij}|\} \le \|C_n\|$ and A is invertible the result follows.

Lemma 2. $(H^{-1}Z'W_l Z H^{-1})^{-1} = \frac{1}{\varphi(x_\alpha, \overline{X}_l)} S^{-1}\left(I + O_p\left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}}\right)\right)$ *uniformly, where* W, Z *and* S *are defined above and* $H = diag(h^{i-1})_{i=1,\cdots,p+1}$.

Proof: The elements of $H^{-1}Z'WZH^{-1}$ can all be expressed in the form

$$\frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{i\alpha}-x_\alpha\right) L_g\left(\bar{X}_i-\bar{X}_l\right)\left(\frac{x_{i\alpha}-x}{h}\right)^m$$

$$= E\left[K_h\left(X_{i\alpha}-x_\alpha\right) L_g\left(\bar{X}_i-\bar{X}_l\right)\left(\frac{x_{i\alpha}-x}{h}\right)^m\right] + o_p(n^{\frac{-1}{2}})$$

$$= \int u^m K(u) L\left(\overline{u}\right)\varphi\left(x_\alpha+hu, \bar{X}_l+g\overline{u}\right) du d\overline{u} + o_p(n^{\frac{-1}{2}})$$

$$= \varphi(x_\alpha, \bar{X}_l)\int u^m K(u)du + O_p\left(h+\frac{\ln n}{\sqrt{nhg^{d-1}}}\right).$$

The result is obtained by applying lemma 1. It should be noted that the $o_p(1)$ is uniform over the interior of the support of $\varphi\left(x,\overline{x}\right)$, by the bandwidth conditions and Silverman (1986).

Proof of Theorem 1:

Define $E_i[W] = E[W \mid X_i]$ and $E_*[W] = E[W \mid X_1, \cdots, X_n]$. Let $\varphi_{\underline{\alpha}}(\cdot)$ be the marginal density of $\bar{X}_1$. Let

$$F_i = \begin{bmatrix} f_\alpha\left(x_\alpha\right)+\sum_{\beta\neq\alpha} f_\beta\left(X_{i\beta}\right) \\ f'_\alpha\left(x_\alpha\right) \\ \vdots \\ \frac{1}{p!}f^p_\alpha\left(x_\alpha\right) \end{bmatrix},$$

where $\delta_i^j = 1$ if $i = j$ and 0 otherwise. The difference between the function and the estimate can be written as

$$
\begin{aligned}
\frac{1}{\nu!}\left(\hat{f}_\alpha^\nu(x_\alpha) - f_\alpha^\nu(x_\alpha)\right) &= \frac{1}{n}\sum_{i=1}^{n} E_\nu' \left(Z'W_iZ\right)^{-1} Z'W_iY - \frac{1}{\nu!}f_\alpha^\nu(x_\alpha) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\nu' \left(Z'W_iZ\right)^{-1} Z'W_iY - E_\nu' \left(Z'W_iZ\right)^{-1} Z'W_iZF_i + O(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\nu' \left(Z'W_iZ\right)^{-1} Z'W_i\left[Y - ZF_i\right] + O(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\nu' \left(Z'W_iZ\right)^{-1} Z'W_i\left[Y - ZF_i\right] + O(n^{-1/2}) \\
&= \frac{1}{h^\nu n}\sum_{i=1}^{n} E_\nu' \left(H^{-1}Z'W_iZH^{-1}\right)^{-1} H^{-1}Z'W_i\left[Y - ZF_i\right] + O(n^{-1/2}) \\
&= \frac{1}{h^\nu n}\sum_{i=1}^{n} \frac{1}{\varphi(x_\alpha, \underline{X}_i)} E_\nu' S^{-1}\left(I + O_p\left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}}\right)\right) H^{-1}Z'W_i \\
&\quad \left[Y - ZF_i\right] + O(n^{-1/2})
\end{aligned}
$$

Writing the above in terms of sums gives

$$
\begin{aligned}
\frac{1}{\nu!}\left(\hat{f}_\alpha^\nu(x_\alpha) - f_\alpha^\nu(x_\alpha)\right) &= \frac{1}{h^\nu n}\sum_{i=1}^{n} \frac{1}{\varphi(x_\alpha, \underline{X}_i)} \frac{1}{n}\sum_{l=1}^{n} K_{\nu h}^*\left(X_{l\alpha} - x_\alpha\right) L_g\left(\overline{X}_l - \overline{X}_i\right) \\
&\times \left(1 + O_p\left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}}\right)\right) \left[\sum_{\beta\neq\alpha}\left\{f_\beta\left(X_{l\beta}\right) - f_\beta\left(X_{i\beta}\right)\right\} + \frac{f_\alpha^{p+1}(x)}{(p+1)!}\right. \\
&\times \left.\left(X_{l\alpha} - x_\alpha\right)^{p+1} + O\left(\left(X_{l\alpha} - x_\alpha\right)^{p+2}\right) + \epsilon_l\right] + O\left(n^{-1/2}\right)
\end{aligned} \tag{8}
$$

It can be seen that he kernel, $K_{\nu h}^*(\cdot)$, is of order $(\nu, p+1)$, so that

$$
\int u^q K_{\nu h}^*(u)\, du = \left\{\begin{array}{ll} 0 & q \leq p, \quad q \neq \nu \\ 1 & q = \nu \\ \Lambda \neq 0 & q = p+1 \end{array}\right\}.
$$

where Λ is some constant. The last condition follows from $p - \nu$ odd.

The proof of the theorem partly follows Chen, Härdle, Linton, Severance-Lossin (1995). We separate (8) into a systematic "bias" and a stochastic "variance".

$$\frac{1}{n}\sum_{i=1}^{n}\frac{E_i(\widehat{a}_i)}{\varphi(x_\alpha,\bar{X}_i)}+\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{a}_i-E_i(\widehat{a}_i)}{\varphi(x_\alpha,\bar{X}_i)}+O_p\left(\frac{h}{\sqrt{nhg^{d-1}}}+\frac{\ln n}{nhg^{d-1}}\right)$$

where,

$$\widehat{a}_i=\frac{1}{h^\nu n}\sum_{l=1}^{n}K^*_{\nu h}\left(X_{l\alpha}-x_\alpha\right)L_g\left(\overline{X}_l-\overline{X}_i\right)$$
$$\times\left[\frac{f_\alpha^{p+1}(x_\alpha)}{(p+1)!}\left(X_{l\alpha}-x_\alpha\right)^{p+1}+O\left(\left(X_{l\alpha}-x_\alpha\right)^{p+2}\right)+\sum_{\beta\neq\alpha}\left\{f_\beta\left(X_{l\beta}\right)-f_\beta\left(X_{i\beta}\right)\right\}+\epsilon_l\right]$$

It remains to work with the first order approximations.

Let

$$T_{1n}=\frac{1}{n}\sum_{i=1}^{n}\frac{E_i(\widehat{a}_i)}{\varphi(x_\alpha,\bar{X}_i)}\quad ;\quad T_{2n}=\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{a}_i-E_i(\widehat{a}_i)}{\varphi(x_\alpha,\bar{X}_i)}.$$

We prove the theorem by showing:

I. $T_{1n}=n^{\frac{-(p+1-\nu)}{2p+3}}b_\alpha\left(x_\alpha\right)+O_p(h^{p+2-\nu})$

II. $T_{2n}=\sum_{j=1}^{n}w_{j\alpha}\epsilon_j+O_p(n^{-1/2})$,

where $w_{j\alpha}=\frac{1}{h^\nu n}K^*_{\nu h}(x_\alpha-X_{j\alpha})\frac{\varphi_{\underline{\alpha}}(X_{j\underline{\alpha}})}{\varphi(x_\alpha,X_{j\underline{\alpha}})}$, and $n^{\frac{p+1-\nu}{2p+3}}\sum_{j=1}^{n}w_{j\alpha}\epsilon_j$ obeys a Central Limit Theorem with asymptotic variance as stated in Theorem 1. To see this note that

$$E\left[\left\{n^{\frac{p+1-\nu}{2p+3}}\sum_{j=1}^{n}w_{j\alpha}\epsilon_j\right\}^2\right]=n^{\frac{2(p+1-\nu)}{2p+3}}\sum_{j=1}^{n}E\left[w_{j\alpha}^2\epsilon_j^2\right]=n^{\frac{4p+5-\nu}{2p+3}}E\left[w_{1\alpha}^2\epsilon_1^2\right],$$

since $w_{j\alpha}\epsilon_j$ are mean zero and i.i.d., and

$$E\left[w_{1\alpha}^2\epsilon_1^2\right]\ =\frac{1}{n^2}\int\sigma^2(z,w)K^{*2}_{\nu h}(x_\alpha-z)\frac{\varphi^2_{\underline{\alpha}}(w)}{\varphi^2(x_\alpha,w)}\varphi\left(z,w\right)dzdw.$$

Changing variables to $u = \frac{x_\alpha - z}{h}$ gives

$$
\begin{aligned}
E\left[w_{1\alpha}^2 \epsilon_1^2\right] &= \tfrac{1}{n^2 h^{2\nu+1}} \int \sigma^2(x_\alpha + hu, w) K_v^{*2}(u) \tfrac{\varphi_{\underline{\alpha}}^2(w)}{\varphi^2(x_\alpha, w)} \varphi\left(x_\alpha + hu, w\right) du dw \\
&= n^{-\frac{4p+5-\nu}{2p+3}} \left\|K^*\right\|_2^2 \int \sigma^2(x_\alpha, w) \tfrac{\varphi_{\underline{\alpha}}^2(w)}{\varphi(x_\alpha, w)} dw + o(n^{-\frac{4p+5-\nu}{2p+3}}),
\end{aligned}
$$

by assumption (A4) and the bandwidth conditions. To establish the Lindeberg condition, required for the CLT, note that

$$
\frac{w_{1\alpha}^2 \epsilon_1^2}{E\left[w_{1\alpha}^2 \epsilon_1^2\right]} 1\!\!1 \left[\frac{w_{1\alpha}^2 \epsilon_1^2}{E\left[w_{1\alpha}^2 \epsilon_1^2\right]} \geq \delta n\right] \leq D \epsilon_1^2
$$

for some constant D, by assumptions (A1), (A4) and (A5). The Lindeberg condition then follows from the Lebesgue Dominated Convergence Theorem.

We now establish the approximations in **I** and **II**.

I. Consider $\varphi(x_\alpha, \bar{X}_i)^{-1} E_i(\widehat{a}_i)$, which is, in fact, an approximation of the conditional bias of the Nadaraya-Watson estimator at $(x_\alpha, \bar{X}_i)$. This is,

$$
\begin{aligned}
\varphi(x_\alpha, \bar{X}_i)^{-1} E_i(\widehat{a}_i) &= E_i\Bigg[\frac{1}{\varphi(x, \overline{x}_i)} h^{-\nu} n^{-1} \sum_{l=1}^{n} L_g(\bar{X}_l - \bar{X}_i) K_{\nu h}^*(X_{l\alpha} - x_\alpha) \\
&\quad \times \Bigg(\frac{f_\alpha^{p+1}(x_\alpha)}{(p+1)!} (X_{l\alpha} - x_\alpha)^{p+1} + O\left((X_{l\alpha} - x_\alpha)^{p+2}\right) \\
&\quad + \sum_{\beta \neq \alpha} f_\beta(X_{l\beta}) - \sum_{\beta \neq \alpha} f_\beta(X_{i\beta}) + \epsilon_l\Bigg)\Bigg] \\
&= \frac{h^{-\nu}}{\varphi(x_\alpha, \bar{X}_i)} \int L_g(w - \bar{X}_i) K_{\nu h}^*(z - x_\alpha) \varphi(z, w) \Bigg(\frac{f_\alpha^{p+1}(x_\alpha)}{(p+1)!} \times \\
&\quad (z - x_\alpha)^{p+1} + O\left((z - x_\alpha)^{p+2}\right) + \sum_{\beta \neq \alpha} f_\beta(w) - \sum_{\beta \neq \alpha} f_\beta(X_{i\beta})\Bigg) dw dz,
\end{aligned}
$$

since $E_*[\epsilon_i] = 0$. We now change variables to $u = \frac{z-x_\alpha}{h}$ and $v = \frac{w-\bar{X}_i}{g}$, where v is a $d-1$-dimensional vector with β^{th} component v_β, so that

$$\begin{aligned}
\varphi(x_\alpha, \bar{X}_i)^{-1} E_i(\widehat{a}_i) &= \frac{h^{-\nu}}{\varphi(x_\alpha, \bar{X}_i)} \int L(v) K_\nu^*(u) \varphi(x_\alpha + hu, \bar{X}_i + gv) \left(\frac{f_\alpha^{p+1}(x)}{(p+1)!} (hu)^{p+1} \right. \\
&\quad \left. + O\left((hu)^{p+2}\right) + \sum_{\beta \neq \alpha} f_\beta(X_{i\beta} + gv_\beta) - \sum_{\beta \neq \alpha} f_\beta(X_{i\beta}) \right) du dv \\
&= h^{p+1-\nu} \mu_{p+1}(K^*) \left\{ \frac{1}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha) \right\} + \delta_0^\nu \sum_{\beta \neq \alpha} f_\beta(X_{i\beta}) \\
&\quad + o_p(h^{p+1-\nu}) + O_p(g^q),
\end{aligned}$$

by assumptions (A1), (A2), (A3) and (A5). Since the $\varphi(x, \overline{x}_i)^{-1} E_i(\widehat{a}_i)$ are independent and bounded we have

$$\begin{aligned}
T_{1n} &= h^{p+1-\nu} \mu_{p+1}(K^*) \left\{ \tfrac{1}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha) \right\} + o_p(h^{p+1-\nu}) + O_p(g^q) + \delta_0^\nu O_p(n^{-1/2}) \\
&= n^{\frac{-(p+1-\nu)}{2p+3}} b_\alpha(x_\alpha) + o_p\left(h^{p+1-\nu}\right).
\end{aligned}$$

II. We now turn to the stochastic term,

$$T_{2n} = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{a}_i - E_i(\widehat{a}_i)}{\varphi(x_\alpha, \bar{X}_i)}.$$

We further write

$$\widehat{a}_i - E_i(\widehat{a}_i) = \widehat{a}_i - E_*(\widehat{a}_i) + E_*(\widehat{a}_i) - E_i(\widehat{a}_i).$$

We show that $\frac{1}{h^\nu n} \sum_{i=1}^n \frac{\widehat{a}_i - E_*(\widehat{a}_i)}{\varphi(x_\alpha, \overline{x}_i)} = \sum_{j=1}^n w_{j\alpha} \epsilon_j + O_p\left(n^{-1/2}\right)$, where

$$\widehat{a}_i - E_*(\widehat{a}_i) = h^{-\nu} n^{-1} \sum_{j=1}^n K_{\nu h}^*(x_\alpha - X_{j\alpha}) L_g(\bar{X}_i - \bar{X}_j) \epsilon_j.$$

Therefore,

$$\begin{aligned}
\tfrac{1}{n} \textstyle\sum_{i=1}^n \frac{\widehat{a}_i - E_*(\widehat{a}_i)}{\varphi(x_\alpha, \bar{X}_i)} &= \tfrac{1}{h^\nu n} \textstyle\sum_{i=1}^n \frac{1}{\varphi(x_\alpha, \bar{X}_i)} n^{-1} \sum_{j=1}^n K_{\nu h}^*(x_\alpha - X_{j\alpha}) L_g(\bar{X}_i - \bar{X}_j) \epsilon_j \\
&= h^{-\nu} n^{-1} \textstyle\sum_{j=1}^n K_{\nu h}^*(x_\alpha - X_{j\alpha}) \epsilon_j \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x, \overline{x}_i)} L_g(\bar{X}_i - \bar{X}_j) \right\} \quad (9) \\
&= \textstyle\sum_{j=1}^n w_{j\alpha} \epsilon_j \left\{ 1 + \tau_i \right\},
\end{aligned}$$

where $\tau_i = o_p(h)$ and independent of all the ϵ_j's. The last equality is demonstrated as follows. Let

$$\eta_j = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\varphi(x_\alpha, \bar{X}_i)}L_g(\bar{X}_i - \bar{X}_j) \ ,$$

and break η_j into $E_j[\eta_j] + \{\eta_j - E_j[\eta_j]\}$. Then,

$$\begin{aligned} E_j[\eta_j] &= \int \frac{1}{\varphi(x_\alpha, z)} L_g(z - \bar{X}_j)\varphi_{\underline{\alpha}}(z)\,dz \\ &= \int \frac{1}{\varphi(x_\alpha, \bar{X}_j + gu)} L(u)\varphi_{\underline{\alpha}}(\bar{X}_j + gu)du \\ &= \frac{\varphi_{\underline{\alpha}}(\bar{X}_j)}{\varphi(x_\alpha, \bar{X}_j)} + O_p(g^q). \end{aligned}$$

Also,

$$\begin{aligned} E_j\left[\{\eta_j - E_j[\eta_j]\}^2\right] &\le \frac{1}{n}\int\left\{\frac{1}{\varphi(x_\alpha, z)}L_g(z - \bar{X}_j) - \frac{\varphi_{\underline{\alpha}}(\bar{X}_j)}{\varphi(x_\alpha, \bar{X}_j)}\right\}^2 \varphi_{\underline{\alpha}}(z)\,dz + O_p\left(n^{-1}g^{2q}\right) \\ &= \frac{1}{n}\int\left\{\frac{1}{\varphi(x_\alpha, z)}L_g(z - \bar{X}_j)\right\}^2 \varphi_{\underline{\alpha}}(z)\,dz + O_p\left(n^{-1}\right). \end{aligned}$$

By a change of variables we get

$$\begin{aligned} E_j\left[\{\eta_j - E_j[\eta_j]\}^2\right] &\le \frac{1}{ng^{d-1}}\int\left\{\frac{1}{\varphi(x_\alpha, \bar{X}_j + gv)}L(v)\right\}^2 \varphi_{\underline{\alpha}}(\bar{X}_j + gv)dv + O_p\left(n^{-1}\right) \\ &= \frac{1}{ng^{d-1}}\frac{\varphi_{\underline{\alpha}}(\bar{X}_j)}{\varphi^2(x_\alpha, \bar{X}_j)}\left\|L\right\|_2^2 + O_p\left(n^{-1}\right) = o_p(h), \end{aligned}$$

by the assumptions (A1), (A2), and (A5). Thus the last line in (9) is shown.

By the same conditioning arguments as in (II.2) in the proof of Theorem 1 in Chen, Härdle, Linton and Severance-Lossin (1995) It can be shown that $\frac{1}{n}\sum_{i=1}^{n}\frac{E_*(\widehat{a}_i) - E_i(\widehat{a}_i)}{\varphi(x_\alpha, \bar{X}_i)} = O_p(n^{-1/2})O(h^{p+1-\nu} + g^q) + O_p(n^{-1/2})O_p((ng^{d-1}h^{-2v-1})^{-1/2}) = o_p\left((n^{-1}h^{-2v-1})^{-1/2}\right)$, which establishes **II** and thus proves the theorem. ∎

Proof of Theorem 2:

An argument identical to that in the proof of Theorem 2 in Chen, Härdle, Linton and Severance-Lossin (1995) shows that the asymptotic covariance between the estimates of $\widehat{f}_\alpha(x_\alpha)$ and $\widehat{f}_\beta(x_\beta)$ is negligible compared with their variances. ∎

REFERENCES

[1] Blackorby, C., D. Primont, and R. Russell. "On Testing Separability Restrictions with Flexible Functional Forms." *Journal of Econometrics* 5(1977):195-209.

[2] Buja, A, Hastie, T.J. and Tibshirani, R.J. (1989), Linear smoothers and additive models, *Ann. Statist.*, 17, 453-510.

[3] Chen, Härdle, Linton and Severance-Lossin (1995), Nonparametric Estimation of Additive Seperable Regression Models

[4] Diewert, W. and T. Wales. "Flexible Functional Forms and Global Curvature Restrictions." *Econometrica* 55(1987):43-68.

[5] Driscoll, P. and Boisvert. "Dual Second- and Third-Order Translog Models of Production" *American Journal of Agricultural Economics* 73(1991):1146-1160.

[6] Driscoll, P., A. McGuirk, and J. Alwang."Testing Hypotheses of Functional Structure: Some Rules for Determining Flexiblity of Restricted Production Models." *American Journal of Agricultural Economics* 74(1992):100-108.

[7] Fan, J. (1992). Design-adaptive nonparametric regression, *J. Amer. Stat. Assoc.,* 87, 998-1004.

[8] Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Stat.,* 21, 196-216.

[9] Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and Engel, J. (1993). Local Polynomial Fitting: A Standard for Nonparametric Regression, *CORE Discussion Paper 9315.*

[10] Guikey, D., C. Lovell, and R. Sickles. " A comparison of the Performance of Three Flexible Functional Forms." *International Economic Review* 24(1983):591-616.

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

[12] Härdle, W. Klinke, S. and Turlach, B.A. (1995) XploRe - an interactive statistical computing environment. Springer Verlag, New York .

[13] Härdle, W and Marron, J.S. (1991), Bootstrap simultaneous error bars for nonparametric regression, *Ann. Stat.,* 19, 778-796.

[14] Härdle, W. and Tsybakov, A.B. (1995) Additive nonparametric regression on principal components, *J. Nonparametric Statistics*, in press

[15] Hastie, T. J. and R. J. Tibshirani (1990), Generalized additive models, Chapman and Hall: London

[16] Linton, O. and J. P. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika* **82**, 93-101.

[17] Leontief, W. (1947) Introduction to a theory of an internal structure of functional relationships. *Econometrika,* 15, 361-373.

[18] Masry, E. and Tjøstheim, D. (1993), Nonparametric estimation and identification of ARCH and ARX nonlinear time series. Convergence properties and rates. *working paper, University of Bergen, Dept. of Math.*

[19] Nadaraya, E.A. (1964), On estimating regression. *Theor. Probab. Appl.*, 9, 141-142.

[20] SILVERMAN, B.W. (1986), Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

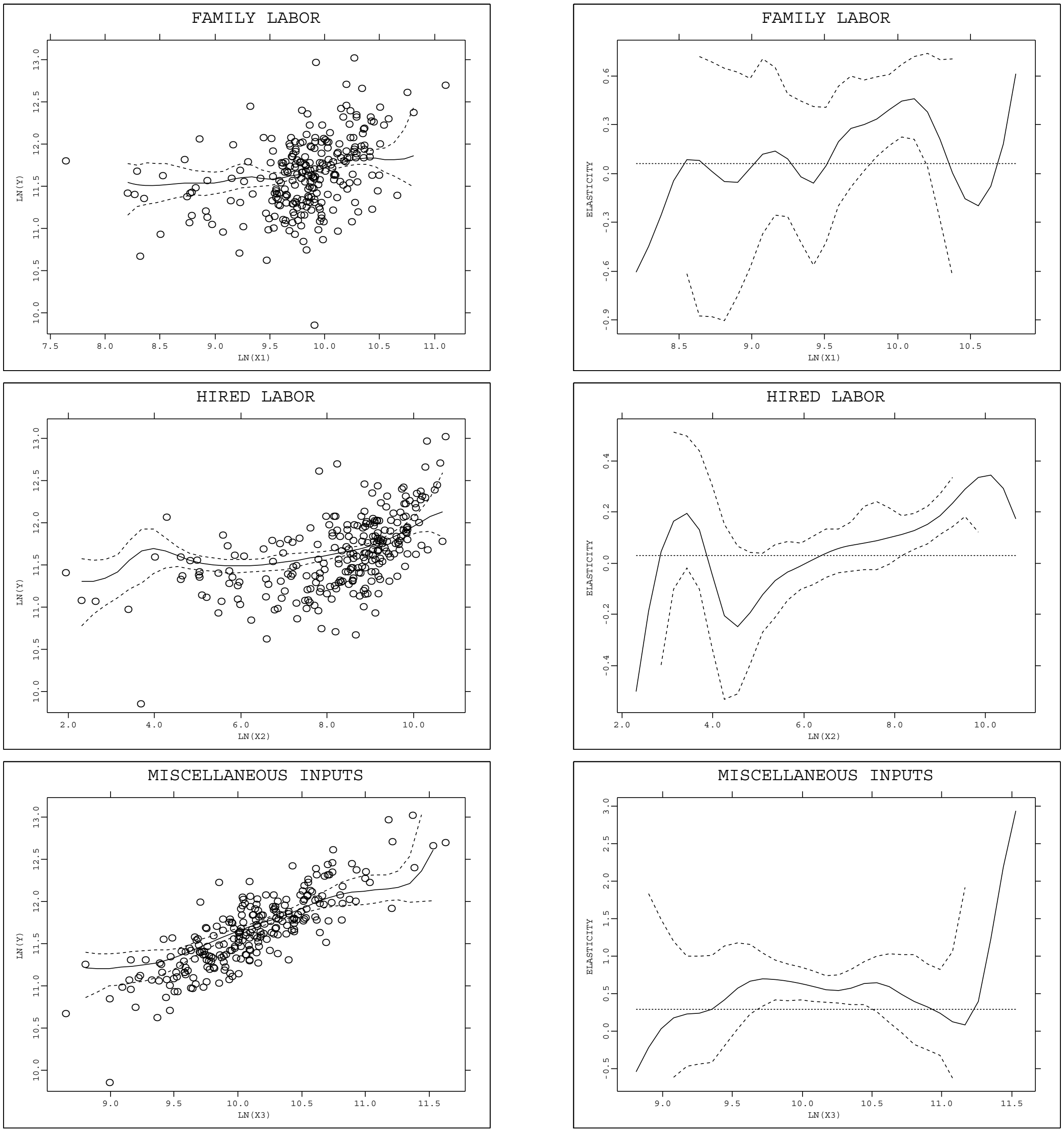
[21] STONE, C.J. (1985), Additive regression and other nonparametric models, *Ann. Statist.* 13, 689 - 705

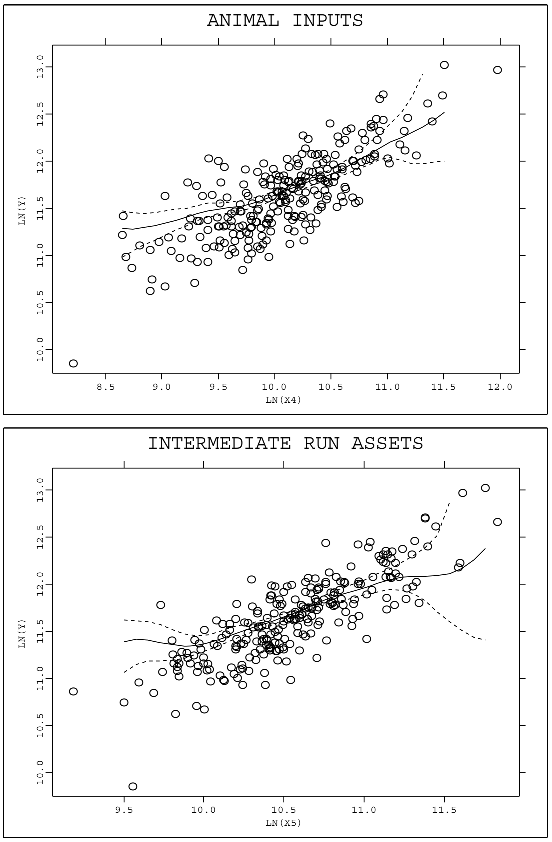
[22] STONE, C.J. (1986), The dimensionality reduction principle for generalized additive models, *Ann. Statist.* 14, 590- 606

[23] TJØSTHEIM, D. AND AUESTAD, B.H. (1994), Nonparametric identification of nonlinear time series: projections, *J. American Statistical Association,* 89, 1398-1409

[24] WALES, T. "On the Flexibility of Flexible Functional Forms: An Empirical Approach" *Journal of Econometrics* 5(1977):183-193.

[25] WATSON, G.S. (1964), Smooth regression analysis. *Sankhyā A*, 26, 359-372


FAMILY LABOR
LN(Y)
LN(X1)
FAMILY LABOR
ELASTICITY
LN(X1)
HIRED LABOR
LN(Y)
LN(X2)
HIRED LABOR
ELASTICITY
LN(X2)
MISCELLANEOUS INPUTS
LN(Y)
LN(X3)
MISCELLANEOUS INPUTS
ELASTICITY
LN(X3)



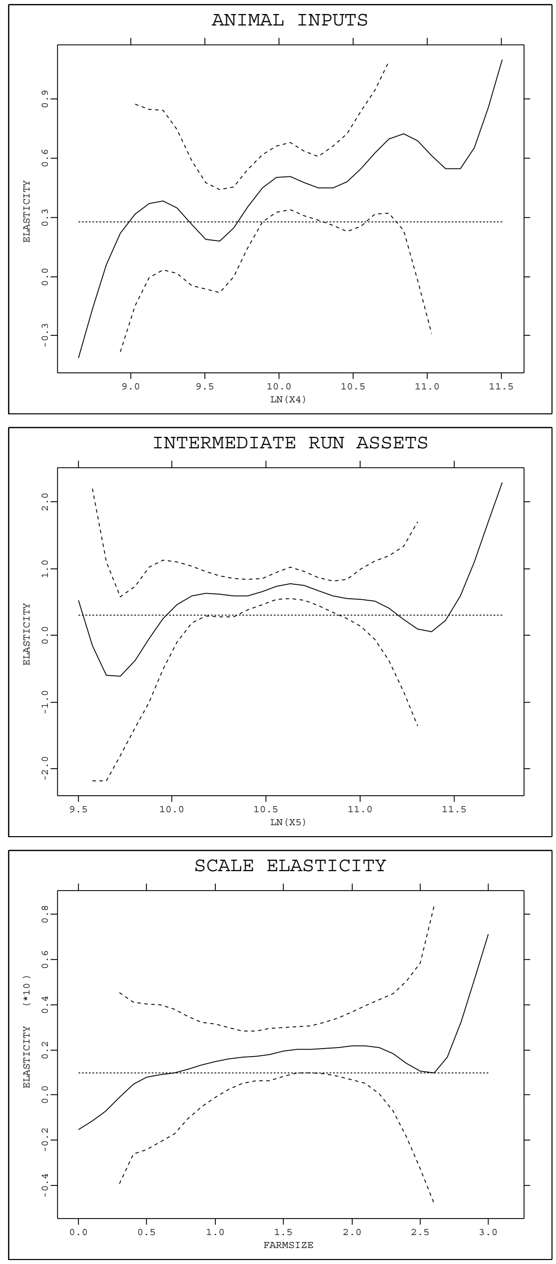


Figure 1 (left): *Additive functions with 90% pointwise confidence intervals and data points*

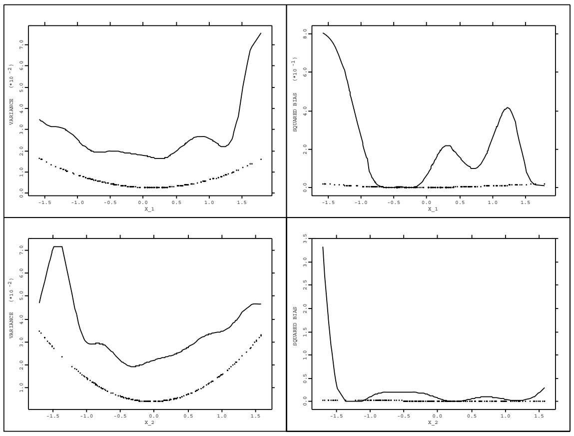
Figure 2 (right): *Derivative estimates with 90% pointwise confidence intervals and linear estimates*

Table 1

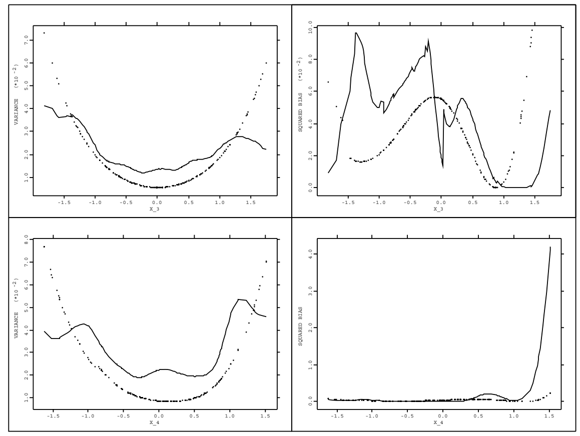
		for the add. functions				for the first derivatives			
		$\hat{g}_1$	$\hat{g}_2$	$\hat{g}_3$	$\hat{g}_4$	$\hat{g'}_1$	$\hat{g'}_2$	$\hat{g'}_3$	$\hat{g'}_4$
$N(.2)$	nad.wat.	.110	.191	.140	.101	.138	.516	.082	.313
	loc.pol.	.069	.083	.117	.081	.167	.018	.036	.153
$N(.2)$	nad.wat.	.169	.046	.071	.062	.212	.372	.064	.447
	loc.pol.	.075	.022	.073	.045	.013	.029	.058	.084

Table 2

		for the add. functions				for the first derivatives			
		$\hat{g}_1$	$\hat{g}_2$	$\hat{g}_3$	$\hat{g}_4$	$\hat{g'}_1$	$\hat{g'}_2$	$\hat{g'}_3$	$\hat{g'}_4$
U^4	nad.wat.	.212	.280	.128	.203	.708	1.271	.111	2.896
	loc.pol.	.095	.110	.319	.081	.007	.021	.463	.729
U^4	nad.wat.	.512	.215	.287	.504	.362	1.687	.214	4.410
	loc.pol.	.137	.060	.206	.212	.134	.011	.285	.415

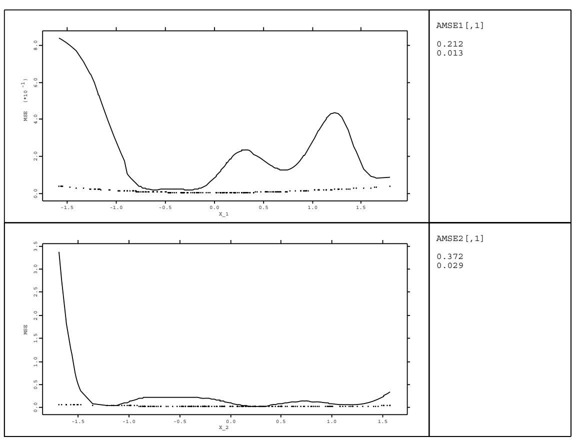


The dashed lines are the loc.pol., the solid lines Nad.Wat.. On the left side are plotted the variances of $\hat{g}'_1$ (upper) and $\hat{g}'_2$ (lower), on the right the squared biases.

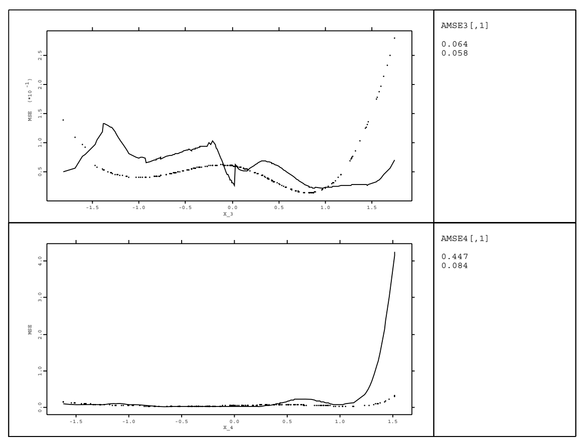


The dashed lines are the loc.pol., the solid lines with Nad.Wat.. On the left side are plotted the variances of $\hat{g}'_3$ (upper) and $\hat{g}'_4$ (lower), on the right the squared biases.

FIGURES 3-4



Dashed lines are the loc.pol., solid Nad.Wat. On the left are the MSEs of $\hat{g}'_1$ and $\hat{g}'_2$, on the right the averaged MSEs.



Dashed lines are the loc.pol., solid Nad.Wat. On the left are the MSEs of $\hat{g}'_1$ and $\hat{g}'_2$, on the right the averaged MSEs.

FIGURES 5-6