Nonparametric Estimation of Additive Models with Homogeneous Components

Wolfgang Härdle

Woocheol Kim

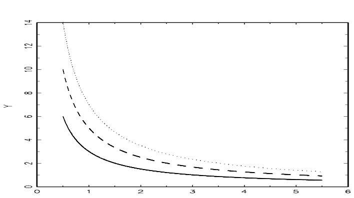
Gautam Tripathi

Sonderforschungsbereich 373

Wirtschaftswissenschaftliche Fakultät

Humboldt–Universität zu Berlin

sfb.wiwi.hu-berlin.de



A function, $f(\cdot)$ is homogeneous of degree α, if

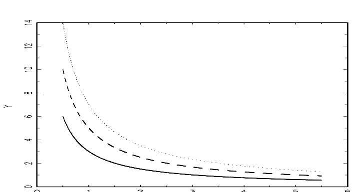
$$f(\lambda x_1, .., \lambda x_d) = \lambda^\alpha f(x_1, .., x_d).$$

Examples:

i) Linear models: $f(x) = x^T \beta \ \ (\alpha = 1)$

ii) Cobb-Douglas : $f(x) = x_1^{\alpha_1} x_2^{\alpha_2}, \ (\alpha = \alpha_1 + \alpha_2)$

iii) CRS-technology: $f(x) = a(x_1^\rho + x_2^\rho)^{1/\rho}, (\alpha = 1)$



Why is the concept of homogeneity important ?

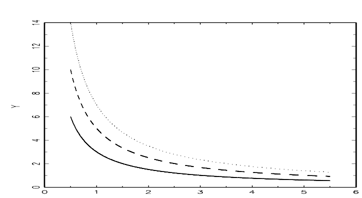
- Characterization of production functions

$\alpha < 1$		decreasing	
$\alpha = 1$	$\Longleftrightarrow$	constant	returns to scale
$\alpha > 1$		increasing	

- In the theory of producers, cost-minimizing(profit-maximizing) behavior of competitive firms implies their cost(profit) functions are linearly homogeneous in input (and output) prices.

$$C = c(y, p_I) \text{ s.t. } c(y, \lambda p_I,) = \lambda c(y, p_I)$$

$$\pi = \pi(p_I, p_O) \text{ s.t. } \pi(\lambda p_I, \lambda p_O) = \lambda \pi(p_I, p_O)$$



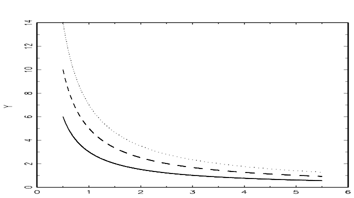
Nonparametric Models with Homogeneous Restriction

- The estimation has been carried out only in parametric forms. Christensen and Greene (1976) analyzed the cost function of electricity generation in the US with inputs of capital, labor, and fuel.
- Partial Linear Model. Tripathi (2000) 'efficiency bound for β' with homogeneous $f(\cdot)$:

$$Y_i = Z_i^T \beta + f(X_i) + \varepsilon_i$$

- Nonparametric Model. Tripathi and Kim (1999) with homogeneous $f(\cdot)$:

$$Y_i = f(X_i) + \varepsilon_i$$



Objective

Analyze nonparametric *additive* models where at least one component is restricted to be homogeneous.

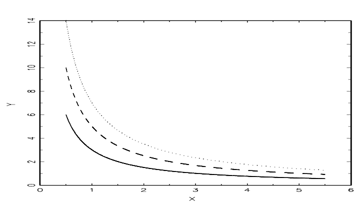
$$Y_i = f_1(X_i) + f_2(Z_i) + \varepsilon_i, \quad (\varepsilon_i : \text{i.i.d.}),$$
$$\text{where } f_1(\cdot) \text{ is homogeneous.}$$

Example:

Y_i : total costs

$f_1(X_i)$: variable costs (capital, labor,..)

$f_2(Z_i)$: fixed costs



Extension

an option pricing model

Consider a *nonparametric* option pricing model,

$$\Pi_t = f_1\left(S_t, K, T-t, X_t\right),$$

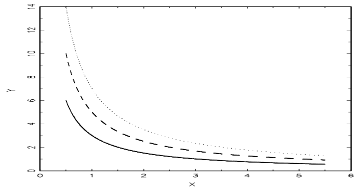
$$\begin{aligned} \Pi_t &= \text{option price} \\ S_t &= \text{price of underlying asset} \\ K &= \text{exercise price} \\ T-t &= \text{time to expiration} \\ X_t &= \text{other var. } (S_{t-1} \text{ or volatility}). \end{aligned}$$

Garcia and Renault (1996) showed $f_1(\cdot)$ is homogeneous of degree one in (S_t, K).

Under multiplicative assumption, the pricing model is

$$\Pi_t = f_1(S_t, K) f_2\left(T-t, X_t\right),$$

where $f_1(\cdot)$ is linearly homogeneous.



Imposing Homogeneity

Numeraire Approach

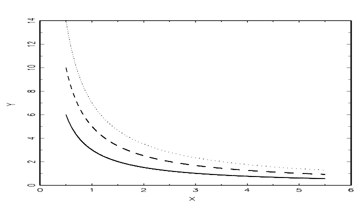
From the homogeneity,

$$f_1\left(X_{1i},..,X_{di}\right)=X_{di}^{\alpha_1}f_1\left(X_{1i}/X_{di},..,X_{(d-1)i}/X_{di},1\right).$$

By defining
$\beta_1\left(U_i\right)=f_1\left(X_{1i}/X_{di},..,X_{(d-1)i}/X_{di},1\right)$ with
$U=\left(X_{1i}/X_{di},..,X_{(d-1)i}/X_{di},1\right)$, reparametrize into

$$Y_i=X_{di}^{\alpha_1}\beta_1\left(U_i\right)+f_2\left(Z_i\right)+\varepsilon_i. \tag{1}$$

Since α is known, we only estimate $\beta_1\left(\cdot\right)$ and construct $\widehat{f_1}\left(x\right)=x_d^{\alpha_1}\widehat{\beta}_1\left(u\right)$.



General Model

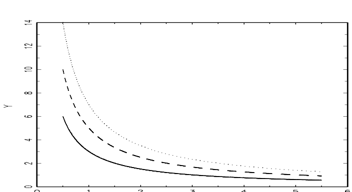
Assume $f_2(\cdot)$ is also homogeneous, then,

$$Y_i = X_{di}^{\alpha_1} \beta_1 (U_i) + Z_{si}^{\alpha_2} \beta_2 (V_i) + \varepsilon_i. \quad (2)$$

With $Z_{si} = 1$ and $V_i = Z_i$, (2) includes (1) as a special case.

Additional Contribution: We extend the theory for Varying-Coefficients Models by Hastie and Tibshirani (1997) or Functional Coefficients AR models by Tsay (1993).

$$
\begin{aligned}
Y_i &= \sum_{k=1}^{d} X_{ki} \beta_k \left(X_{(d+1)i} \right) + \varepsilon_i, \\
Y_i &= \sum_{k=1}^{d} Y_{i-k} \beta_k \left(Y_{i-d'} \right) + \varepsilon_i
\end{aligned}
$$



Two-Step Estimation Procedure

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

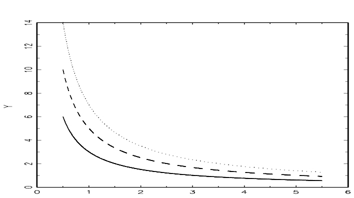
Local Linear Fit :First Step

After locally approximating $\beta_1(\cdot)$ and $\beta(\cdot)$ by linear equations,

$$\min_{b_{1k}\text{'s},\, b_{2k}\text{'s}} \frac{1}{n} \sum_{i=1}^{n} K_h (W_i - w) \times [\ \ y_i - \{b_{10} +$$

$$\sum_{k=1}^{d-1} b_{1k}(\frac{U_{ki} - u_k}{h_1})\} X_{di}^{\alpha_1} - \{b_{20} + \sum_{k=1}^{s-1} b_{2k}(\frac{V_{ki} - v_k}{h_2})\} Z_{si}^{\alpha_2}]^2,$$

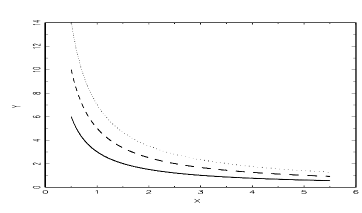
where $w = (u, v)$ and $W_i = (U_i, V_i)$.



- Note that $\widehat{b}_{10}$, the estimate of $\beta_1(u)$, also depends on the value of v. Thus, we denote the level estimates by

$$\begin{bmatrix} \widehat{\beta}_1(u,v) \\ \widehat{\beta}_2(u,v) \end{bmatrix} = \begin{bmatrix} \widehat{b}_{10} \\ \widehat{b}_{20} \end{bmatrix}.$$

- These estimates are *consistent*, but their convergence rates ($n^{\frac{2}{4+(d+s-2)}}$) are *not optimal*, slower than $n^{\frac{2}{4+(d-1)}}$ or $n^{\frac{2}{4+(s-1)}}$. This is a natural result due to the use of the kernel weights, $K_h(W_i - w)$, of dimension, $(d+s-2)$ in our smoothing method.



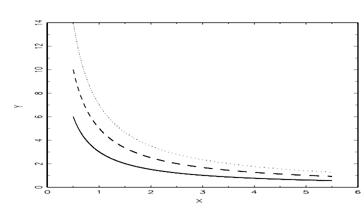
Marginal Integration: Second Step

- For the optimal convergence rate, marginally integrate the pilot estimates of $\widehat{\beta}_{10}(u, V_i)$ over V_i $i = 1, .., n$, i.e.,

$$\widehat{\beta}^*_{10}(u) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\beta}_{10}(u, V_i),$$

similarly,

$$\widehat{\beta}^*_{20}(v) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\beta}_{20}(U_i, v).$$

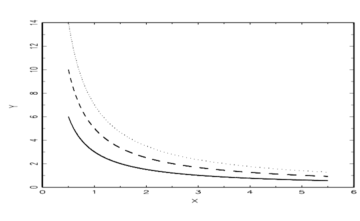


Marginal Integration

Newey (1994), Tjøstheim and Auestadt (1994), and Linton and Nielsen (1995)

- Advantage: *theoretical tractability* in deriving asymptotic properties, in contrast to *backfitting*
- Weakness: *high costs of computations*
- alternative: *Instrumental Variable approach by Kim (1998)*

Finally, for the regression surface, we use
$\widehat{f}^*(x, z) = x_d^{\alpha_1} \widehat{\beta}_1^*(u) + z_s^{\alpha_2} \widehat{\beta}_2^*(v)$

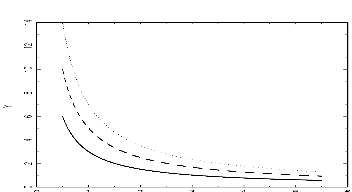


Conditions

A1. $\{Y_i, X_i, Z_i\}_{i=1}^{n}$ *is a random sample, and* ε_i is *i.i.d. with* $E(\varepsilon|X,Z) = 0$ *and* $E\left(\varepsilon^2|X,Z\right) = \sigma^2(X,Z) < \infty$.

A2. (**Continuity and Differentiability**) *The functions of the components, varying-coefficients, and conditional variance, together with the densities(marginal or joint)-*$f_1(\cdot)$, $f_2(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$ $\sigma(\cdot)$, $p_X(\cdot)$, $p_Z(\cdot)$ and $p_{X,Z}(\cdot)$ *are continuous (and hence bounded on the compact support) and twice differentiable with bounded partial derivatives.*

A3. (**Density Functions**) $p_X(\cdot)$, $p_Z(\cdot)$ and $p_{X,Z}(\cdot)$ *are bounded away from zero on the compact supports. Also, conditional density exists and is bounded.*

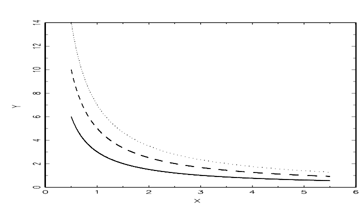


A4. *The matrix* $E\left(W^TW|X_d = x_d, Z_s = z_s\right)$ *is of full rank, and* $E\left(W^TW|X_d = x_d, Z_s = z_s\right)^{-1}$ *is bounded element-wise in a neighborhood of* (x_d, z_s).

A5. (**Kernel Functions**) *The kernel function* K *is positive, compactly supported bounded function, with* $\int K(u)\,du = 1$ *and* $\int uK(u)\,du = 0$. $|K(x_1) - K(x_2)| < c|x_1 - x_2|$ *for all* x_1 *and* x_2 *in its support.*

A6. (**Bandwidth Condition 1**) $h_1 = h_2 = h \to 0$ *and* $nh^{d+s-2} \to \infty$.

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



Main Results I

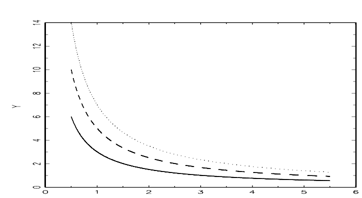
Notation:
$w = (w_1, w_2) = (u, v)$, $\widehat{\beta}_0(w) = \left(\widehat{\beta}_{10}(w), \widehat{\beta}_{20}(w)\right)$

Theorem 1. *Assume that the conditions of A.1 through A.6 hold. Then,*

$$\sqrt{nh^{d+s-2}}\left[\widehat{\beta}_0(w) - \beta_0(w) - BIAS\right] \xrightarrow{\mathcal{L}} N\left(0, \frac{||K||_2^2}{p_W(w)}\Sigma_\beta\right)$$

where

$\mu_K^2 = \int K(u)\, u^2 du$, *and* $||K||_2^2 = \int K^2(r)\, dr$.

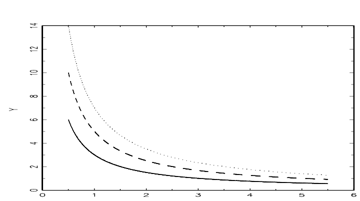


$$BIAS = \frac{h^2}{2}\mu_K^2 \times$$

$$\left[\begin{array}{c} tr\left(D^2\beta_1(w_1)\right) + \frac{E\left(X_d^{\alpha_1} Z_s^{\alpha_2} | W=w\right)}{E\left(X_d^{2\alpha_1} | W=w\right)} tr\left(D^2\beta_2(w_2)\right) \\ tr\left(D^2\beta_2(w_2)\right) + \frac{E\left(X_d^{\alpha_1} Z_s^{\alpha_2} | W=w\right)}{E\left(Z_s^{2\alpha_2} | W=w\right)} tr\left(D^2\beta_1(w_1)\right) \end{array}\right],$$

$$\Sigma_\beta(W) \equiv$$

$$\left[\begin{array}{cc} \frac{E_{|W}(X_d^{2\alpha_1}\sigma_\varepsilon^2(W,X_d,Z_s))}{E_{|W}^2(X_d^{2\alpha_1})} & \frac{E_{|W}(X_d^{\alpha_1} Z_s^{\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s))}{E_{|W}(X_d^{2\alpha_1})E_{|W}(Z_s^{2\alpha_2})} \\ \frac{E_{|W}(X_d^{\alpha_1} Z_s^{\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s))}{E_{|W}(X_d^{2\alpha_1})E_{|W}(Z_s^{2\alpha_2})} & \frac{E_{|W}(Z_s^{2\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s))}{E_{|W}^2(Z_s^{2\alpha_2})} \end{array}\right]$$

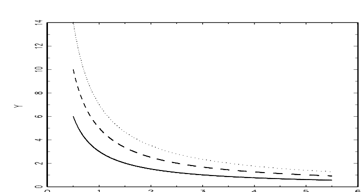


Remark 2

- the convergence rate, $\sqrt{nh^{d+s-2}}$, from using the kernel function which is defined on $\mathbb{R}^{d-1} \times \mathbb{R}^{s-1}$.

- the bias of $\widehat{\beta}_{10}(u, v)$ is similar to the local linear fit in Fan (1992), a function of "second derivatives only", except that it depends on $D^2\beta_2(v)$, which is a natural extension of Tripathi and Kim (1999) dealing with $Y_i = X_{di}^{\alpha_1}\beta_1(U_i) + \varepsilon_i$.

- For homoscedastic errors, the variance is $||K||_2^2\sigma_\varepsilon^2/p_W(w)$, the standard result.



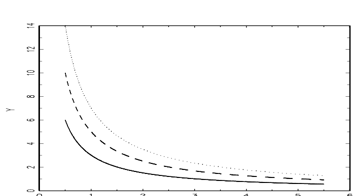
From $\widehat{f}(x,z) = x_d^{\alpha_1}\widehat{\beta}_1(u) + z_s^{\alpha_2}\widehat{\beta}_2(v)$

Corollary 3. *Under the same conditions of Theorem 1,*

$$\sqrt{nh^{d+s-2}}\left[\widehat{f}(x,z) - f(x,z) - BIAS_f\right] \xrightarrow{\mathcal{L}} N\left(0,\ \frac{||K||_2^2}{p_W(w)}\Sigma_f\right),$$

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

$$\Sigma_f = \frac{x_d^{2\alpha_1} E\left(X_d^{2\alpha_1}\sigma_\varepsilon^2(W,X_d,Z_s)|W=w\right)}{E^2\left(X_d^{2\alpha_1}|W=w\right)} + 2\frac{x_d^{\alpha_1} z_s^{\alpha_2} E\left(X_d^{\alpha_1} Z_s^{\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s)|W=w\right)}{E\left(X_d^{2\alpha_1}|W=w\right)E\left(Z_s^{2\alpha_2}|W=w\right)} + \frac{z_s^{2\alpha_2} E\left(Z_s^{2\alpha_2}\sigma_\varepsilon^2(W,X_d,Z_s)|W=w\right)}{E^2\left(Z_s^{2\alpha_2}|W=w\right)}.$$



Main Results II

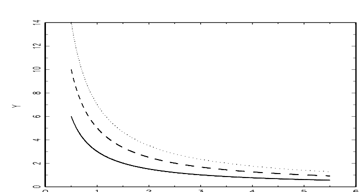
Notation: $\widehat{\beta}_1^*(u) = \frac{1}{n}\sum_{j=1}^{n} \widehat{\beta}_{10}(u, V_j)$

Theorem 4 *Under the conditions of A.1 through A.5 and A.7,*

i) $\sqrt{nh_1^{d-1}} \left[\widehat{\beta}_1^*(u) - \beta_1(u) - BIAS^*(u)\right]$
$\xrightarrow{\mathcal{L}} N\left(0, ||K||_2^2 \Sigma_{\beta_1}\right),$

$$\Sigma_{\beta_1} = \int \frac{p_V^2(s_2)}{p_W(u,s_2)} \frac{E\left(X_d^{2\alpha_1}\sigma_\varepsilon^2(W,X_d)|W=(u,s_2)\right)}{E^2\left(X_d^{2\alpha_1}|W=(u,s_2)\right)} ds_2,$$

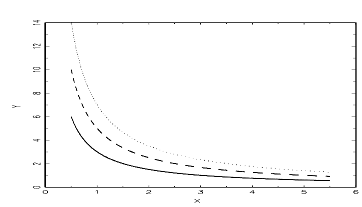
$$BIAS^*(u) = \mu_K^2 [\frac{h_1^2}{2} tr\left(D^2\beta_1(u)\right) + \frac{h_2^2}{2}\int p_V(v) \frac{E\left(X_d^{\alpha_1}Z_s^{\alpha_2}|W=u,v\right)}{E\left(X_d^{2\alpha_1}|W=u,v\right)} tr\left(D^2\beta_2(v)\right) dv],$$



ii) $\sqrt{nh_1^{d-1}}\left[\widehat{f}_1^*(x) - f_1(x) - BIAS^*_{f_1}(x)\right]$
$\xrightarrow{\mathcal{L}} N\left(0,\ ||K||_2^2\Sigma_{f_1}\right),$

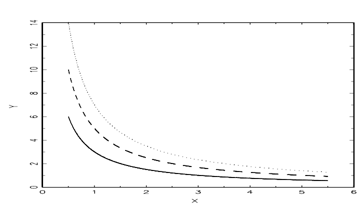
$$BIAS^*_{f_1}(x) = x_d^{\alpha_1} BIAS^*(u)$$

$$\Sigma_{f_1} = x_d^{2\alpha_1} \int \frac{p_V^2(s_2)}{p_W(u,s_2)} \frac{E\left(X_d^{2\alpha_1}\sigma_\varepsilon^2(W,X_d)|W=(u,s_2)\right)}{E^2\left(X_d^{2\alpha_1}|W=(u,s_2)\right)} ds_2.$$



Remark 5

- Undersmoothing in a nuisance direction, $h_2^2/h_1^2 \to 0,\ BIAS^*(u) = \frac{h_1^2}{2}\mu_K^2 tr\left(D^2\beta_1(u)\right).$
- For homoscedastic errors, the variance is $||K||_2^2\sigma_\varepsilon^2 \int \frac{p_V^2(s_2)}{p_W(u,s_2)} ds_2.$
- the same results from usual marginal integration in additive models with LLF as pilot estimate.



Application: livestock production function in Wisconsin

Data Set: Farm Credit Service of Saint Paul, Minnesota (1987)

the number of observations, $N = \mathbf{250}$

y: livestock

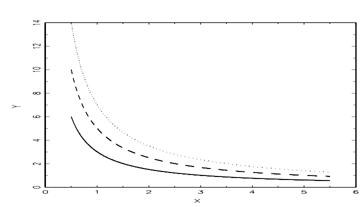
x: family labor

z_1: miscellaneous inputs (repairs, rent, supplies, gas, oil utilities)

z_2: intermediate assets

z_3: hired labor

z_4:animal inputs (purchased feed, breeding, veterinary services)



OLS based on **Cobb-Douglas**

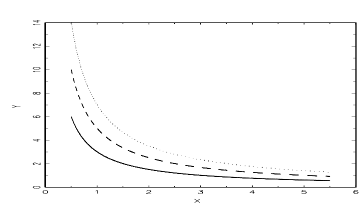
$$f(l) = c \prod_{i=1}^{5} l_i^{\beta_i}$$

$$\begin{aligned}
\widehat{\log y} &= \underset{(0.289)}{1.886} + \underset{(0.020)}{0.063} \log x + \underset{(0.025)}{0.289} \log z_1 \\
&\quad + \underset{(0.034)}{0.305} \log z_2 + \underset{(0.007)}{0.031} \log z_3 + \underset{(0.023)}{0.277} \log z_4, \\
R^2 &= 0.900 \\
\sum_{i=1}^{5} \widehat{\beta}_i &= 0.965.
\end{aligned}$$

At 1% level, we cannot reject the hypothesis that $\sum_{i=1}^{5} \beta_i = 1$, that is, cannot reject the hypothesis of CRS under a Cobb-Douglas specification.

Problems: the functional misspecification,

homogeneity only on 'variable input', not on 'fixed input'



- Nonparametric Modeling Assumption:

fixed variable: family labor(x)

variable input: other inputs($z_1, .., z_4$)

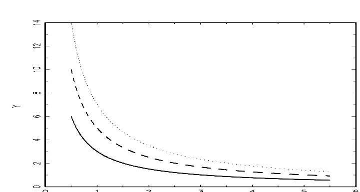
$$
\begin{aligned}
y &= f_1(x) + f_2(z) + \varepsilon \qquad \text{: additivity} \\
&= f_1(x) + z_4 f_2(z_1/z_4, z_2/z_4, z_3/z_4, 1) + \varepsilon \\
&: \quad \text{linear homogenity} \\
&= f_1(x) + z_4 g_2(w_1, w_2, w_3) + \varepsilon,\ w_i = z_i/z_{4.}
\end{aligned}
$$

* Severance-Lossin and Sperich (1997):

componentwise additivity

no interaction between bariable inputs

$$y = \sum_{i=1}^{5} h_i(l_i)$$



- Results : **elasticity of scale** measures the percent increase in output due to one percent increase in all inputs.

$$e(x,z) = \sum_{i=1}^{5} \frac{\partial \log f(l_i)}{\partial \log l_i}$$

1. Unrestricted Model

$$e(x,z) = \frac{x' \nabla f_x(x,z) + z' \nabla f_z(x,z)}{f(x,z)}$$

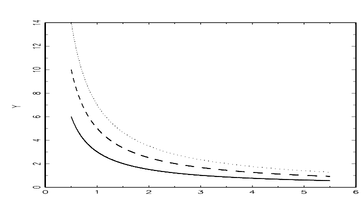
2. Restricted Model: with $f_2(z)$ homogeneous of degree r

$$e(x,z) = \frac{x' \nabla f_x(x,z) + r' f_2(z)}{f(x,z)},$$

by Euler's theorem

3. Parametric Cobb-Douglas

$$e(x,z) = \sum_{i=1}^{5} \beta_i$$



- Scale Elasticities for Livestock Production in Wisconsin Farms

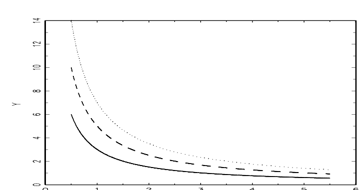
	(Full	Sample)	(Excl.	Outliers)
	Mean	Med.	Mean	Med
RM	1.067	1.018	1.060	1.016
UM	0.994	1.011	1.011	1.012
CD	0.965	(fixed)		

1. $\widehat{e}(x_i, z_i)$ fluctuates around 1

2. closeness of average or median scale elasticity between two models

$\Longrightarrow$indirect evidence for the validity of restriction

3. $\widehat{e}(x_i, z_i)$'s from the restricted model are more centered around 1 than those from the unrestricted, while they are fixed as $\sum_{i=1}^{5} \widehat{\beta}_i = 0.965$ under Cobb-Douglas.



Conclusion

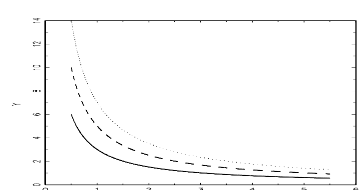
- Nonparametric Estimation of Additive Models with Homogeneous Components

nonparametric	:	flexibilty
additivity	:	reduction in Dimension
homogeneity	:	economic restriction

- Asymptotic Theory of Two-Step Estimators:

local linear fit	:	1st step
marginal integration	:	2nd step

properties :asymptotic normality, optimal convergence rate



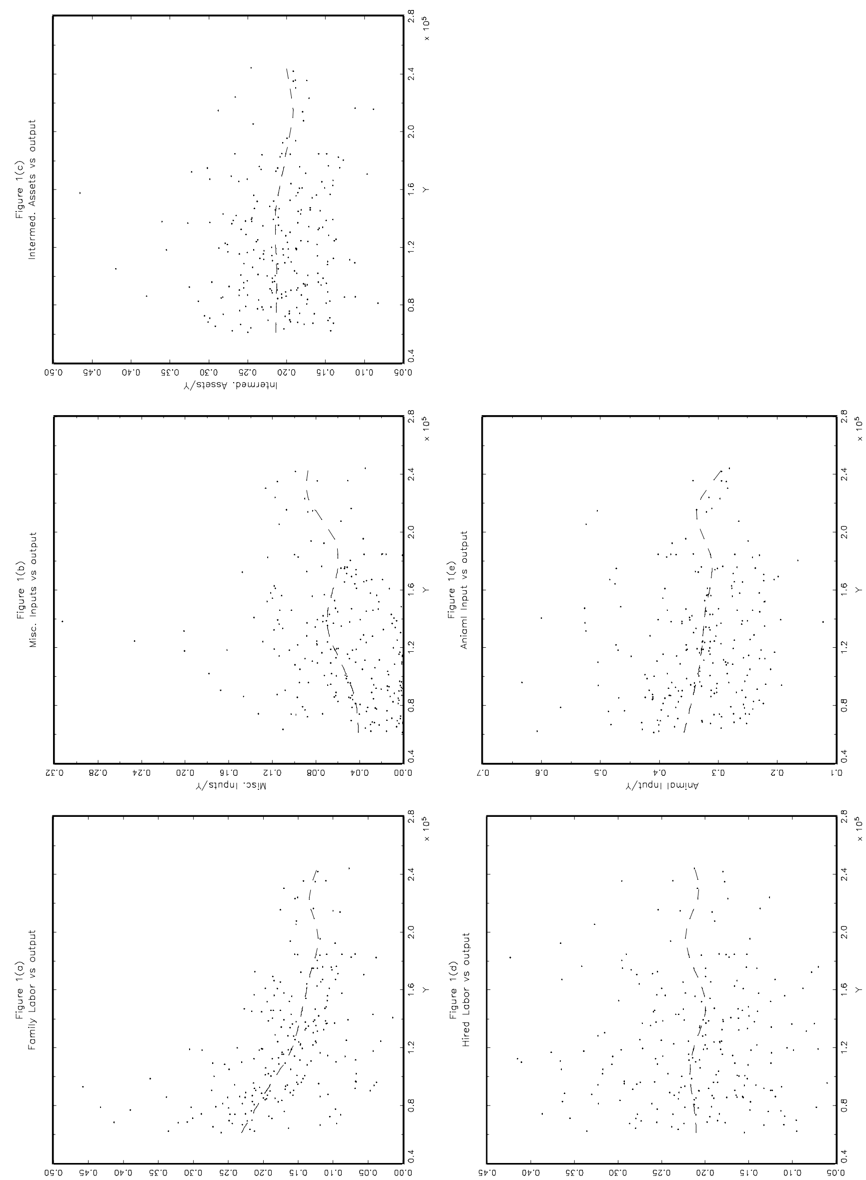


Figure 1(a)
Family Labor vs output
Family Labor/Y
Figure 1(b)
Misc. Inputs vs output
Misc. Inputs/Y
Figure 1(c)
Intermed. Assets vs output
Intermed. Assets/Y
Figure 1(d)
Hired Labor vs output
Hired Labor/Y
Figure 1(e)
Animal Input vs output
Animal Input/Y
Y

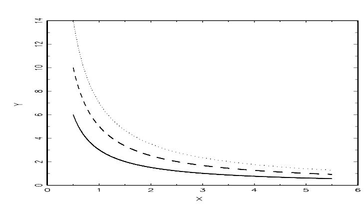


Figure 2(a)
Density for Scale Elasticity from Rest.

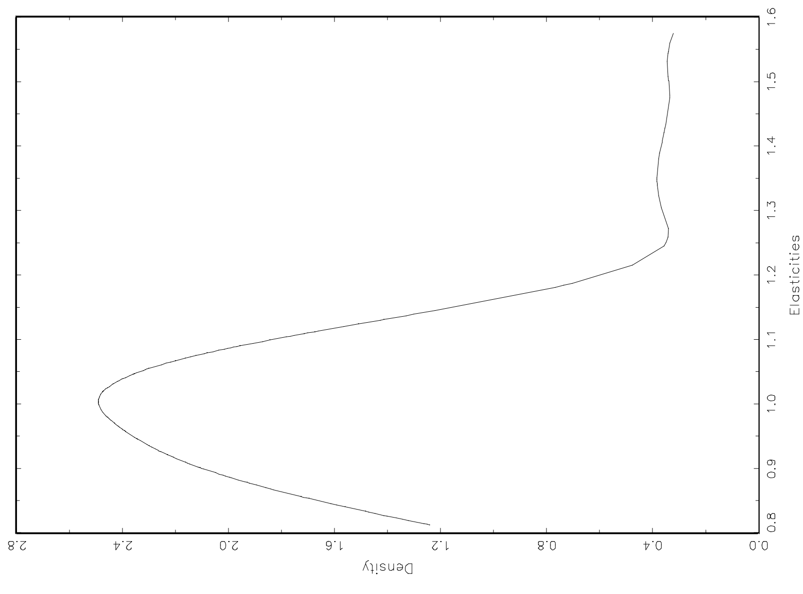


Figure 2(b)
Density for Scale Elasticity from UnRest.