# Nonparametric Estimation of Homothetic and Homothetically Separable Functions

**Arthur Lewbel** — Boston College

**Oliver Linton** — LSE

**Yale University, October, 2004**

Outline

1. Statement of the Problem
2. Literature Review
3. Estimation Idea in old version
4. A more general approach
5. Discussion on Distribution Theory
6. Simulations and Applications
7. Extensions

## The Problem

- A given function $r(x, w)$ has the following structure. There exist functions $h$ and $g$ such that
$$r(x, w) = h[g(x), w],$$
where
  - $g$ is linearly homogeneous, i.e., $g(cx) = cg(x)$ for $c \in \mathbb{R}$
  - $h$ is strictly monotonic in $g$, i.e., $\partial h / \partial g > 0$.
- Homotheticity:
$$r(x) = h[g(x)]$$
- Linear homogeneity vs any other nonzero known degree (or any known monotonic transformation of $g$) is WLOG.
- Goal: consistent and asymptotically normal estimator of $h$ and $g$ based on some estimator $\widehat{r}(x, w)$ of $r(x, z)$ when $h, g$ are unknown but continuous/smooth.

## Literature Review

- Homothetic and homothetically separable functions are common in models of consumer preferences and firm production.
- $r(x, w)$ could be a utility, cost function or production function, either directly estimated or recovered from consumer or factor demand equations.
- Examples: Blackorby, Primont, and Russell (1978), Chiang (1984), Zellner and Ryu (1998), Matzkin (1994). Zellner and Revankar (1969)

$$Ye^{\theta Y} = AK^{\alpha(1-\delta)}L^{\alpha\delta}$$

- Linear index models like standard censored, truncated, or discrete response models are homothetic functions, with $g(x) = x^\top\beta$. Replacing $x^\top\beta$ with an arbitrary linearly homogeneous function $g(x)$ is a natural generalization for contexts like price indices or constant returns to scale technologies.

Other Homogeneity related estimators

- Matzkin (1992) consistent estimator for
$$y = I[g(x) + \varepsilon \geq 0],$$
$g(x)$ homogeneous, $\varepsilon \perp\!\!\!\perp x$. Newey and Matzkin (1993) similar to ours, no $w$, more steps, incomplete.
- Matzkin (2003)
$$y = m(x, \varepsilon)$$
with $\varepsilon \perp\!\!\!\perp x$ and e.g., $m$ homogeneous in $x, \varepsilon$.
- Nonparametric homogeneous functions: Matzkin (1992), Tripathi and Kim (2001).
- Yatchew and Bos (1997) consider estimating some homothetic demand models by nonparametric least squares.

Other Separability related estimators

- Weak separability: $r(x,w) = h[g(x), w]$ without $g$ homogeneous. Gorman (1959), Goldman and Uzawa (1964), Blackorby, Primont, and Russell (1978). Pinkse (2001) estimates $g$ up to monotonic transformation.
- Strong or additive separability:
$$r(x,w) = g(x) + t(w).$$
Härdle, Kim, and Tripathi (2001), Friedman and Stutzle (1981), Breiman and Friedman (1985), Andrews (1991), Tjøstheim and Auestad (1994), Linton and Nielsen (1995), Stone (1986).
- Generalized additive separability:
$$r(x,w) = H\left(g(x) + t(w)\right)$$
for some known or unknown $H$. Hastie and Tibshirani (1990), Linton and Härdle (1996), Horowitz (2001).

## Estimation Idea
## Matching

- Normalize $g(x_0) = 1$.
- For a given $x, w$, find $u_{xw}$ such that
  $$r\left(x, w\right) = r(u_{xw}x_0, w),$$
  a match.
- Then
  $$g\left(x\right) = g(u_{xw}x_0) = u_{xw}g(x_0) = u_{xw},$$
  so
  $$g(x) = u_{xw}.$$

# More Matching

- By homogeneity, if know $g(vx)$, then know $g(x)$, because
$$g(x) = g(vx)/v \text{ for any } v.$$
- For each $v$, find the match $u_{xw}(v)$ such that
$$r\,(vx, w) = r(u_{xw}(v)x_0, w).$$
Then
$$g(x) = u_{xw}(v)/v.$$
- Alternatively, for each $v$, find the match $u_{xw}(v)$ such that
$$r\,(vx, w) = r(u_{xw}(v)vx_0, w).$$
Then
$$g(x) = u_{xw}(v).$$
- Alternatively, for each $v$ find the match $u_{xw}(v)$ such that
$$r\,(u_{xw}(v)x, w) = r(vx_0, w).$$
Then
$$g(x) = v/u_{xw}(v).$$

## Averaging

- Reduce dimension and speed convergence of estimator by averaging over $v$ and $w$:

$$g(x) = \int_{\Psi_{v,w|x}} \frac{u_{xw}(v)}{v} dF_{v,w|x}(v, w|x)$$

  where $F_{v,w|x}$ is any distribution function with support $\Psi_{v,w|x}$ such that
  - for all $(v, w) \in \Psi_{v,w|x}$, $(vx, w) \in \Psi_{x,w}$ and $(u_{xw}(v)x_0, w) \in \Psi_{x,w}$.
- Could instead have averaged $r$ over $w$ before matching.

## Identifying $h$

- The function $h$ could be identified by

$$r(\gamma x_0, w) = h(g(\gamma x_0), w) = h(\gamma g(x_0), w) = h(\gamma, w)$$

  but this would lead to a very inefficient estimator because it ignores all $x$'s except for those on the same ray from the origin as $x_0$.

- It will be better to recover the function $h$ given the estimated $g(x)$ by setting

$$h(\gamma, w) = \int r(x, w) 1(g(x) = \gamma) dx$$

- In practice it will be desirable to do local averaging, smoothing over the indicator

$$h_\delta(\gamma, w) = \int r(x, w) 1(g(x) \in [\gamma - \delta, \gamma + \delta]) dx$$

  for some small $\delta$.

- An alternative is to write

$$h(\gamma, w) = E[r(X, W) \mid g(X) = \gamma, W = w],$$

  which amounts to similar local averaging in practice.

## Where to locate?

- Don't need to find a match for every $x$, only for vectors of length one. Replace $x$ by $x/||x||$ everywhere above. By homogeneity, if know $g(x/||x||)$, then know $g(x)$, because
$$g(x) = ||x||g(x/||x||).$$
- Replacing $x$ by $x/||x||$ simplifies support considerations. Only need $\Psi_{v,w|x/||x||}$. If $\Psi_X$ contains a ball around zero, then can choose $F_{v,w|x}$ to be same for all $x$.
- Replacing $x$ by $x/||x||$ also imposes homogeneity on $g$.

Estimation Algorithm

1. For each $x$, draw $V_i, W_i$, for $i = 1, ..., n$ from some distribution $F_{v,w|x}$.
2. For each $V_i, W_i$, find the match $\widehat{U}_i$ such that $\widehat{U}_i = \arg\min_u \left[\widehat{r}(V_i x, W_i) - \widehat{r}(u x_0, W_i)\right]^2$.
3. Let
$$\widehat{g}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{U}_i}{V_i}.$$
4. Do this for $x = X_i$, $i = 1, \ldots, n$
5. For any $c = (\gamma, w)$ with $\gamma \in \{g(x) : x \in \Psi_X\}$, let $\widehat{h}(\gamma, w)$ be the intercept from a local polynomial regression of order $q$ of $\widehat{r}(Z_i)$ on $\widehat{C}_i = (\widehat{g}(X_i), W_i)$.

## A More General Approach
## Polar Coordinates

- Estimating $g(x/||x||)$ by averaging over $x$'s that lie on the same ray from the origin as the given $x$ can be conveniently expressed in polar coordinates. Write $x$ in polar coordinates as $\rho, \theta$, where $\rho$ is length and $\theta$ is direction, so $\theta$ contains the same information as $x/||x||$. Let $\theta_0$ denote the direction of $x_0$
- Define the functions $R$ and $G$ that are just the functions $r$ and $g$ expressed in polar coordinates

$$R(\rho, \theta, w) = r(x, w) \text{ and } G(\theta) = g(x/||x||),$$

- Any function $G$ automatically corresponds to a homogeneous function $g$, defined by

$$g(x) = \rho G(\theta).$$

- Without loss of generality take $x, x_0$ to have unit length, and for simplicity ignore $w$. Our estimation trick uses the fact that
  $R(v, \theta) = R(u_0 v, \theta_0) \Longrightarrow u_0 = G(\theta)$.
  This is true for all $v$. Replace $v \mapsto \rho$
- Can think of this as a moment condition with
  $M(u; \rho, \theta) = R(\rho, \theta) - R(u\rho, \theta_0)$
  so that
  $$M(u; \rho, \theta) = 0$$
  if and only if $u = u_0$. True for all $\rho$.
- We do this matching for each $\rho$ and then average the answer.

Quadratic Forms

- For a given $\theta$, we could estimate $u_0 = G(\theta)$ by minimizing the quadratic form
$$Q(u) = \int M(u;\rho,\theta)^2 f_\rho(\rho) d\rho$$
which is averaged over $\rho$. As long as $h$ is invertible on its first element for some positive measure subset of the support of $w$, this criterion function will have a unique zero at $u = u_0$.
- The first order condition is that
$$\int M(u_0;\rho,\theta_0)\frac{\partial R}{\partial \rho}(u_0\rho,\theta_0)\rho f_\rho(\rho) d\rho = 0$$
- i.e., you are matching on some integral of $R(\rho,\theta)$ over $\rho$
$$\int R(\rho,\theta)\frac{\partial R}{\partial \rho}(u_0\rho,\theta_0)\rho f_\rho(\rho) d\rho$$
$$= \int R(u_0\rho,\theta_0)\frac{\partial R}{\partial \rho}(u_0\rho,\theta_0)\rho f_\rho(\rho) d\rho.$$

- This can be taken one stage further. Let us define the function $u_0(\cdot)$ as the unique minimizer of

$$Q(u(\cdot)) = \int [M(u(\theta); \rho, \theta)]^2 V(\rho, \theta) f_{\rho,\theta}(\rho, \theta) d\rho d\theta,$$

  where

  – $V(\rho, \theta)$ is some weight function
  – $f_{\rho,\theta}(\rho, \theta)$ is the joint density of the random variables $\rho, \theta$.

- Should choose $V f_{\rho,\theta}$ to be the inverse variance of estimates of $R(\rho, \theta) - R(u_0\rho, \theta_0)$.

- Letting
$$u(\theta) = u_0(\theta) + \epsilon h(\theta),$$
where $h$ is any measurable function of $\theta$ and $\epsilon$ is a scalar, we have the first order condition

$$\begin{aligned} 0 &= \int [R(\rho,\theta) - R(u_0(\theta)\rho,\theta_0)] \frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0) \\ &\quad \times \rho V(\rho,\theta) h(\theta) f_{\rho,\theta}(\rho,\theta) d\rho d\theta \\ &= E\left[[R(\rho,\theta) - R(u_0(\theta)\rho,\theta_0)] \right. \\ &\quad \left. \times \rho \frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0) V(\rho,\theta) h(\theta)\right] \\ &= \int E\left[[R(\rho,\theta) - R(u_0(\theta)\rho,\theta_0)] \frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0) \right. \\ &\quad \left. \rho V(\rho,\theta) | \theta\right] \times h(\theta) f_\theta(\theta) d\theta \end{aligned}$$

by standard calculus of variations arguments, where $f_\theta(\theta)$ is the marginal density of the random variable $\theta$.

- Taking $h(\theta)$ to be the Dirac delta function $\delta_\theta(.)$, this implies that

$$E\left[R(\rho,\theta)\frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0)\rho V(\rho,\theta)|\theta\right]$$
$$= E\left[R(u_0(\theta)\rho,\theta_0)\frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0)\rho V(\rho,\theta)|\theta\right]$$

for all $\theta$. This is a nonlinear equation in $u_0(\theta)$, but it is 'separable' in $\theta$.

- In practice we would obtain

$$\pi(\rho,\theta)=\frac{\partial R}{\partial \rho}(u_0(\theta)\rho,\theta_0)\rho V(\rho,\theta)f_{\rho|\theta}(\rho|\theta)$$

from some preliminary estimation and then match by finding the $u_0(\theta)$ that solves

$$\int R(\rho,\theta)\pi(\rho,\theta)d\rho=\int R(u_0(\theta)\rho,\theta_0)\pi(\rho,\theta)d\rho$$

for each $\theta$.

- That is, you compute

$$m(u,\theta)=\int[R(\rho,\theta)-R(u\rho,\theta_0)]\pi(\rho,\theta)d\rho$$

for each $\theta$ and then find the value of $u$, $u_0(\theta)$, that zeros $m(u,\theta)$.

- The optimal weighting function $V(\rho,\theta)f_{\rho,\theta}(\rho,\theta)$ should be the inverse of the asymptotic variance of
$$\widehat{R}(\rho,\theta) - \widehat{R}(u_0(\theta)\rho,\theta_0)$$
and can be estimated as well.

- Basically, for each $\theta$ you are matching on some integral of $R$ over $\rho$ where the weighting is optimally determined from some preliminar procedure, just like GMM.

## Distribution Theory

- Pointwise asymptotic normality for $\widehat{g}$ and $\widehat{h}$ at rates $K_x - 1$ and $\max\{K_w + 1, K_x - 1\}$.
- The analysis for $\widehat{g}$ involves some arguments like that of estimates of mode of a density or regression function to derive the properties of the inverse function $\widehat{s}$
- Argument for $\widehat{h}$ is a very complicated generated regressors problem.
- Messy formulae, simplified by using polar coordinates for $x$; analytic standard errors.
- Assumptions on $\widehat{R}$, smoothness and support conditions.
- Performance comparison between the integrate first then invert versus invert first then integrate. Answer depends.

## Consistency for each $v$

- For each $v, w$ can think of as an optimization estimator

$$Q_n(u;v) = [M_n(u;v,\theta)]^2 = \left[\widehat{R}(v,\theta) - \widehat{R}(uv,\theta_0)\right]^2$$

  such that

$$\widehat{u}(v) \in \arg\min_{u \in U} Q_n(u;v)$$

  is an estimator of $G(\theta)$.

- Provided $\widehat{R}$ is continuous along the two rays there exists a minimizer, but it may not be unique for every sample.

- Consistency of $\widehat{u}(v)$ follows from uniform consistency of $\widehat{R}$ along the two rays

$$\sup_v |\widehat{R}(v,\theta) - R(v,\theta)| \; = \; o_p(1)$$

$$\sup_v |\widehat{R}(v,\theta_0) - R(v,\theta_0)| \; = \; o_p(1)$$

  along with the identifiable uniqueness coming from the strict monotonicity of $h$.

- Consistency of the average of $\widehat{u}(v)$ over $v$ follows under minimal conditions on the averaging measure.

Asymptotic Distribution for each $v$

- Recall
$$M_n(u; v, \theta) = \widehat{R}(v, \theta) - \widehat{R}(uv, \theta_0)$$
is the moment condition.

- By Taylor expanding around the true value in this case denoted $u_0(\theta)$, can show that
$$\widehat{u}_v(\theta) - u_0(\theta)$$
$$\simeq -\left[\frac{\partial M_n(u; v, \theta)}{\partial u}\right]^{-1}_{u=u_0(\theta)} [M_n(u_0(\theta); v, \theta)]$$
$$\simeq \left[\frac{\partial R(\rho, \theta_0)}{\partial \rho}\rho\right]^{-1}_{\rho=vu_0(\theta)} \times$$
$$\left[\widehat{R}(v, \theta) - R(v, \theta) - (\widehat{R}(u_0(\theta)v, \theta_0) - R(u_0(\theta)v, \theta_0))\right],$$

- Furthermore,
$$\left[\frac{\partial R(\rho, \theta_0)}{\partial \rho}\rho\right]_{\rho=vu_0(\theta)} = h'(vu_0(\theta))vu_0(\theta)$$
because by assumption $u_0(\theta_0) = 1$.

- For asymptotic normality need some additional conditions on $\widehat{R}(v,\theta)$. We assume they are like local polynomial polar co-ordinate regression estimators.

- Then we have

$$(nh^d)^{1/2}[\widehat{u}_v(\theta) - u_0(\theta) - b^2\beta(v,\theta)] \Longrightarrow N(0, s^2(v,\theta)),$$

where

$$\begin{aligned} s^2(v,\theta) &= ||K||^2 \frac{1}{[h'(vu_0(\theta))vu_0(\theta)]^2} \times \\ &\quad \left[\frac{\sigma^2(v,\theta)}{f_{\rho,\theta}(v,\theta)} + \frac{\sigma^2(u_0(\theta)v,\theta_0)}{f_{\rho,\theta}(u_0(\theta)v,\theta_0)}\right] \\ \beta(v,\theta) &= \frac{\mu_2(k)}{2} \frac{-1}{[h'(vu_0(\theta))vu_0(\theta)]} \times \\ &\quad [\nabla_2 R(v,\theta) - \nabla_2 R(u_0(\theta)v,\theta_0)]\,. \end{aligned}$$

- Here, $\nabla_2 R(v,\theta)$ is the trace of the Hessian evaluated at $v,\theta$.

Asymptotics for General Class of Estimators

- Suppose that
$$\widehat{m}(\widehat{u}(\theta), \theta) = 0$$
with
$$\widehat{m}(u, \theta) = \int [\widehat{R}(\rho, \theta) - \widehat{R}(u\rho, \theta_0)]\pi(\rho, \theta) d\rho$$
for some weight function $\pi$.

- Then by Taylor expansion
$$\begin{aligned} & \widehat{u}(\theta) - u_0(\theta) \\ \simeq & -\left[\int \frac{\partial R}{\partial \rho}(u_0(\theta)\rho, \theta_0)\rho\pi(\rho, \theta) d\rho\right]^{-1} \\ & \times \int [\widehat{R}(\rho, \theta) - \widehat{R}(u_0(\theta)\rho, \theta_0)]\pi(\rho, \theta) d\rho \end{aligned}$$

- This has the asymptotic distribution

$$(nh^{d-1})^{1/2}[\widehat{u}(\theta) - u_0(\theta) - b^2\boldsymbol{\beta}_\pi(\theta)] \Longrightarrow N(0, \mathbf{s}^2_\pi(\theta)),$$

where:

$$\boldsymbol{\beta}_\pi(\theta) = \frac{-\int \beta(\rho,\theta)\pi(\rho,\theta)d\rho}{\int h'(\rho u_0(\theta))\rho u_0(\theta)\pi(\rho,\theta)d\rho}$$

$$\mathbf{s}^2_\pi(\theta) = \frac{\int s^2(\rho,\theta)\pi^2(\rho,\theta)d\rho}{\left[\int h'(\rho u_0(\theta))\rho u_0(\theta)\pi(\rho,\theta)d\rho\right]^2}$$

$$s^2(\rho,\theta) = ||K||^2 \times \left[\frac{\sigma^2(\rho,\theta)}{f_{\rho,\theta}(\rho,\theta)} + \frac{\sigma^2(u_0(\theta)\rho,\theta_0)}{f_{\rho,\theta}(u_0(\theta)\rho,\theta_0)}\right]$$

$$\beta(\rho,\theta) = \frac{\mu_2(k)}{2}\left[\nabla_2 R(\rho,\theta) - \nabla_2 R(u_0(\theta)\rho,\theta_0)\right].$$

- Optimize with respect to $\pi$ from earlier, that is, take

$$\pi(\rho,\theta) = h'(\rho u_0(\theta))\rho u_0(\theta)V(\rho,\theta)f_{\rho|\theta}(\rho|\theta)$$

$$V(\rho,\theta)f_{\rho,\theta}(\rho,\theta) = 1/s^2(\rho,\theta)$$

- Then you get

$$\mathbf{s}^2_{\pi_{opt}}(\theta) = \frac{1}{\int [h'(\rho u_0(\theta))\rho u_0(\theta)]^2 s^{-2}(\rho,\theta)d\rho}$$

## Simulations

- The design is $g(x) = ||x||/\sqrt{2}$, $h(g) = \exp(g)$, and $y = h[g(x)] + \varepsilon$ so there is no $w$. The distribution of $X$ is uniform over the donut

  $$\Psi_X = \{x : 0.2 < ||x||^2 \leq 2\}.$$

  We take $V = ||X||$ and $x_0 = (1, 1)$, which makes $g(x_0) = 1$ and $\Psi_V = (0.08, \sqrt{2})$.
- local constant kernel regressions with a Gaussian kernel. 100 reps
- Four fit criteria: integrated mean squared error IMSE, integrated mean absolute error IMAE, pointwise mean squared error PMSE (at central point), and pointwise mean absolute error PMAE.

## Results

- Table 1 shows that for estimation of $g$ the largest bandwidth produces superior estimates in all designs, with error criteria reduced by roughly 1/2 to 1/4 relative to estimates based on the smallest bandwidth.
- When estimating $g$ using the large bandwidth, most criteria in most designs are approximately halved when the sample size is increased from 100 to 500 observations.
- Estimates of $h$, Table 2, are generally less accurate than the estimates of $g$.
- The best choice of bandwidth for $b$ varies across designs, but the differences in fit across different bandwidths is less pronounced for $h$ than for $g$.
- The improvement in the fit of $h$ when increasing the sample size from 100 to 500 varies across designs, with decreases in the error measures ranging from about 40% to 80%.

## Application to Nonparametric Production Function Estimation

- Let $y$ be the log output of a firm and $x$ be a vector of inputs, and suppose that
$$E(y|x) = r(x) = h[g(x)]$$
with linearly homogeneous $g$.
- A property of production that is empirically important is returns to scale, defined as
$$S(g) = \frac{\partial h(g)}{\partial \ln g}$$
- Other important properties are measures of substitutability of inputs, such as the technical rate of substitution and the elasticity of substitution. When $x$ consists of just two elements, for example, capital $K$ and labor $L$, then a simple measure of substitutability is
$$\alpha(K/L) = \frac{\partial \ln g\,(K/L, 1)}{\partial \ln(K/L)}$$
Note in interpreting this measure that $g\,(K/L, 1) = g\,(K, L)\,/L$.

- The substitutability measure $\alpha(K/L)$ equals a constant $\alpha$ when $g(x) = K^{\alpha}L^{1-\alpha}$, that is, when the production function $r(x)$ is a monotonic transformation of a Cobb Douglas, which is a common specification for homothetic production.
- Observations of chemical manufacturing firms in mainland China in two time periods, 1995 and 2001. For each firm, we observe
  - the net value of real fixed assets $K$
  - the number of employees $L$
  - $Y$ defined as the log of value-added real output.
- Output and capital are measured in thousands of Yuan converted to the base year 2000 using a general price deflator for the Chinese chemical industry. A total sample size of 1638 firms in 2001 and 1560 firms in 1995.

- We consider both nonparametric and parametric estimates of the production function $r(K, L)$. The parametric model we employ is a homothetic Translog production function, in which log output

$$Y = h[g(K, L)] + \epsilon$$

$$g(K, L) = \left(\frac{K}{L}\right)^{\alpha} L$$

$$h(g) = \beta_0 + \beta_1 \ln(g) + \beta_2 \ln(g)^2$$

- Fitting this model by nonlinear least squares in each of the years of data yields the parameter estimates reported in Table 3 (standard errors are in parentheses).

TABLE 3: Parametric Translog Estimates

| | $\alpha$ | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| 2001 Translog | 0.696 | 9.815 | 0.783 | 0.036 |
| | (0.043) | (0.031) | (0.028) | (0.012) |
| 1995 Translog | 0.478 | 9.585 | 0.961 | 0.045 |
| | (0.046) | (0.024) | (0.041) | (0.017) |

- Figures 2 and 3 show homothetic Translog and homothetic nonparametric estimates $\widehat{g}(K/L, 1)$ and $\widehat{h}(g)$ in 2001.
- Figure 3 also shows fits from a simple nonhomothetic kernel regression of $Y$ on $K, L$, that is, the initial unconstrained estimator of the function $r$.
- For simplicity, at each nonparametric estimation step we used ordinary kernel regressions with a normal kernel and bandwidth given by Silverman's rule.

- The nonparametric fits of $r$ and those of $h$ shown in Figure are quite similar, indicating that the imposition of homotheticity is reasonable for this data set.
- The nonparametric estimates of the functions $g$ and $h$ are roughly similar to the parametric Translog model estimates, but show quite a bit more curvature, departing most markedly from the parametric model for $g$ at low capital to labor ratios and from the model for $h(g)$ at low values of $g$.
- These differences are greatly magnified when one calculates the returns to scale $S(g)$ and the substitution measure $\alpha(K/L)$. For the Translog model,
  $S(g) = \beta_1 + 2\beta_2 \ln(g)$ and $\alpha(K/L) = \alpha$.
  For the nonparametric model we use the approximation
  $\widehat{S}(\widehat{g}_i) \approx [\widehat{h}(\widehat{g}_{i+1}) - \widehat{h}(\widehat{g}_{i-1})]/(\widehat{g}_{i+1} - \widehat{g}_{i-1})$
  after sorting the data by $\widehat{g}_i$ for each firm $i$, and similarly for $\widehat{\alpha}(K/L)$.

- Unlike the popular homothetic Translog model, which assumes $\alpha$ constant, the nonparametric estimates have $\alpha$ sharply increasing at low capital labor ratios and leveling off only at high levels. This result indicates likely inadequacies of the parametric model. The assumption of a constant $\alpha$ may be more reasonable for advanced economies like the United States, which tend to have higher capital labor.
- The models also differ in returns to scale $S(g)$. Both models imply similar returns to scale on average, but the parametric model has $S(g)$ mildly increasing, based on a small but statistically significant positive estimate of $\widehat{\beta}_2$. In contrast, the nonparametric estimates are roughly U shaped, with a majority of the data in the decreasing part. Given the substantial variability of the nonparametric $\widehat{S}$, it is difficult to draw conclusions about the dependence of $S$ on $g$.

- The estimates based on 1995 data are broadly similar to 2001. The major difference between the two years is that average returns to scale appear to have declined over time, from approximately constant returns with average $S$ near one in 1995, to decreasing returns with $S$ near 0.8 in 2001.
- This finding could be an artifact of substantial ownership reform during this period. Many larger firms in the Chinese chemical industry may still be state-owned in 2001, while many smaller enterprises were privatized after 1995 and so could have substantially restructured, thereby enhancing their productivity. Combining these into a single cross section might then create the appearance of decreasing returns on average.
- This could explain the overall difference in mean $S$ between the two years, but would explain the observed patterns in $S(g)$

within each year, though as noted above these departures of $S(g)$ from a constant are at best weakly estimated.

- Changes over time may more generally be due to changes in technology, demand, and other aspects of China's increasing economic liberalization and growth over this time period.

## Some Extensions
## Endogenous Regressors

- Assume
$$y = H[g(x), w, \varepsilon],$$
and elements of $X, W$ are endogenous, correlated with $\varepsilon$. Let
$$U = (X, W) - E(X, W \mid Z).$$
Then $\varepsilon \mid X, W, Z \backsim \varepsilon \mid U, Z$. Define
$$\Upsilon(X, W, U) = E(Y \mid X, W, U).$$
Assume that $\varepsilon \mid U, Z \backsim \varepsilon \mid U$. Then
$$\Upsilon(x, w, u) = h[g(x), w, u].$$

- Let $\widehat{U}$ be residuals from nonparametrically regressing $X, W$ on $Z$. Let $\widehat{\Upsilon}$ be a nonparametric regression of $Y$ on $X, W, \widehat{U}$. Then apply the homotheticity estimator to $\widehat{\Upsilon}$ to get $\widehat{g}$.

- Assumption $\varepsilon \mid U, Z \backsim \varepsilon \mid U$ is like control functions of Blundell,Powell (2000,1) nonparametric triangular system of Newey, Powell,Vella (1999), Imbens and Newey (2001), Chesher (2001).

## Another Model

- Now consider
$$r(v, z, w) = h[m(z) + v, w]$$
$m$ need not be homogeneous. This is the same as above taking $m(z) = \ln(G(\theta))$, $v = \ln(\rho)$ and $h = \exp(h)$ from the polar notation of our old model.

- Example: partly linear index models, reservation price and willingness to pay models such as $y = I[-m(x) + \varepsilon \leq v]$ where $v$ is the price.

- Also censored regression. Suppose that we observe $Y, X$ where
$$\begin{aligned} Y^* &= g(X) - \varepsilon \\ Y &= \max\{Y^*, 0\}. \end{aligned}$$
Then
$$\Pr(Y \leq y | X = x) = F_\varepsilon (y - g(x))$$
for all $y \geq 0$ and all $x$.

## Estimate the Scale

- Have defined a class of estimators for homothetic functions with known scale of homogeneity $\tau = 1$.
- Can also allow scale $\tau$ to be unknown, i.e.,
$$g(cx) = c^{\tau} g(x)$$
for some unknown $\tau$. In polar co-ordinates
$$r(\rho, \theta) = H(\rho^{\tau} G(\theta)).$$
- Algorithm is as before for each $\tau \in [\underline{\tau}, \overline{\tau}]$ except there is a final stage where you optimize over $\tau$ using some criterion. For regression function $r$ we could minimize
$$Q_n(\tau) = n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \widehat{H}_{\tau}(\rho_i^{\tau} \widehat{G}_{\tau}(\theta_i)) \right\}^2,$$
where $\widehat{H}_{\tau}, \widehat{G}_{\tau}$ are our estimates for a specific $\tau$.

TABLE 1: Monte Carlo fit criteria for $\widehat{g}(x)$

| $\frac{\sigma_x}{\sigma_x+\sigma_\varepsilon}$ | | 0.75 | | | 0.5 | | | 0.25 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | | 100 | 300 | 500 | 100 | 300 | 500 | 100 | 300 | 500 |
| IMSE | $b_1$ | 0.0494 | 0.0436 | 0.0416 | 0.1280 | 0.1006 | 0.0867 | 0.4004 | 0.3193 | 0.2797 |
| | $b_2$ | 0.0189 | 0.0138 | 0.0122 | 0.0462 | 0.0369 | 0.0322 | 0.1622 | 0.1184 | 0.1025 |
| | $b_3$ | 0.0172 | 0.0102 | 0.0090 | 0.0306 | 0.0225 | 0.0191 | 0.1021 | 0.0843 | 0.0661 |
| IMAE | $b_1$ | 0.0292 | 0.0251 | 0.0246 | 0.0775 | 0.0576 | 0.0555 | 0.2402 | 0.1978 | 0.1604 |
| | $b_2$ | 0.0116 | 0.0087 | 0.0077 | 0.0289 | 0.0224 | 0.0203 | 0.0952 | 0.0749 | 0.0661 |
| | $b_3$ | 0.0104 | 0.0063 | 0.0056 | 0.0186 | 0.0138 | 0.0118 | 0.0638 | 0.0521 | 0.0427 |
| PMSE | $b_1$ | 0.0487 | 0.0432 | 0.0244 | 0.1237 | 0.0855 | 0.0781 | 0.3976 | 0.3180 | 0.2204 |
| | $b_2$ | 0.0198 | 0.0133 | 0.0109 | 0.0492 | 0.0357 | 0.0271 | 0.1910 | 0.1152 | 0.1046 |
| | $b_3$ | 0.0183 | 0.0098 | 0.0077 | 0.0294 | 0.0213 | 0.0158 | 0.1023 | 0.7483 | 0.0592 |
| PMAE | $b_1$ | 0.0302 | 0.0270 | 0.0173 | 0.0779 | 0.0541 | 0.0473 | 0.2238 | 0.1972 | 0.1435 |
| | $b_2$ | 0.0116 | 0.0081 | 0.0066 | 0.0294 | 0.0215 | 0.0178 | 0.0979 | 0.0709 | 0.0632 |
| | $b_3$ | 0.0111 | 0.0060 | 0.0049 | 0.0177 | 0.0128 | 0.0103 | 0.0638 | 0.0587 | 0.0392 |

TABLE 2: Monte Carlo fit criteria for $\widehat{h}(g)$

| $\frac{\sigma_x}{\sigma_x+\sigma_\varepsilon}$ | | 0.75 | | | 0.5 | | | 0.25 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | | 100 | 300 | 500 | 100 | 300 | 500 | 100 | 300 | 500 |
| IMSE | $b_1$ | 0.1103 | 0.0953 | 0.0918 | 0.2140 | 0.1724 | 0.1711 | 0.4911 | 0.3807 | 0.3525 |
| | $b_2$ | 0.1284 | 0.0946 | 0.0828 | 0.1703 | 0.1279 | 0.1150 | 0.3597 | 0.2702 | 0.2481 |
| | $b_3$ | 0.2101 | 0.1551 | 0.1357 | 0.2242 | 0.1638 | 0.1442 | 0.3378 | 0.2645 | 0.2183 |
| IMAE | $b_1$ | 0.0765 | 0.0624 | 0.0579 | 0.1570 | 0.1218 | 0.1144 | 0.3740 | 0.2834 | 0.2581 |
| | $b_2$ | 0.0894 | 0.0603 | 0.0511 | 0.1275 | 0.0898 | 0.0801 | 0.2771 | 0.1991 | 0.1803 |
| | $b_3$ | 0.1552 | 0.1045 | 0.0884 | 0.1686 | 0.1150 | 0.0988 | 0.2614 | 0.1678 | 0.1625 |
| PMSE | $b_1$ | 0.0963 | 0.0706 | 0.0537 | 0.1972 | 0.1451 | 0.1440 | 0.4527 | 0.3500 | 0.3523 |
| | $b_2$ | 0.0908 | 0.0568 | 0.0418 | 0.1376 | 0.0913 | 0.0800 | 0.3310 | 0.2462 | 0.2146 |
| | $b_3$ | 0.1528 | 0.0968 | 0.0760 | 0.1733 | 0.1083 | 0.0826 | 0.3124 | 0.2214 | 0.1662 |
| PMAE | $b_1$ | 0.0678 | 0.0490 | 0.0397 | 0.1382 | 0.1057 | 0.0989 | 0.3558 | 0.2600 | 0.2597 |
| | $b_2$ | 0.0653 | 0.0329 | 0.0251 | 0.1116 | 0.0666 | 0.0593 | 0.2607 | 0.1917 | 0.1609 |
| | $b_3$ | 0.1088 | 0.0580 | 0.0479 | 0.1286 | 0.0708 | 0.0566 | 0.2457 | 0.1842 | 0.1201 |

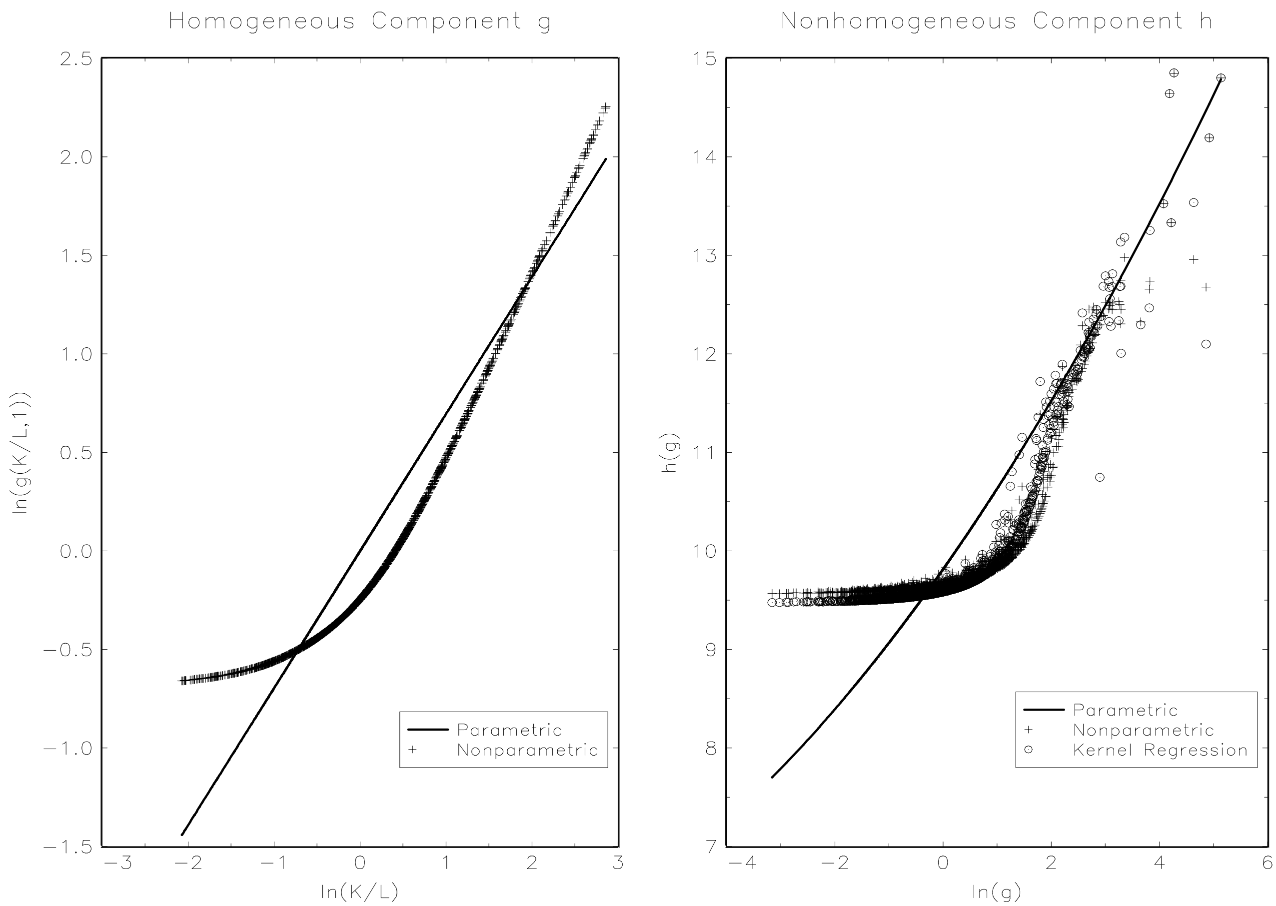
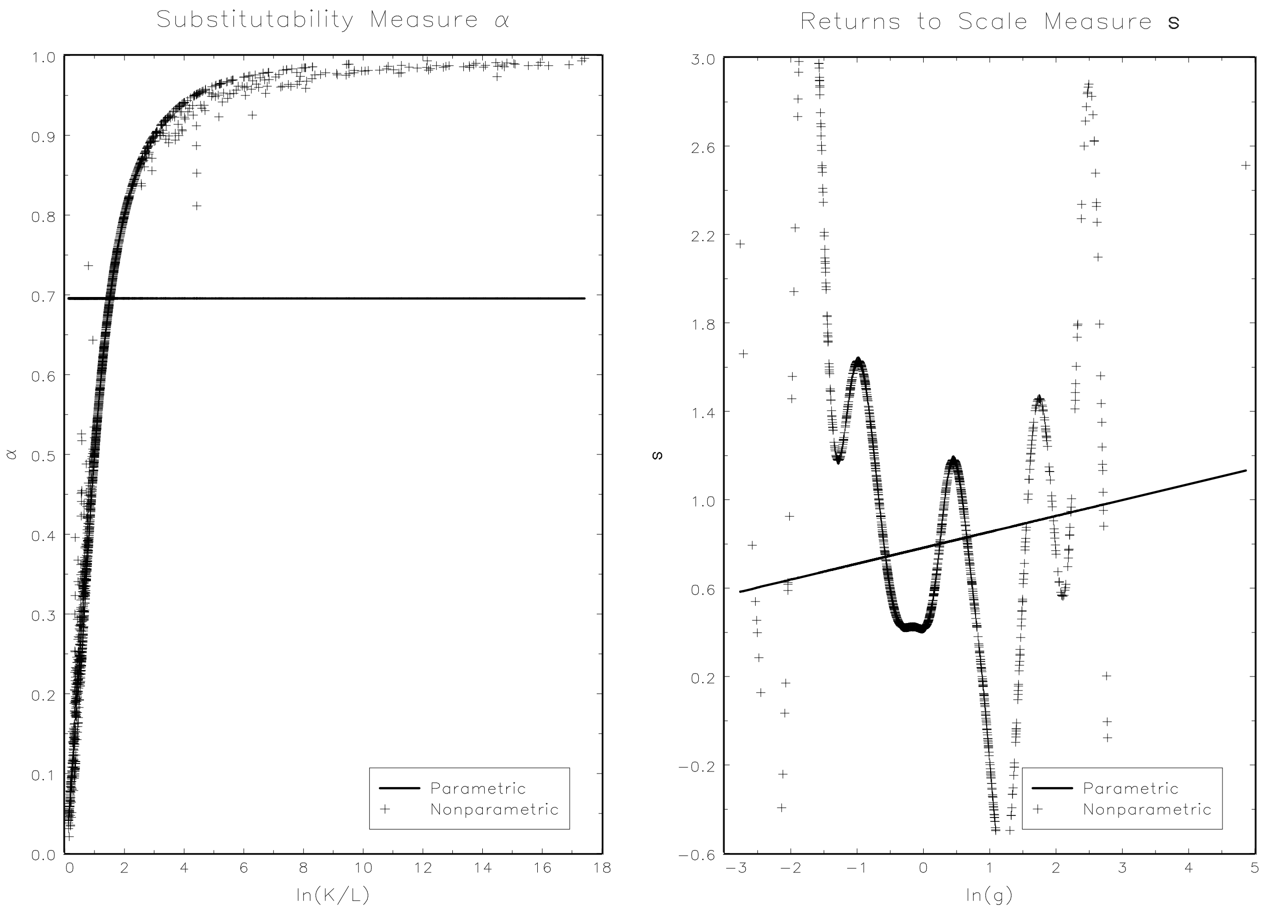

Homogeneous Component g
ln(g(K/L,1))
ln(K/L)
Parametric
Nonparametric
Nonhomogeneous Component h
h(g)
ln(g)
Parametric
Nonparametric
Kernel Regression

TABLE 4: Substitutability and Returns to Scale Estimates

| | $\alpha$ parametric | $\alpha$ nonparametric | $S$ parametric | $S$ nonparametric |
|---|---|---|---|---|
| mean 2001 | 0.696 | 0.537 | 0.788 | 0.821 |
| standard deviation 2001 | 0.000 | 0.281 | 0.082 | 1.286 |
| mean 1995 | 0.478 | 0.562 | 0.968 | 1.101 |
| standard deviation 1995 | 0.000 | 0.225 | 0.072 | 1.528 |


Substitutability Measure α
Returns to Scale Measure s
α
s
ln(K/L)
ln(g)
Parametric
Nonparametric
Parametric
Nonparametric