SOME ECONOMETRIC OPTIONS FOR DEALING WITH UNKNOWN FUNCTIONAL FORM[1]

C.J. O'Donnell
School of Economics, University of Queensland, Brisbane QLD 4072.
email: c.odonnell@economics.uq.edu.au

Version 13 February 2006

Abstract

Economic theory often provides information on the variables to be included in economic relationships (e.g., demands are functions of prices) and sometimes provides information on the signs and magnitudes of first- and second-order derivatives (e.g., homogeneity and concavity information). However, it rarely provides information concerning functional forms. In the absence of this information, it is common to assume a specific functional form (e.g., translog) and subsume errors of approximation into a disturbance term. Unfortunately, the estimated parameters of these approximating relationships do not consistently estimate the economically-relevant characteristics of the true relationship unless the latter is of the approximating class (White, 1980). Practical econometric solutions to the problem are now becoming available. This paper discusses kernel regression (KR), flexible least squares (FLS), generalized restricted least squares (GRLS) and latent class (LC) estimators. The empirical performance of all four estimators is assessed using an artificially-generated data set. Three of the estimators are then used to estimate characteristics of a labour demand function for US agriculture.

[1] Paper presented to the 50th Annual Conference of the Australian Agricultural and Resource Economics Society, 8-10 February, 2006.

1. INTRODUCTION

Consider an economic relationship of the form

(1) $\quad Y = m(Z) + E$

where Y is (the logarithm of) a random variable, Z is a $K \times 1$ vector of (logarithms of) non-stochastic exogenous variables, $m(.)$ is an unknown real-valued continuous function, and E is random variable representing the combined effects of all those factors that influence Y but have not been included in $m(Z)$. Most models of demand, supply, cost and profit appearing in the economics literature are models of this type. For purposes of inference it is common to assume E is independently normally distributed with zero mean and constant variance (the normality assumption is often justified using central limit theorems). The objects of inference are usually the conditional mean and first-derivatives of Y with respect to the elements of Z. If the variables are measured in logarithms then these first-order derivatives (gradients) are elasticities.

Estimation and inference is complicated by the fact that the mathematical form of the function $m(.)$ is generally unknown[2]. The usual solution is to assume a flexible functional form such as the generalized Leontief form of Diewert (1971) or the translog form of Christensen, Jorgenson and Lau (1973). These forms are flexible in the sense that they are capable of providing a local second-order approximation to an arbitrary twice-continuously differentiable function at a point. The fact that the approximation properties are local is a drawback of the approach, not least because the point of approximation is unknown and may not be represented in the data set (White, 1980).

Problems with flexible functional forms have led to interest in the nonparametric estimation approach. Let z be any fixed point that lies within the range of values of Z. The nonparametric approach to estimating the conditional mean $E(Y \mid Z = z) = m(z)$ involves taking a weighted average of Y values corresponding to Z values in the neighbourhood of z. Much of the nonparametric literature is concerned with choosing between weighting functions that can provide local or global approximations to the conditional mean (at the point z). The

---

[2] There are some instances where the functional form is known. For example, totally differentiating a function $Q = m(Z_1, ..., Z_K)$ and dividing through by Q yields a linear-in-the-parameters equilibrium displacement model, $Y = \eta_1 X_1 + \eta_2 X_2 + ... + \eta_K X_K$, where $\eta_k \equiv \partial \ln Q / \partial \ln Z_k$ denotes the elasticity of Q with respect to Z_k and $X_k \equiv d \ln Z_k = dZ_k / Z_k$ denotes the percentage change in Z_k. Other examples can be found in the pure theory of international trade.

major drawback of the approach is that very large samples are typically required to obtain reasonable estimates of the conditional mean (or the gradients). Part of the problem is what Silverman (1986) calls the empty-space phenomenon, where few points lie around the mean of a joint probability density when the number of variables is large[3]. The problem is sometimes referred to as the "curse of dimensionality".

This paper considers an alternative estimation approach that includes the flexible functional form and nonparametric approaches as limiting special cases. To motivate the approach, we note that most of the flexible functional forms used in the economics literature can be interpreted as second-order Taylor's series approximations to an unknown function at a single point (i.e., local approximations). To endow such models with greater flexibility, we effectively take Taylor's series expansions of the unknown function at each of the N points represented in the data set. The resulting model is an observation-varying parameter model in which the parameters are non-stochastic. Two econometric estimators are available for models of this type. If we have time-series data then parameter estimates can be computed using the flexible least squares estimator of Kalaba and Tefatsion (1989). This estimator finds time paths for the parameters that minimize an "incompatibility cost" function. Alternatively, and irrespective of the type of data that has been collected, a latent class estimator is available. The latent class approach involves dividing the data set into $J \leq N$ subsets, or classes, and estimating a separate flexible functional form model for each class. Thus, the latent class estimator offers a compromise between the conventional flexible functional form approach (corresponding to $J = 1$) and the nonparametric approach (corresponding to $J = N$).

The paper is essentially divided into two parts. The first part explains the nonparametric (NP), conventional flexible functional form (FFF), flexible least squares (FLS) and latent class (LC) estimators of the general model given by (1). The discussion of nonparametric estimators (Section 2) draws from Pagan and Ullah (1999) and references therein. The discussion of the conventional flexible functional form estimator (Section 3) is used to motivate an observation-varying parameter model that can be estimated using flexible least squares and latent class estimators (Section 4). An important contribution of the paper is to show how these different estimators should be modified to impose linear equality constraints involving the first-order derivatives of (1) (Section 5). Among other things, this part of the paper shows that the restricted LC estimator does not generally collapse to the restricted FFF estimator when the number of classes is $J = 1$. Rather, it collapses to the

[3] For example, less than 2% of the mass of a 10-dimensional standard normal distribution lies at points that are within a distance of 1.6 of the origin, compared to about 90% in the univariate case.

Generalised Restricted Least Squares (GRLS) estimator of Doran, O'Donnell and Rambaldi (2003).

The second part of the paper assesses the performance of the various estimators using artificially-generated time-series data (Section 6) and U.S. Department of Agriculture (USDA) panel data on aggregate labour usage in U.S. agriculture (Section 7). The simulated data is constructed using the data generating process previously used by White (1980) and Byron and Bera (1983) to investigate the approximation properties of the ordinary least squares (OLS) estimator. The simulation experiment reveals that the (conventional) restricted FFF estimator tends to outperform alternative estimators when evaluating characteristics of the economic relationship at the variable means. However, matters are not so straightforward when evaluating quantities of interest at other points represented in the data set.

2. NONPARAMETRIC ESTIMATION

The econometric model corresponding to (1) is

(2) $$y_i = m(z_i) + e_i$$

(3) $$e_i \sim iidN\left(0, \sigma^2\right)$$

where y_i and z_i denote the i-th observations on Y and Z, and e_i is the corresponding error term (i = 1, …, N). The primary object of inference is the mean of Y conditional on $Z = z$, formally defined as

(4) $$E(Y \mid Z = z) = m(z) = \frac{\int y\, p(y,z)dy}{p(z)} = \frac{q(z)}{p(z)}$$

where $p(y,z)$ is the joint probability density function (pdf) of Y and Z evaluated at $Y = y$ and $Z = z$, and $p(z)$ is the marginal pdf of Z evaluated at $Z = z$. We are also interested in the gradients, defined as

(5) $$m_k(z) = p(z)^{-1}\left[q_k(z) - p_k(z)m(z)\right] \qquad \text{for } k = 1, \ldots, K,$$

where $q_k(z)$ and $p_k(z)$ denote the first-order derivatives of $q(z)$ and $p(z)$ with respect to the k-th element of z.

The nonparametric approach to estimating the conditional mean (and, subsequently, the gradients) essentially involves replacing $p(y,z)$ and $p(z)$ in (4) with estimators, and then evaluating the resulting integral.

Kernel Density Estimators

Consider the problem of estimating $p(z)$ when $K = 1$. Suppose we use our observations on Z to construct a simple histogram with I bins, such as that depicted[4] in figure 1. The rectangles in figure 1 have areas that sum to $f_1 h + \ldots + f_I h = Nh$, where f_i is the frequency count in the i-th bin and h is the bin width, or bandwidth. Thus, dividing each frequency count by Nh yields rectangles with areas that sum (integrate) to one. The heights of these scaled rectangles can therefore be used as density estimates. Formally, the local histogram estimator of the density at $Z = z$ is

$$(6) \qquad \hat{p}(z) = \frac{1}{Nh}\sum_{i=1}^{N} I\left(-½ < \frac{z_i - z}{h} < ½\right)$$

where $I(.)$ is an indicator function that takes the value one when the argument is true and zero otherwise.

The local histogram estimator (6) assigns each observation a weight of one if it lies within half a bin width of z, otherwise it gets a weight of zero. Unfortunately, this weighting system yields an estimated density function that is neither smooth nor continuous. Rosenblatt (1956) overcame the problem by replacing the indicator function with a continuous kernel[5]. Formally, the kernel density estimator is

$$(7) \qquad \hat{p}(z) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{z_i - z}{h}\right)$$

where $K(.)$ is an arbitrary kernel. A common choice of kernel is the standard normal probability density function (pdf)[6].

---

[4] The histogram in figure 1 has been constructed using $N = 200$ observations on $Z \sim 0.6N(0,1) + 0.4N(4,4)$. The sample standard deviation is $s = 2.619$.

[5] A real positive function that is symmetric around zero and integrates to one.

[6] If the uniform kernel $K(z) = f_U(z \mid -½, ½)$ is chosen then (7) collapses to (6). Thus, the local histogram estimator can also be viewed as a kernel density estimator.

The availability of multivariate pdfs means the kernel density estimator (7) is also available when $K > 1$. For example, the multivariate density $p(y, z)$ can be estimated using

$$(8) \qquad \hat{p}(y,z) = \frac{1}{Nh}\sum_{i=1}^{N} G\left(\frac{y_i - y}{h}, \frac{z_i - z}{h}\right)$$

where $G(.)$ is a multivariate kernel with the property $\int G(y,z)\,dy = K(z)$.

Kernel Regression Estimators

The nonparametric approach to estimating $E(Y \mid Z = z)$ involves taking a weighted average of all Y values corresponding to Z values in the neighbourhood of z. To formally derive such an estimator, Pagan and Ullah (1999, p.83) use the estimated joint density (8) to derive the following kernel estimator of the numerator in (4):

$$(9) \qquad \hat{q}(z) = \frac{1}{Nh}\sum_{i=1}^{N} y_i K\left(\frac{z_i - z}{h}\right)$$

Substituting (7) and (9) for $p(z)$ and $q(z)$ in (4) yields the Nadaraya-Watson kernel regression estimator of the conditional mean:

$$(10) \qquad \hat{m}(z) = \frac{\hat{q}(z)}{\hat{p}(z)} = \sum_{i=1}^{N} w_i y_i$$

where

$$(11) \qquad w_i \equiv K\left(\frac{z_i - z}{h}\right) \Bigg/ \sum_{i=1}^{N} K\left(\frac{z_i - z}{h}\right)$$

is a weight that increases as z_i approaches z. A natural estimator of the first-order derivative (5) is

$$(12) \qquad \hat{m}_k(z) = \hat{p}(z)^{-1}\left[\hat{q}_k(z) - \hat{p}_k(z)\hat{m}(z)\right]$$

where $\hat{p}_k(z)$ and $\hat{q}_k(z)$ denote the first-order derivatives of $\hat{p}(z)$ and $\hat{q}(z)$ with respect to the k-th variable.

Selecting the Bandwidth and the Kernel

The problem of selecting the bandwidth and kernel in the regression estimator (10) and (11) is equivalent to the problem of selecting the bandwidth and kernel in the density estimator (7). Common approaches to this problem are based on knowledge of the local properties of kernel density estimators. Again, it is useful to begin with the case where $K = 1$.

When statisticians talk about the local properties of density estimators they are referring to the accuracy with which $\hat{p}(z)$ estimates $p(z)$ (i.e., the accuracy of the estimator at the point $Z = z$). Under mild regularity conditions, $\hat{p}(z)$ is consistent and asymptotically normal. However, the estimator is generally[7] biased in finite samples. If we choose a very small bandwidth then the bias is small but the variance of the estimator is large – pictorially, the estimated density function is very rough. Conversely, if we choose a very large bandwidth then the bias is large but the variance is small – the estimated density is overly smooth. A compromise is to choose the bandwidth that minimizes approximate mean integrated square error (AMISE),

(13) $$AMISE = \frac{1}{4}\lambda_1 h^4 + \lambda_2 (Nh)^{-1}$$

where λ_1 and λ_2 are known functions of the kernel and the second derivative (curvature) of the true (unknown) density. The AMISE is clearly minimized when $h = (N\lambda_2 / \lambda_1)^{-1/5}$. Thus, the AMISE-minimising bandwidth also depends on the kernel and the true density. For example, if i) the kernel is a standard normal pdf, and ii) the true density is normal, then the AMISE-minimising bandwidth is $h = 1.059 sN^{-1/5}$, where s is the sample standard deviation. There is evidence that this choice of bandwidth and kernel also tends to perform well when the true density is not normal, except perhaps when it is heavily bimodal or skewed (see Silverman, 1986).

We can also select an optimal kernel by minimising AMISE. We simply substitute our expression for the optimal bandwidth, $h = (N\lambda_2 / \lambda_1)^{-1/5}$, back into (13) and minimise the result with respect to $K(.)$, subject to the constraint that $K(.)$ has all the properties of a kernel. The solution is the Epanechnikov (1969) kernel:

(14) $$K(z) = \begin{cases} 0.75(1 - z^2) & \text{if } |z| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

[7] The bias is negligible if the unknown density is linear in the neighbourhood of $Z = z$.

In practice, most kernels appear to perform nearly as well as the Epanechnikov kernel (see Silverman, 1986) so the choice of kernel is often made on the grounds of computational convenience.

All of these results for univariate densities generalise to the multivariate case. When $K > 1$, the AMISE-minimising value of h is $h = cN^{-1/(4+K)}$ where c again depends on the kernel and the true density. For example, if i) the kernel is multivariate standard normal, ii) the true density is multivariate normal, and iii) all K variables have been scaled to have unit variance[8], then $c = [4/(2K+1)]^{1/(K+4)}$. The AMISE-minimising choice of kernel is a multivariate Epanechnikov kernel. Again, the choice of kernel is often made on the grounds of computational convenience, with attention focusing on kernels that can be written as the product of K univariate densities. Epanechnikov (1969) shows that the kernel that minimises AMISE over the class of so-called "product kernels" is the multivariate standard normal pdf. If this kernel is used then it is common to set the bandwidth for each univariate standard normal kernel to $h_k = 1.059 s_k N^{-1/5}$, where s_k is the sample standard deviation of the k-th variable.

Standard Errors

To compute standard errors, note that the y_i in (10) are independently distributed and the weights are nonstochastic. Thus, the variance of the estimated conditional mean is

$$(15) \qquad Var\left(\hat{m}(z)\right) = \sum_{i=1}^{N} w_i^2 Var(y_i) = \sum_{i=1}^{N} w_i^2 \sigma^2 .$$

A consistent estimator of σ^2 is (e.g., Pagan and Ullah, p. 181)

$$(16) \qquad \hat{\sigma}^2 = \sum_{i=1}^{N} w_i \left(y_i - \hat{m}(z_i)\right)^2 .$$

The elements of the covariance matrix of the first derivatives can be estimated using

$$(17) \qquad Cov\left(\hat{m}_k(z), \hat{m}_j(z)\right) = \frac{\hat{p}_k(z)\hat{p}_j(z)}{\hat{p}(z)^2} Var\left(\hat{m}(z)\right).$$

[8] The kernel method may not be scale-invariant, so it is common to standardize the data in this way.

Practical Issues

Kernel density (and associated regression) estimators are not without their problems. A minor problem is that the bandwidth is independent of the point at which the density is evaluated, sometimes resulting in too much smoothing over some data points and too little smoothing over others. To make the bandwidth depend on the data density in a neighbourhood of an evaluation point *z*, it is common to let $d_1(z) \leq ... \leq d_N(z)$ denote the distances (arranged in ascending order) between *z* and each of the *N* data points. Replacing *h* in (7) with $2d_k(z)$ yields the *k*-th nearest-neighbour density estimator. Unfortunately, this estimator does not integrate to one, so the associated estimated density function is not a proper pdf. Moreover, even though the estimated density function is continuous, it is not smooth.

A more serious problem with kernel density estimation is that very large samples are often required to obtain reasonable estimates of multivariate densities. Silverman (1986) reports a case where a sample of size 842,000 was needed to estimate *p*(0) to a desired degree of accuracy when *K* = 10, compared to a sample of size 4 when *K* = 1. In the remainder of this paper we consider estimation approaches that allow us to avoid this so-called curse of dimensionality.

Finally, some popular econometric software packages, including EViews and SHAZAM, have options for nonparametric estimation of econometric models. Unfortunately, EViews does not currently provide for estimation of multiple regression models, while SHAZAM does not compute standard errors for nonparametric estimators.

3. FIXED-PARAMETER FLEXIBLE FUNCTIONAL FORMS

Two methods for approximating unknown functions that are used in the empirical literature are Taylor's series approximations and members of the class of Fourier series expansions[9]. In this section we consider common flexible functional form models that can be derived as Taylor's series expansions. Details concerning Fourier expansions can be found in Gallant (1981) and Gallant (1984).

Let $z_i = (z_{1i}, ..., z_{Ki})'$ denote the *i*-th observation on *Z*. A second-order Taylor's series expansion of $m(z_i)$ around the point z is (e.g., Greene, 2003, p.838):

$$(18) \qquad m(z_i) = \alpha + \sum_{k=1}^{K} \phi_k z_{ki} + 0.5 \sum_{k=1}^{K} \sum_{j=1}^{K} \gamma_{kj} z_{ki} z_{ji} + a(z_i)$$

where

[9] Fourier series expansions include sine/cosine expansions as well as Hermite and Laguerre expansions.

(19) $\alpha = m(z) - g(z)'z + 0.5z'H(z)z$

(20) $\phi = [\phi_k] = g(z) - H(z)z$

(21) $\Gamma = [\gamma_{jk}] = H(z).$

and $g(z)$ and $H(z)$ are the $K \times 1$ gradient vector and $K \times K$ Hessian matrix evaluated at $Z = z.$ The last term on the right-hand side of (18) is an approximation error with the property[10] $\lim_{z_i \to z} a(z_i) = 0.$ The gradients at the point $Z = z_i$ are:

(22) $m_k(z_i) = \phi_k + \sum_{j=1}^{K} \gamma_{kj} z_{ji} + a_k(z_i)$

where $a_k(z_i) = \partial a(z_i)/\partial z_{ki}.$ Substituting (18) into (2) yields the familiar fixed parameter flexible functional form econometric model

(23) $y_i = x_i'\beta + u_i$

(24) $u_i \sim iidN\left(a(z_i), \sigma^2\right)$

where the vectors

(25) $x_i = \begin{bmatrix} 1 & z_{Ki} & \cdots & z_{Ki} & 0.5z_{1i}^2 & z_{1i}z_{2i} & \cdots & z_{1i}z_{Ki} & 0.5z_{2i}^2 & z_{2i}z_{3i} & \cdots & z_{Ki}z_{Ki} \end{bmatrix}'$

and

(26) $\beta = \begin{bmatrix} \alpha & \phi_1 & \cdots & \phi_K & \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{KK} \end{bmatrix}'$

are of row dimension $K^* = 1 + 1.5K + 0.5K^2$, and $u_i = a(z_i) + e_i$ is a composite error term.

The problem with this model is that, unless all approximation errors are zero, the disturbance term is correlated with the explanatory variables and has an unknown observation-varying mean. The usual way forward is to implicitly assume

(27) $a(z_i) = a^* \Rightarrow a_k(z_i) = 0 \Rightarrow m_k(z_i) = \phi_k + \sum_{j=1}^{K} \gamma_{kj} z_{ji}$ for $i = 1, \ldots, N.$

[10] The approximation error is also zero if all third- and higher-order derivatives of the unknown function are zero.

Under this assumption, the usual menu of econometric estimators is available. The possibility that $a^* \neq 0$ is of no practical consequence because a fixed non-zero mean can be subsumed into the intercept term (see, for example, Greene, 2003, p.14). Since the error is normally distributed by assumption, the OLS estimator is also the maximum likelihood (ML) estimator.

The properties of these conventional estimators depend critically on whether the assumption $a(z_i) = a^*$ is valid. If not, they are generally biased and inconsistent. – for details concerning the OLS estimator, see White (1980) and Byron and Bera (1983). One solution is to subsume $a(z_i)$ into the deterministic part of the regression model, resulting in a model in which the parameters are observation-varying. Another motivation for an observation-varying parameter model involves taking Taylor's series expansions at points represented in the data set.

4. OBSERVATION-VARYING PARAMETER MODELS

If the fixed parameter regression model given by (23) and (24) is to provide a second-order approximation to the unknown function at the observed point z_i then equations (19) to (21) imply

(28) $\alpha = m(z_i) - g(z_i)'z_i + 0.5z_i'H(z_i)z_i \equiv \alpha_i$

(29) $\phi = g(z_i) - H(z_i)z_i \equiv \left[\phi_{ki}\right]$ and

(30) $\Gamma = H(z_i) \equiv \left[\gamma_{jki}\right]$

Moreover, the limiting property of the approximation error implies $a(z_i) = 0$. Thus, by permitting the model to approximate the unknown function at every point in the data set, the gradients (22) become

(31) $m_k(z_i) = \phi_{ki} + \sum_{j=1}^{K} \gamma_{kji} z_{ji}$

and the econometric model takes the form

(32) $y_i = x_i'\beta_i + e_i$

(3) $e_i \sim iidN\left(0, \sigma^2\right)$

where

$$(33)\qquad \beta_i = [\; \alpha_i \quad \phi_{1i} \quad \cdots \quad \phi_{Ki} \quad \gamma_{11i} \quad \gamma_{12i} \quad \cdots \quad \gamma_{1Ki} \quad \gamma_{22i} \quad \gamma_{23i} \quad \cdots \quad \gamma_{KKi} \;]'.$$

The potential for this model to approximate the unknown function at every point in the data set is illustrated in figures 2 and 3. Figure 2 depicts two linear functions providing first-order approximations to an (unknown) nonlinear function at two points. Panels (b) to (d) in figure 3 depict three linear surfaces that approximate the (unknown) nonlinear surface depicted in panel (a).

The model given by (32) and (3) differs from the fixed-parameter model given by (23) and (24) in two respects. First, the error term is well-behaved. Second, every parameter is observation-varying. This second feature of the model is problematic because there are only N observations with which to estimate NK^* parameters. Estimation of the model in a random coefficients framework is inappropriate because the unknown parameters are nonstochastic. Instead, we must use an estimator that imposes some structure on the β_i. In this section we consider a flexible least squares estimator and a latent class estimator.

Flexible Least Squares (FLS)

If time-series data are available then there is a natural ordering of the observations and the model can be estimated using the FLS estimator of Kalaba and Tesfatsion (1989). This estimator finds time paths for the parameters that minimize the "incompatibility cost" function

$$(34)\qquad c(\beta_1,\ldots,\beta_N,\delta) = \frac{1}{1-\delta}\left[\delta\sum_{i=1}^{N-1}(\beta_{i+1}-\beta_i)'(\beta_{i+1}-\beta_i)' + (1-\delta)\sum_{i=1}^{N}(y_i - x_i'\beta_i)^2\right]$$

where $0 < \delta < 1$ is a smoothness parameter. Relatively high values of the smoothness parameter have the effect of placing relatively high weight on holding the parameters constant and less weight on minimizing the sum of squared measurement errors. Indeed, it can be seen from (34) that FLS → OLS as $\delta \to 1$.

Practical implementation of the FLS approach involves choosing a value for the smoothness weight. Unfortunately, the literature provides little guidance in this regard. It is common practice to set $\delta = 0.5$, thus giving equal weight to the dynamic evolution of the β_is and the residual error. There are two other difficulties with FLS that limit the usefulness of the approach. First, no closed form solutions are available for the covariance matrix of the

FLS estimator – the few econometrics software packages that compute FLS estimates do not provide estimated standard errors (e.g., SHAZAM). Second, and perhaps more importantly, the approach cannot be used for making out-of-sample predictions.

Latent Class (LC) Estimators

An alternative estimation approach that is more widely applicable involves dividing the set of data points into $J \leq N$ subsets, or classes. Specifically, we identify subsets $S_1, ..., S_J$ with the property that $\beta_i = \delta_j$ for all $y_i \in S_j$. Then the observation-varying parameter model becomes

(35) $y_i = x_i'\delta_j + e_i$ for $i \in S_j$ and

(3) $e_i \sim iidN\left(0, \sigma^2\right)$.

If the data points can be assigned to classes *a priori* then this model can be estimated using conventional techniques: if J = 1 an appropriate estimator is the ML estimator; if J = N a suitable estimator is the kernel regression estimator; and if $1 < J < N / K^*$ the model can be estimated within a conventional seemingly unrelated regression framework. Unfortunately, the observations can rarely be classified into subsets *a priori*, so these conventional estimators are unavailable. Instead, the model must be estimated in a mixtures, or latent class, framework. The use of the term "mixtures" derives from the fact that the random variable y_i has a distribution that is a mixture of normal random variables having the same variance but different means. The dependent variable in (35) is a mixture of a finite number of normal random variables, so the model is an example of a "finite mixture of normals" model.

Mixtures models can be estimated by the method of maximum likelihood. The log-likelihood for the observed sample is

(36) $$\ln L = \sum_{i=1}^{N} \ln p(y_i \mid z_i, \sigma^2) = \sum_{i=1}^{N} \ln \left\{ \sum_{j=1}^{J} \pi_{ij} \, p(y_i \mid z_i, \delta_j, \sigma^2) \right\}$$

where $\pi_{ij} = \Pr(y_i \in S_j)$ is the (unobserved) prior probability that the random variable y_i belongs to the j-th class. For estimation purposes, it is common to specify this probability to be of the multinomial logit form:

$$\pi_{ij} = \frac{\exp(x_i'\eta_j)}{\sum_{m=1}^{J}\exp(x_i'\eta_m)} \tag{37}$$

where η_j is a $K^* \times 1$ vector of unknown parameters[11]. It is also common to specify $\pi_{ij} = \pi_j$.

The fact that σ^2 is class-invariant means the log-likelihood function (36) is bounded. Thus, in principle, maximum likelihood estimates can be obtained using conventional optimization algorithms such as DFP. Unfortunately, direct maximization of the likelihood function often proves difficult, partly because the likelihood function is not necessarily concave (so there may be several local maxima), but also because the large numbers of parameters in latent class models render standard maximization algorithms unreliable. A more reliable alternative to direct maximization of the likelihood is the EM algorithm of Dempster, Laird and Rubin (1977). This algorithm involves writing the log-likelihood function in the form

$$\ln L_c = \sum_{j=1}^{J}\sum_{i=1}^{N} d_{ji}\left[\ln \pi_j + \ln f_j(y_i \mid \theta_j)\right] \tag{38}$$

where $\theta_j \equiv (\delta_j' \quad \sigma^2)'$; d_{ji} is an unobserved dummy variable that takes the value 1 if and only if y_i belongs to the *j*-th mixtures component (class); and $f_j(y_i \mid \theta_j)$ is the density of y_i when $d_{ji} = 1$. The EM algorithm for maximizing (38) is an iterative algorithm involving two steps: E (for expectation) and M (for maximisation). Details are provided in the Appendix.

Implementing the EM algorithm involves selecting a set of starting values and a stopping rule. The algorithm usually converges slowly and the likelihood function may have several local maxima, so the algorithm needs to be implemented using several sets of carefully-chosen starting values. If the errors are independently distributed, it is common to select starting values by randomly partitioning the data into classes and then estimating the parameters using conventional techniques. The EM algorithm is then applied using each of these "random starts". To stop the algorithm, let $L_c^{(k)}$ denote the value of the likelihood function after the *k*-th iteration. The E- and M-steps are usually repeated until $L_c^{(k+1)} - L_c^{(k)}$ changes by an arbitrarily small amount. Define

[11] If x_i contains a constant then the constraint $\eta_J = 0$ is necessary for identification purposes. This model is sometimes known as a mixture of experts model (e.g., McLachlan and Peel, 2000, p.167).

(39) $$D^{(k)} = L_c^{(k)} + \frac{\left(L_c^{(k+1)} - L_c^{(k)}\right)}{\left(1 - a^{(k)}\right)}$$

where $a^{(k)} = \left(L_c^{(k+1)} - L_c^{(k)}\right) / \left(L_c^{(k)} - L_c^{(k-1)}\right)$. Böhning et al (1994) suggest the EM algorithm can be stopped if $\left|D^{(k+1)} - D^{(k)}\right|$ is less than a desired tolerance.

Irrespective of the method used to compute the latent class estimates, implementation of the method also involves selecting the number of classes, or components. A natural way of doing this is to use the likelihood ratio statistic to test for the smallest value of J that is compatible with the data. Unfortunately, not enough regularity conditions hold for the likelihood ratio statistic to have the usual chi-squared distribution, so it is common to simply select the number of classes using a conventional information criterion, such as the Akaike Information Criterion (AIC) or the Bayesian information criterion (BIC). The AIC has a tendency to fit too many components, while the BIC has a tendency to fit too few components when the model for the component densities is valid and the sample size is small – see Celeux and Soromenho (1996). An alternative criterion that has performed well in simulation experiments is the Integrated Classification Likelihood Criterion (ICL) of Biernacki, Celeux and Govaert (1998).

Unfortunately, none of the well-known econometrics packages have latent class options that implement the EM algorithm or compute the ICL. LIMDEP does have an option for maximizing the latent class likelihood function directly, but it is only available for use with panel data.

5. RESTRICTIONS

Economic theory very often provides information concerning linear functions of the first-order derivatives of (1). This information can be written in the form

(40) $$R(Z)g(Z) = r(Z)$$

where $g(Z)$ is the $K \times 1$ gradient vector with $m_k(Z)$ in the k-th row; $R(Z)$ is a known $J \times K$ matrix; and $r(Z)$ is a known $J \times 1$ vector. Constant returns to scale production functions, input and output distance functions, and all demand, supply, cost and profit functions must satisfy homogeneity constraints that take this form. The econometric model incorporating these constraints is given by equations (2), (3) and

(41) $\quad R(z_i)g(z_i) = r(z_i).$

In this section, we show how different estimators can be used to incorporate this information into the estimation process.

Linear equality constraints involving the first-order derivatives have no bearing on the nonparametric estimation of the conditional mean, but do impact on estimation of the gradients. The restricted nonparametric estimator of the gradient vector is

(42) $\quad \tilde{g}(z) = \hat{g}(z) - R(z)'\left(R(z)R(z)'\right)^{-1}\left(R(z)'\hat{g}(z) - r(z)\right)$

where $\hat{g}(z)$ is the unrestricted estimator of $g(z)$, with elements $\hat{m}_k(z)$ computed using (12) .

The linear equality constraints (41) can often be incorporated into fixed-parameter flexible functional form econometric models, but in a very restrictive way. To see this, observe that the gradients (22) are linear in the unknown parameters, so the constraints (41) take the form

(43) $\quad R_i\beta = r_i \qquad$ for $i = 1, \ldots, N$,

where $R_i \equiv R(z_i)$ is a known $J \times K^*$ matrix, and $r_i \equiv r(z_i)$ is a $J \times 1$ vector that is known if and only if $a(z_i) = a^*$ (recall that this assumption is usually implicit when estimating models of this type). Unfortunately, these observation-varying constraints represent a system of *JN* equations in K^* unknowns, for which there is no general solution. The usual way forward is to devise a set of observation-invariant constraints that are sufficient but not necessary for (41) to hold:

(44) $\quad R\beta = r$

where R is $J \times K^*$ and r is $J \times 1$. Examples include the adding-up constraints that are usually imposed on the parameters of flexible functional forms in order to impose homogeneity. These constraints can be imposed using conventional econometric estimators, including restricted least squares (RLS). A problem with this approach is that sufficient conditions of the form (44) may not be available. In any event, the constraints (44) are sufficient but not necessary for the constraints (41) to hold, so conventional econometric estimators potentially over-constrain the parameter space, leading to biased and inconsistent estimates.

The observation-varying constraints (41) are most easily incorporated into a flexible functional form model in which the parameters are also observation-varying. To estimate the observation-varying parameter model given by (32), (3) and (41) it is convenient to write the constraints (41) in the form

(45) $R_i \beta_i = r_i$ for $i = 1, \ldots, N$,

which is the same as (43) except that the parameter vector is observation-varying . Moreover, there is no approximation error in this model, so R_i and r_i are always known. If R_i is of full rank then the general solution to (45) is (Graybill, 1969, p.142)

(46) $\beta_i = R_i^+ r_i + H_i \gamma_i$

where γ_i is an arbitrary $K^* \times 1$ vector, R_i^+ is the unique $K^* \times J$ Moore-Penrose generalised inverse of R_i, and $H_i \equiv I_{K^*} - R_i^+ R_i$ is a symmetric idempotent $K^* \times K^*$ matrix. A vector β_i will satisfy the observation-varying constraints (45) if and only if it has the form given by (46). Substituting (46) into the regression equation (32) yields an estimating equation of the form

(47) $y_i^R = x_i'^R \gamma_i + e_i$

where $y_i^R = y_i - x_i' R_i^+ r_i$ is $N \times 1$ and $x_i'^R = x_i' H_i$ is $N \times K^*$. The observation-varying parameter model is now given by (47) and (3). Importantly, the parameter vector γ_i in this model is theoretically unconstrained, so the unconstrained flexible least squares and latent class estimators discussed in the previous section are available. Letting $\tilde{\gamma}_i$ denote a flexible least squares or latent class estimator of γ_i, an estimator of β_i is then $\tilde{\beta}_i = R_i^+ r_i + H_i \tilde{\gamma}_i$. If we use a latent class estimator with $J = 1$ then $\tilde{\beta}_i$ collapses to the Generalised Restricted Least Squares (GRLS) estimator[12] of Doran, O'Donnell and Rambaldi (2003). The GRLS estimator derives its name from the fact that it collapses to the conventional RLS estimator if and only if R_i and r_i are observation-invariant. Thus, whereas the unrestricted latent class estimator collapses to OLS whenever $J = 1$, the restricted latent class estimator only collapses to RLS when $J = 1$ and the constraints are of a very simple (i.e., observation-invariant) form.

[12] The GRLS estimator is identical to the Singular Value Decomposition (SVD) estimator of O'Donnell, Rambaldi and Doran (2001). Doran, O'Donnell and Rambaldi (2003) present the estimator in a more general framework and establish some additional theoretical properties.

For more details concerning the GRLS estimator and its properties, see Doran, O'Donnell and Rambaldi (2003).

6. SIMULATION EXPERIMENT

To assess the empirical performance of different estimators, we took the two-input stochastic production function used in experiments by White (1980, p.151), Byron and Bera (1983, p.255) and Pagan and Ullah (1999, p.190):

$$(48) \qquad y_i = -\frac{1}{\gamma_1}\ln\left(e^{-\gamma_1 z_{1i}} + \gamma_2 e^{-\gamma_1 z_{2i}}\right) + e_i \qquad\qquad i = 1, \ldots, 200$$

where y_i represents the *i*-th observation on the logarithm of output and z_{ki} represents the *i*-th observation on the logarithm of the *k*-th input (k = 1, 2). We set $\gamma_1 = 5$, $\gamma_2 = 2$, $z_1 = (0.5 \quad 0.5)'$, and drew another 199 observations on the two log-inputs from independent standard uniform distributions[13]. We then drew 200 observations on e_i from a normal distribution with mean zero and variance 0.01. These settings are identical to those used by White (1980) and others, apart from the fact that we set the elements of z_1 equal to the population means of the log-inputs. These other authors evaluated the performance of different estimators at this particular point and, in order for us to assess the performance of the FLS estimator, we needed to ensure this point was included in the data set (recall that the FLS estimator cannot be used to obtain out-of-sample predictions). Note that the variables are measured in logarithms, so the parametric models all have translog functional forms[14].

The economic quantities of primary interest are the conditional mean and the two elements of the gradient vector, all evaluated at the mean log-inputs. Four unrestricted estimators were used to estimate these quantities:

KR	Kernel regression
FFF	Flexible functional form
FLS	Flexible least squares
LC(3)	Latent class estimator with J = 3.

[13] Pagan and Ullah generate the X_js from "two independently distributed uniform random variables with mean .5 and variance 1/12" (p.191). A uniform random variable defined over the interval [0,1] has mean 0.5 and variance 1/12.

[14] A special case of the translog is the Cobb-Douglas functional form. As it happens, the production surface in this experiment is the surface depicted in panel (a) of Figure 3. The surfaces depicted in panels (b) to (d) in Figure 3 are the estimated surfaces obtained using a Cobb-Douglas functional form (and an unrestricted latent class estimator with J = 3).

The number of classes in the LC model was chosen without computing any information criteria, but in the knowledge that mixtures of J = 3 normal densities are sufficiently flexible to approximate a very large range of density shapes. In any event, the results reported in this section are robust to this choice.

The production function (48) exhibits constant returns to scale (CRS). Accordingly, we also estimated the economic quantities of interest subject to the constraint:

$$\text{(49)} \qquad \sum_{k=1}^{K} m_k(z_i) = 1.$$

This is the necessary and sufficient condition for homogeneity established by Euler's Theorem. Acronyms for the restricted estimators are:

KRR	Restricted kernel regression
FFFR	Restricted flexible functional Form (translog)
FLSR	Restricted flexible least squares
LCR(3)	Restricted latent class estimator with J = 3.
GRLS	Generalised restricted least squares (restricted LC estimator with J = 1).

Obtaining the KRR estimates was straightforward – the econometric model is given by (2), (3) and (41) with $R(z_i) = [1 \quad 1]$ and $r(z_i) = 1$. Restricted estimation of the fixed-parameter flexible functional form model was also reasonably straightforward (under the assumption that the approximation error is observation-invariant). For this model the CRS restriction (49) becomes

$$\text{(50)} \qquad \sum_{k=1}^{K} \phi_k + \sum_{k=1}^{K}\sum_{j=1}^{K} \gamma_{kj} z_{ji} = 1$$

which will be satisfied at all data points if and only if:

$$\text{(51)} \qquad \sum_{k=1}^{K} \phi_k = 1 \qquad \text{and} \qquad \sum_{j=1}^{K} \gamma_{kj} = 0 \qquad \text{for all } k.$$

Thus, the CRS flexible functional form econometric model is given by (23), (24) and (44) with

$$(52)\qquad R=\begin{bmatrix}0&1&1&0&0&0\\0&0&0&1&1&0\\0&0&0&0&1&1\end{bmatrix}\quad\text{and}\quad r=\begin{bmatrix}1\\0\\0\end{bmatrix}.$$

Importantly, the constraints (51) are sufficient but not necessary for the CRS property (49) to hold. To impose this property within a flexible functional form framework, the parameters of the model must be observation-varying. Then the necessary and sufficient condition for CRS is

$$(53)\qquad \sum_{k=1}^{K}\phi_{ki}+\sum_{k=1}^{K}\sum_{j=1}^{K}\gamma_{kji}z_{ji}=1.$$

This constraint is identical to (50) except that the parameters are observation-varying. The CRS observation-varying parameter model is given by (32), (3) and (45) with

$$(54)\qquad R_i=\begin{bmatrix}0 & 1 & 1 & z_{1i} & (z_{1i}+z_{2i}) & z_{2i}\end{bmatrix}$$

and $r_i=1$. Substituting the restriction into the estimating equation yields the (unrestricted) estimating equation (47), which can be estimated using flexible least squares or a latent class estimator. The FLSR, LCR and GRLS estimates reported in this section were obtained in this way.

The covariance matrix of the unrestricted kernel regression estimator was computed using (15) to (17). The restricted kernel regression estimator of the gradient vector is a linear transformation of the unrestricted estimator, so its covariance matrix was computed using a simple transformation of the unrestricted covariance matrix (e.g., Greene, 2003, p.869).

Standard errors for the unrestricted and restricted flexible least squares estimates were computed using a simple bootstrapping procedure involving re-sampling from the (artificial) data set with replacement. The standard errors reported below were computed using 500 bootstrap samples. Standard errors for the latent class estimators were obtained as the diagonal elements of the inverse of the following estimate of the information matrix:

$$(55)\qquad \hat{I}(\theta)=E\left[\left(\frac{\partial \ln L_c}{\partial\theta}\right)\left(\frac{\partial \ln L_c}{\partial\theta}\right)'\right]\Bigg|_{\theta=\tilde{\theta}}=E\left[\sum_{i=1}^{N}s(y_i;\theta)s(y_i;\theta)'\right]\Bigg|_{\theta=\tilde{\theta}}$$

where $\tilde{\theta}$ is an estimate of θ and the score vector is:

$$(56) \quad s(y_i;\theta) = \sum_{j=1}^{J} d_{ji} \left[\frac{\partial \ln f_j(y_i \mid \theta_j)}{\partial \theta} \right]$$

Results

GAUSS was used to estimate all models and compute estimates of all economic quantities of interest[15]. Results are summarised in table 1, which is divided into three sections. The top section of the table reports estimates of the economic quantities of interest evaluated at the population means of the log-inputs. The true value of the conditional mean is $m(z_1) = 0.280$ and the true values of the first derivatives are $m_1(z_1) = 0.333$ and $m_2(z_1) = 0.667$. Estimated standard errors are given in parentheses. Restricted estimates of $m_2(z_1)$ have not been reported because they are redundant (the estimated gradients sum to one).

The second section of the table reports differences between the estimated values and the true values (bias). The unrestricted estimator exhibiting least bias is marked with the symbol "‡", while the restricted estimator with least bias is marked with an asterisk. When bias is used as a criterion, it appears that the unrestricted latent class estimator outperforms all other unrestricted estimators of the conditional mean. The flexible least squares estimator is a superior unrestricted estimator of the first gradient, while the fixed-parameter flexible functional form estimator is a superior estimator of the second gradient. If the CRS restriction is imposed, the fixed-parameter flexible functional form estimator dominates all other estimators in terms of bias. This result is encouraging for economists using fixed-parameter flexible functional form models to evaluate economic quantities at variable means.

The third section of table 1 reports the root mean square error (RMSE) for each estimator. Using this criterion, the latent class estimator outperforms all other estimators of the conditional mean, whether or not the constant returns to scale restriction is imposed. This result should not discourage researchers from using the fixed-parameter FFF estimator, on the grounds that it is better to be vaguely right (small bias) than precisely wrong (small RMSE). In any event, the restricted FFF estimator also has lowest RMSE when it comes to estimation of the gradients.

Finally, table 2 reports results obtained when economic quantities of interest are evaluated at the point $z_{49} = (0.05 \quad 0.03)'$. This is a point where both log-inputs are at the low end of the range of values represented in the data set. The rankings of the estimators are unchanged when it comes to estimation of the conditional mean – the latent class estimator

[15] Where possible, the results were validated using in-built functions in SHAZAM.

has lower bias than the fixed-parameter FFF estimator unless the CRS restriction is imposed. However, the opposite is the case when it comes to estimating the first gradient – the FFF estimator outperforms the LC(3) estimator unless the CRS restriction is imposed.

7. DEMAND FOR LABOUR IN U.S. AGRICULTURE

To illustrate the application of different estimators in a panel data context, we also estimated an aggregate demand function for labour in U.S. agriculture. The data file contained observations on input and output prices and quantities used in agriculture in state *i* in year *t*:

x_{1it} = quantity index of materials
x_{2it} = quantity index of capital
x_{3it} = quantity index of land
x_{4it} = quantity index of labour
w_{1it} = implicit price index of materials
w_{2it} = implicit price index of capital
w_{3it} = implicit price index of land
w_{4it} = implicit price index of labour
q_{it} = quantity index of total output

for the N = 48 contiguous states of the U.S. from 1960 to 1996. Thus, there are T = 37 time periods and a total of NT = 1776 observations in the data set. The data is the same as that described in Ball *et al* (1999) except that it has been extended from 1990 to 1996, and the EKS technique due to Elteto and Koves (1964) and Szulc (1964) has been used to convert binary Tornquist indices into transitive multilateral Tornquist indexes.

All the models and estimation methods discussed in previous sections are applicable in the panel context, with the exception of flexible least squares (FLS is only available in a time-series context). We simply append a *t* subscript to all observation-varying quantities to index time periods. For simplicity, we chose not to assume an error structure that accounted for the panel nature of the data set (i.e., we assumed the covariance matrix for all NT errors was $\sigma^2 I$).

Economic theory prescribes a conditional input demand function for labour that takes the form of (2) with:

(56) $y_{it} \equiv \ln x_{4it}$

(57) $z_{it} \equiv (\ln w_{1it} \quad \ln w_{2it} \quad \ln w_{3it} \quad \ln w_{4i} \quad \ln q_{it})'$.

and

$$(58) \qquad \sum_{k=1}^{4} m_k(z_{it}) = 0.$$

Equation (58) is a necessary and sufficient condition for the demand function to be homogeneous of degree zero in prices. Economic theory also suggests $m_4(z_{it}) \leq 0$ (i.e., the own-price elasticity of demand for labour is negative).

For nonparametric estimation, the homogeneity constraint is written in the form of (41) with $R(z_{it}) = [1 \quad 1 \quad 1 \quad 1 \quad 0]$ and $r(z_{it}) = 0$. For estimation of the fixed-parameter flexible functional form model, the constraint is written in the form of (44) with

$$(59) \qquad R = \begin{bmatrix} . & 1 & 1 & 1 & 1 & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & 1 & 1 & 1 & 1 & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & 1 & . & . & . & 1 & 1 & 1 & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & 1 & . & . & . & 1 & . & . & 1 & 1 & . & . & . & . \\ . & . & . & . & . & . & . & . & . & 1 & . & . & . & 1 & . & . & 1 & . & 1 & . & . \\ . & . & . & . & . & . & . & . & . & . & 1 & . & . & . & 1 & . & . & 1 & . & 1 & . \end{bmatrix}$$

and $r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}'$. Again, this constraint is sufficient but not necessary for zero-degree homogeneity of the unknown input demand function. Finally, in the case of the observation-varying parameter model, the constraint is in the form of (45) with

$$(60) \qquad R_i = [0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad z_{1i} \quad (z_{1i}+z_{2i}) \quad (z_{1i}+z_{3i}) \quad (z_{1i}+z_{4i}) \quad 0$$
$$\cdots \quad z_{2i} \quad (z_{2i}+z_{3i}) \quad (z_{2i}+z_{4i}) \quad 0 \quad z_{3i} \quad (z_{3i}+z_{4i}) \quad 0 \quad z_{4i} \quad 0 \quad 0]$$

and $r_i = 0$. The latent class estimator was estimated with $J = 2$.

Results

GAUSS was again used to obtain the empirical results. The parameter estimates are reported in table 3, where standard errors are presented in parentheses. The unrestricted first-order coefficients reported in the FFF and LC(2) columns of table 3 are estimates of the parameters in (23) and have meaningful interpretations – the output and input price variables were scaled in such a way that these first-order coefficients can be interpreted as elasticities evaluated at the variable means. The FFF estimate of the materials elasticity is 0.275, while the LC(2) estimate is (0.643)(0.561) + (0.357)(-0.132) = 0.314 (i.e., a probability weighted average of the class-specific estimates). Thus, both elasticity estimates are positive, suggesting labour and materials are substitutes in production. In contrast, the unrestricted estimates of the capital elasticity are both negative, suggesting labour and capital are complements.

The first-order coefficients reported in the FFR column in table 3 are restricted estimates of the elasticities evaluated at the variable means, and they sum to zero. However, the estimates reported in the LCR(2) and GRLS columns have no such convenient interpretations, nor do they satisfy any adding-up constraints. This is because the coefficients in these remaining columns are estimates of the theoretically-unconstrained parameters in the transformed model (47). To obtain restricted LCR and GRLS estimates of elasticities at the variable means, we need to apply the transformation given by (46). Such a transformation has been used to generate the results reported in table 4.

Table 4 reports KR, KRR, FF, FFR, LC, LCR and GRLS estimates of economic quantities of interest (expected log-demand and elasticities with respect to all explanatory variables) when evaluated at the variable means. It can be seen from table 4 that imposing homogeneity has led to a sign reversal in the estimated land elasticity. It is also apparent that different restricted estimation approaches yield qualitatively different estimates of quantities of economic interest: the kernel regression estimates of the land elasticity have a different sign to estimates obtained using all other estimators; the LCR and GRLS estimates of the elasticities are large by comparison with the FFR estimates, with the exception of the materials elasticity. Even greater differences emerge when economic quantities of interest are evaluated at points away from the variable means. To illustrate, table 5 reports estimated quantities of interest when evaluated at the first observation in the data set (Alabama in 1960). It can be seen from this table that the LCR and GRLS output predictions are (precise and) almost twice as large as the FFR predictions; the FFR estimates suggest that labour and materials are substitutes in production while the LCR estimates suggest that they are complements; and the FFR estimates suggest that labour demand is own-price elastic while the LCR estimates suggest that it is own-price inelastic. Most of these results can be validated by selectively citing conflicting results reported across studies by Chambers (1982), Andrikopoulos and Brox (1992), Lopez, Shumway, Saez and Gottret (1988) and O’Donnell, Shumway and Ball (1999). The conclusion we draw from this empirical illustration is that the choice of estimation methodology matters when evaluating economic quantities of interest, whether or not they are evaluated at the variable means.

7. CONCLUSION

Common econometric approaches to estimating economic relationships range from nonparametric kernel regression estimators to more conventional estimators of fixed-parameter flexible functional forms. This paper has motivated and empirically evaluated a latent class estimator that includes the nonparametric and flexible functional form estimators as limiting special cases. The unrestricted latent class estimator collapses to the OLS estimator when there is only one class, and is equivalent to a nonparametric estimator when the number of classes coincides with the number of observations in the data set. Unlike conventional flexible functional form models, latent class models can be motivated without making restrictive assumptions concerning the remainder terms in Taylor's series expansions.

Latent class models are typically estimated by the method of maximum likelihood. Empirical examples from the economics literature include Beard, Caudill and Gropper (1991, 1994), Caudill (2003) and Orea and Kumbhakar (2004). These authors have maximized the likelihood function in one of two ways: directly, using well-known gradient methods; or using the EM algorithm of Dempster, Laird and Rubin (1977). Unfortunately, direct maximization of the likelihood function can be problematic – gradient methods often fail to converge, particularly when the number of classes is large. The EM algorithm is guaranteed to converge, but has not yet been programmed into well-known econometrics software packages.

The simulation example presented in the paper provides evidence that fixed-parameter flexible functional form models can provide adequate approximations to economic quantities of interest when the latter are evaluated at or near the variable means. However, matters are not so straightforward when interest extends to points at the outer ranges of values represented in the data set. In these cases, latent class estimators (with more than one class) may outperform fixed-parameter estimators in some, and possibly many, empirical applications.

This paper has explained and evaluated latent class estimators in the context of a model where the error term is normally distributed (a finite mixture of normals model). Latent class estimators can also be used to estimate other types of models, including probit and tobit models, and random-effects stochastic frontiers. Maximum likelihood estimation of such models using the EM algorithm is computationally demanding, not least because there is no closed form solution to the maximization problem in the M-step (so iterative solution methods must be used). Fortunately, these types of models can be conveniently estimated within a Bayesian framework. Koop (2003) provides details concerning Bayesian estimation of finite mixtures of normals models, while O'Donnell and Griffiths (2006) provide details concerning latent class stochastic frontiers.

REFERENCES

Ball, V.E., F. Gollop, A. Kelly-Hawke, and G. Swinand (1999) "Patterns of State Productivity Growth in the U.S. Farm Sector: Linking State and Aggregate Models" *American Journal of Agricultural Economics* 81(1):164-179.

Beard, T.R., Caudill S.B. and D.M. Gropper (1991) "Finite Mixture Estimation of Multiproduct Cost Functions" *Review of Economics and Statistics* 73:654-664.

Beard, T.R., Caudill S.B. and D.M. Gropper (1994) "The Diffusion of Production Processes in the U.S. Banking Industry: A Finite Mixture Approach" *Journal of Banking and Finance* 21:721-740.

Biernacki, C., Celeux, G., and G. Govaert (1998) "Assessing a Mixture Model for Clustering With the Integrated Classification Likelihood" *Technical Report No. 3521*. Rhône-Alpes. INRIA.

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and B. Lindsay (1994) "The Distribution of the Likelihood Ratio for Mixtures of Densities from the One-Parameter Exponential Family" Annals of the Institute of Statistical Mathematics 46:283-303.

Byron, R.P. and A.K. Bera (1983) "Least Squares Approximations to Unknown Regression Functions: A Comment" *International Economic Review* 24:255-260.

Caudill, S.B. (2003) "Estimating a Mixture of Stochastic Frontier Regression Models via the EM Algorithm" *Empirical Economics* 28:581-598.

Celeux, G. and G. Soromenho (1996) "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model" *Classification Journal* 13:195-212.

Christensen, L., Jorgenson, D. and L. Lau (1973) "Transcendental Logarithmic Production Frontiers" *Review of Economics and Statistics* 55:28-45.

Dempster, A.P., Laird, N. and D. Rubin (1977) "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm" *Journal of the Royal Statistical Society, Series B* 39:1-38.

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

Doran, H.E., C.J. O'Donnell, A.N. Rambaldi (2003) "Imposing Observation-Varying Equality Constraints Using Generalised Restricted Least Squares" *University of Queensland School of Economics Discussion Papers Series* ISSN 1446 55232.

Epanechnikov, V.A. (1969) “Nonparametric Estimates of a Multivariate Probability Density” *Theory of Probability and Applications* 14:153-158.

Elteto, O. and P. Koves (1964) "On a Problem of Index Number Computation Relating to International Comparison" *Statisztikai Szemle* 42:507-518.

Gallant, A.R. (1981) “On the Bias in Flexible Functional Forms and an Essentially Unbiased Form” *Journal of Econometrics* 15:211-245.

Gallant, A.R. “The Fourier Flexible Form” *American Journal of Agricultural Economics* 66(2):206-208.

Graybill, F.A. (1969) *Introduction to Matrices with Applications in Statistics*. Belmont CA. Wadsworth.

Greene, W.H. (2003) *Econometric Analysis*. 5th ed. New York. Prentice Hall.

Kalaba, R. and L.Tesfatsion (1989) “Time-Varying Linear Regression Via Flexible Least Squares” *Computers and Mathematics with Applications* 17(8/9):1215-1245.

Koop, G. (2003) *Bayesian Econometrics*. Chichester, UK. Wiley.

McLachlan, G. and D. Peel (2000) *Finite Mixture Models*. New York. Wiley.

O’Donnell, C.J., A.N. Rambaldi and H.E. Doran (2001) "Estimating Economic Relationships Subject to Firm- and Time-Varying Equality and Inequality Constraints" *Journal of Applied Econometrics* 16:709-726.

O'Donnell, C.J. and W.E. Griffiths (2006) "Estimating State-Contingent Production Frontiers" *American Journal of Agricultural Economics* 88(1):249-266.

Orea, L. and S.C. Kumbhakar (2004) “Efficiency Measurement Using a Latent Class Stochastic Frontier Model” *Empirical Economics* 29:169-183.

Pagan, A. and A. Ullah (1999) *Nonparametric Econometrics*. Cambridge. Cambridge University Press.

Rosenblatt, M. (1956) “Remarks on Some Nonparametric Estimates of a Density Function” *Annals of Mathematical Statistics* 27:642-669.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. New York. Chapman and Hall.

Szulc, B.J. (1964) "Indices for Multiregional Comparisons" Prezeglad Statystyczny (Statistical Review) 3:239-254.

White, H. (1980) “Using Least Squares to Approximate Unknown Regression Functions” *International Economic Review* 21:149-170.

APPENDIX

The EM Algorithm

The E-Step

Suppose the values of the π_j and θ_j were known. Then the conditional expectation of $\ln L_c$ (i.e., conditional on the data) would be

$$\text{(A.1)}\quad E\{\ln L_c \mid y_1,\ldots,y_N\} = \sum_{j=1}^{J}\sum_{i=1}^{N} E\{d_{ji} \mid y_1,\ldots,y_N\}\left[\ln \pi_j + \ln f_j(y_i \mid \theta_j)\right].$$

where $E\{d_{ji} \mid y_1,\ldots,y_N\} = P(d_{ji}=1 \mid y_1,\ldots,y_N) = \tau_{ji}$ is the posterior probability that y_i belongs to the j-th component and is given by

$$\text{(A.2)}\quad \tau_{ji} \equiv P(d_{ji}=1 \mid y_i) = \frac{f_j(y_i \mid d_{ji}=1)P(d_{ji}=1)}{f(y_i)} = \frac{\pi_j f_j(y_i \mid \theta_j)}{\sum_{s=1}^{J}\pi_s f_s(y_i \mid \theta_s)}$$

An estimate of τ_{ji} is

$$\text{(A.3)}\quad \tau_{ji}^{(k)} = \frac{\pi_j^{(k)} f_j(y_i \mid \theta_j^{(k)})}{\sum_{s=1}^{J}\pi_s^{(k)} f_s(y_i \mid \theta_s^{(k)})}\,.$$

where $\pi_j^{(k)}$ and $\theta_j^{(k)}$ denote estimates of π_j and θ_j at the k-th iteration of the EM algorithm. Thus, an estimate of (A.1) is

$$\text{(A.4)}\quad \ln L_c^{(k)} = \sum_{j=1}^{J}\sum_{i=1}^{N}\tau_{ji}^{(k)}\left[\ln \pi_j + \ln f_j(y_i \mid \theta_j)\right]$$

The M-Step

The M-step involves maximising (A.4) with respect to π_j and θ_j to obtain $\pi_j^{(k+1)}$ and $\theta_j^{(k+1)}$. The mixing proportions $\pi_j^{(k+1)}$ are updated independently of the $\theta_j^{(k+1)}$. To update the mixing proportions, we note that if the d_{ij}s were observed then the maximum likelihood estimator of π_j would be the sample mean:

(A.5) $\hat{\pi}_j = \frac{1}{N}\sum_{i=1}^{N} d_{ji}$

Recall that in the E-step we replaced each d_{ij} with its conditional expectation. This suggests we should estimate π_j using:

(A.6) $\pi_j^{(k+1)} = \frac{1}{N}\sum_{j=1}^{N} \tau_{ji}^{(k)}$

To update the $\theta_j^{(k+1)}$ we maximise

(A.7) $S = \sum_{j=1}^{J}\sum_{i=1}^{N} \tau_{ji}^{(k)} \ln f_j(y_i \mid \theta_j^{(k)})$

with respect to $\theta_j^{(k+1)}$. If the mixture densities are normal then closed form solutions are available.

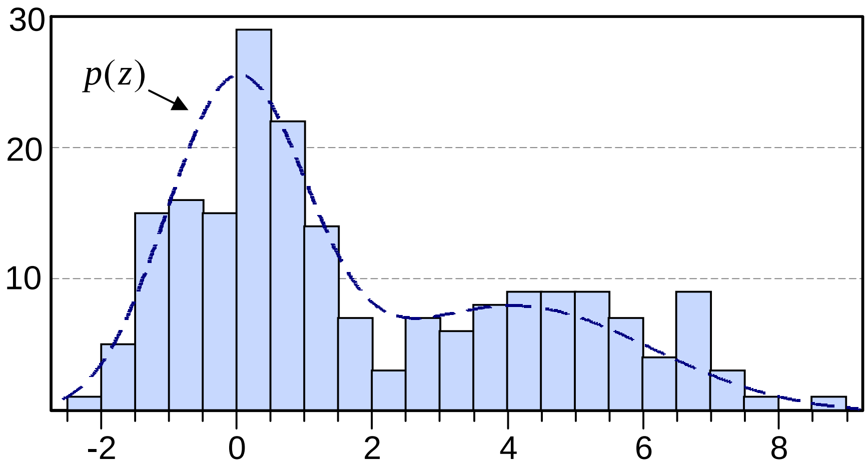


Figure 1. Local Histogram Estimator

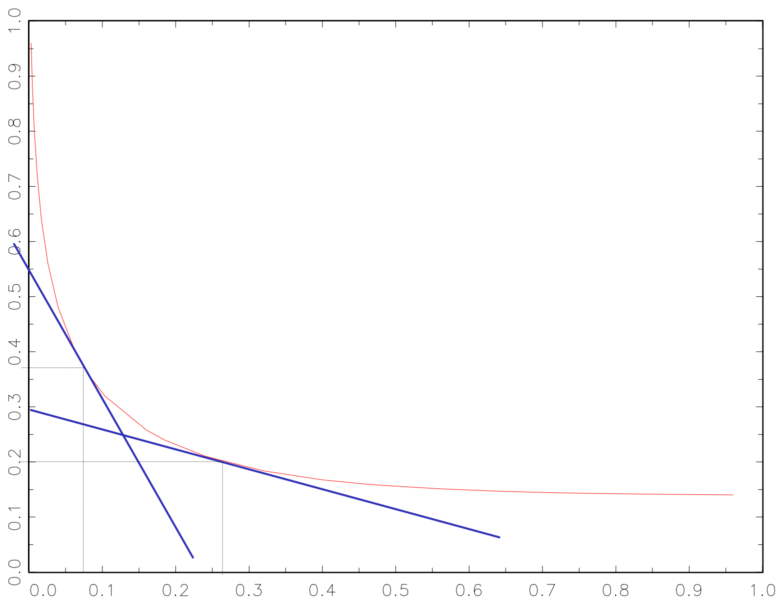
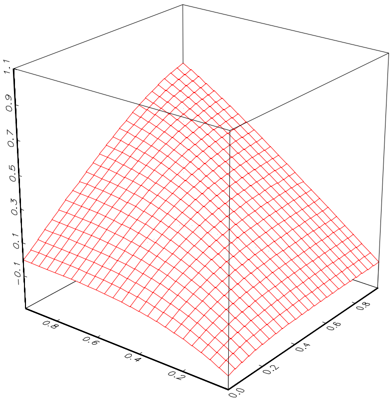
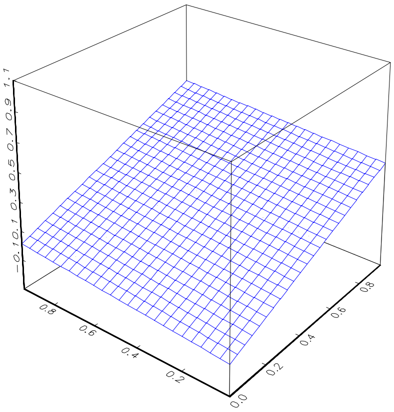


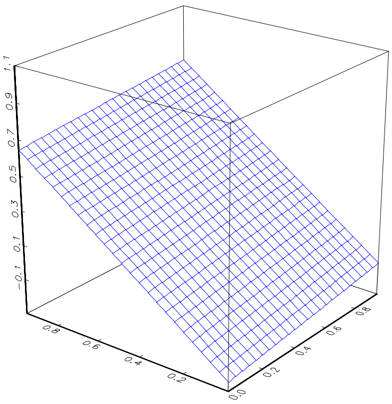
Figure 2. Linear Approximations in Two Dimensions



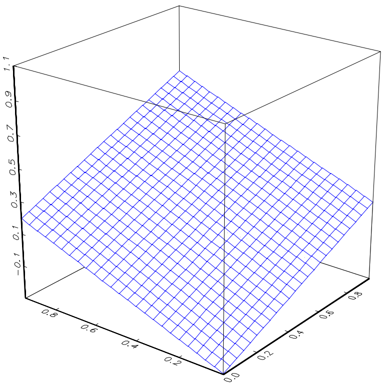
(a)



(b)



(c)



(d)

Figure 3. Linear Approximations in Three Dimensions

Table 1. Economic Quantities of Interest Evaluated at $z_1 = (0.5 \quad 0.5)'$ [(a)]

	$m(z_1)$	$m_1(z_1)$	$m_2(z_1)$
Estimates			
TRUE	0.280	0.333	0.667
KR	0.258 (0.017)	0.440 (0.022)	0.376 (0.001)
FF	0.274 (0.013)	0.404 (0.024)	0.593 (0.026)
FLS	0.231 (0.077)	0.283 (0.122)	0.585 (0.144)
LC	0.278 (0.006)	0.401 (0.010)	0.587 (0.011)
KRR	0.258 (0.017)	0.532 (0.012)	
FFR	0.281 (0.009)	0.405 (0.018)	
FLSR	0.235 (0.073)	0.464 (0.121)	
LCR(3)	0.278 (0.006)	0.492 (0.050)	
GRLS	0.274 (0.014)	0.456 (0.122)	
Bias			
KR	-0.023	0.107	-0.290
FF	-0.006	0.070	-0.074‡
FLS	-0.049	-0.051‡	-0.082
LC(3)	-0.002‡	0.068	-0.079
KRR	-0.023	0.199	
FFR	0.001*	0.072*	
FLSR	-0.046	0.131	
LCR(3)	-0.003	0.158	
GRLS	-0.006	0.123	
RMSE			
KR	0.029	0.109	0.290
FF	0.015	0.074	0.078‡
FLS	0.091	0.132	0.166
LC(3)	0.006‡	0.069‡	0.080
KRR	0.029	0.199	
FFR	0.009	0.074*	
FLSR	0.086	0.178	
LCR(3)	0.006*	0.166	
GRLS	0.015	0.173	

(a) Estimated standard errors in parentheses

Table 2. Economic Quantities of Interest Evaluated at $z_{49} = (0.05 \quad 0.03)'$ [(a)]

	$m(z_{49})$	$m_1(z_{49})$	$m_2(z_{49})$
Estimates			
TRUE	-0.184	0.308	0.692
KR	-0.095	-0.026	0.047
	(0.023)	(0.121)	(0.136)
FF	-0.164	0.301	0.595
	(0.028)	(0.092)	(0.104)
FLS	-0.132	0.393	0.555
	(0.066)	(0.142)	(0.173)
LC	-0.170	0.325	0.623
	(0.012)	(0.038)	(0.043)
KRR	-0.095	0.463	
	(0.023)	(0.008)	
FFR	-0.182	0.387	
	(0.009)	(0.018)	
FLSR	-0.131	0.452	
	(0.068)	(0.142)	
LCR	-0.166	0.373	
	(0.009)	(0.013)	
GRLS	-0.170	0.376	
	(0.022)	(0.031)	
Bias			
KR	0.089	-0.334	-0.645
FF	0.020	-0.007‡	-0.097
FLS	0.052	0.085	-0.138
LC(3)	0.014‡	0.017	-0.069‡
KRR	0.089	0.155	
FFR	0.002*	0.079	
FLSR	0.053	0.144	
LCR(3)	0.018	0.065*	
GRLS	0.014	0.069	
RMSE			
KR	0.092	0.355	0.659
FF	0.034	0.092	0.142
FLS	0.085	0.165	0.221
LC(3)	0.018‡	0.042‡	0.081‡
KRR	0.092	0.156	
FFR	0.010*	0.081	
FLSR	0.086	0.202	
LCR(3)	0.020	0.066*	
GRLS	0.026	0.075	

(a) Estimated standard errors in parentheses

Table 3. Parameter Estimates – US Labour Demand [(a)]

Variable	FFF	FFFR	LC(2)		LCR(2)		GRLS
			$j = 1$	$j = 2$	$j = 1$	$j = 2$	
Constant	0.635	0.689	0.760	0.459	0.464	-0.219	0.566
	(0.018)	(0.016)	(0.013)	(0.043)	(0.021)	(0.045)	(0.029)
ln w_{1it}	0.275	0.991	0.561	-0.132	-1.751	-0.033	0.033
	(0.060)	(0.056)	(0.037)	(0.016)	(0.040)	(0.019)	(0.062)
ln w_{2it}	-0.516	-0.590	-0.499	-0.463	-1.021	0.897	-1.654
	(0.052)	(0.061)	(0.025)	(0.009)	(0.033)	(0.009)	(0.056)
ln w_{3it}	0.150	-0.030	-0.038	0.423	3.132	-0.973	-0.343
	(0.023)	(0.024)	(0.274)	(0.222)	(0.273)	(0.238)	(0.026)
ln w_{4it}	-0.411	-0.372	-0.386	-0.506	-0.243	-1.256	-0.876
	(0.034)	(0.042)	(0.072)	(0.137)	(0.078)	(0.150)	(0.046)
ln q_{it}	0.853	0.844	0.797	0.900	0.370	2.718	0.846
	(0.012)	(0.013)	(0.035)	(0.257)	(0.036)	(0.294)	(0.012)
½ln w_{1it}ln w_{1it}	-0.230	-0.168	-0.457	0.648	-1.492	-0.649	1.865
	(0.380)	(0.339)	(0.069)	(0.128)	(0.073)	(0.142)	(0.378)
ln w_{1it}ln w_{2it}	-0.207	0.824	0.128	-0.271	-0.001	0.684	-0.628
	(0.309)	(0.322)	(0.028)	(0.031)	(0.030)	(0.029)	(0.328)
ln w_{1it}ln w_{3it}	0.225	-0.360	0.377	-0.019	-0.010	-0.033	-0.046
	(0.099)	(0.091)	(0.044)	(0.009)	(0.044)	(0.009)	(0.108)
ln w_{1it}ln w_{4it}	-0.106	-0.295	-0.081	-0.654	1.766	-0.212	-0.283
	(0.190)	(0.182)	(0.107)	(0.019)	(0.119)	(0.019)	(0.208)
ln w_{1it}ln q_{it}	0.160	0.190	0.167	0.138	-0.071	0.643	0.257
	(0.049)	(0.047)	(0.007)	(0.013)	(0.007)	(0.021)	(0.050)
½ln w_{2it}ln w_{2it}	1.824	0.106	1.162	1.987	0.265	-1.560	1.688
	(0.357)	(0.418)	(0.043)	(0.037)	(0.045)	(0.040)	(0.407)
ln w_{2it}ln w_{3it}	-0.675	-0.145	-0.433	-1.089	-0.548	-0.851	-0.840
	(0.096)	(0.108)	(0.016)	(0.025)	(0.019)	(0.033)	(0.101)
ln w_{2it}ln w_{4it}	-0.705	-0.785	-0.747	-0.422	0.783	1.525	-1.088
	(0.178)	(0.212)	(0.009)	(0.274)	(0.009)	(0.273)	(0.197)
ln w_{2it}ln q_{it}	-0.118	-0.174	-0.193	0.054	-0.183	0.295	-0.188
	(0.040)	(0.046)	(0.222)	(0.072)	(0.238)	(0.078)	(0.041)
½ln w_{3it}ln w_{3it}	0.293	0.193	0.080	0.836	-0.050	0.224	0.113
	(0.043)	(0.043)	(0.137)	(0.035)	(0.150)	(0.036)	(0.040)
ln w_{3it}ln w_{4it}	0.200	0.312	0.118	0.130	0.749	-0.608	0.187
	(0.061)	(0.063)	(0.257)	(0.069)	(0.294)	(0.073)	(0.061)
ln w_{3it}ln q_{it}	-0.021	0.041	0.030	-0.032	-1.275	-0.278	-0.003
	(0.012)	(0.013)	(0.128)	(0.028)	(0.142)	(0.030)	(0.013)
½ln w_{4it}ln w_{4it}	0.653	0.768	0.573	1.098	-0.090	0.187	0.589
	(0.149)	(0.181)	(0.031)	(0.044)	(0.029)	(0.044)	(0.164)
ln w_{4it}ln q_{it}	0.054	-0.057	0.036	-0.115	0.055	0.015	0.031
	(0.026)	(0.030)	(0.009)	(0.107)	(0.009)	(0.119)	(0.027)
½ln q_{it}ln q_{it}	-0.007	-0.038	-0.038	-0.053	0.044	-0.041	-0.006
	(0.009)	(0.011)	(0.019)	(0.007)	(0.019)	(0.007)	(0.010)
π_j			0.643	0.357	0.358	0.642	
σ	0.311	0.383	0.224		0.235		0.325

(a) Estimated standard errors in parentheses. Subscripts refer to 1 = materials, 2 = capital, 3 = land, 4 = labour.

Table 4. Expected Log-output and Elasticities Evaluated at Variable Means[(a)]

		Materials	Capital	Land	Labour	Output
	$m(z_{1,1})$	$m_1(z_{1,1})$	$m_2(z_{1,1})$	$m_3(z_{1,1})$	$m_4(z_{1,1})$	$m_5(z_{1,1})$
KR	0.705	0.058	-0.491	-0.022	-0.194	0.424
	(0.025)	(0.034)	(0.001)	(0.004)	(0.014)	(0.000)
FF	0.635	0.275	-0.516	0.150	-0.411	0.853
	(0.018)	(0.060)	(0.052)	(0.023)	(0.034)	(0.012)
LC	0.652	0.314	-0.486	0.127	-0.429	0.834
	(0.009)	(0.032)	(0.027)	(0.012)	(0.018)	(0.006)
KRR	0.705	0.221	-0.329	0.140	-0.032	0.424
	(0.025)	(0.028)	(0.007)	(0.002)	(0.019)	(0.000)
FFR	0.689	0.991	-0.590	-0.030	-0.372	0.844
	(0.016)	(0.056)	(0.061)	(0.024)	(0.042)	(0.013)
LCR	0.579	0.092	-1.629	-0.363	-0.912	0.824
	(0.015)	(0.033)	(0.030)	(0.014)	(0.024)	(0.007)
GRLS	0.566	0.033	-1.654	-0.343	-0.876	0.846
	(0.029)	(0.062)	(0.056)	(0.026)	(0.046)	(0.012)

(a) Estimated standard errors in parentheses.

Table 5. Expected Log-output and Elasticities Evaluated at Alabama Prices and Outputs in 1960[(a)]

		Materials	Capital	Land	Labour	Output
	$m(z_{1,1})$	$m_1(z_{1,1})$	$m_2(z_{1,1})$	$m_3(z_{1,1})$	$m_4(z_{1,1})$	$m_5(z_{1,1})$
KR	1.034	-0.161	0.026	0.067	-0.395	0.394
	(0.041)	(0.016)	(0.002)	(0.050)	(0.122)	(0.015)
FF	1.240	0.139	0.590	-0.338	-1.161	0.834
	(0.049)	(0.180)	(0.183)	(0.056)	(0.115)	(0.023)
LC(2)	1.208	0.222	0.645	-0.423	-1.042	0.837
	(0.026)	(0.104)	(0.102)	(0.031)	(0.065)	(0.013)
KRR	1.034	-0.045	0.141	0.183	-0.279	0.394
	(0.041)	(0.055)	(0.037)	(0.011)	(0.082)	(0.015)
FFR	1.256	1.448	0.456	-0.636	-1.267	0.911
	(0.060)	(0.170)	(0.188)	(0.057)	(0.140)	(0.026)
LCR(2)	2.101	-0.552	1.263	-0.053	-0.658	0.825
	(0.027)	(0.166)	(0.203)	(0.061)	(0.102)	(0.014)
GRLS	2.117	-0.160	0.930	0.071	-0.840	0.816
	(0.051)	(0.312)	(0.382)	(0.115)	(0.192)	(0.026)

(a) Estimated standard errors in parentheses.