Kernel Regression in Empirical Microeconomics

Author(s): Richard Blundell and Alan Duncan

Source: *The Journal of Human Resources*, Winter, 1998, Vol. 33, No. 1 (Winter, 1998), pp. 62-87

Published by: University of Wisconsin Press

Stable URL: https://www.jstor.org/stable/146315

Kernel Regression in Empirical Microeconomics

Richard Blundell
Alan Duncan

ABSTRACT

We consider the implementation of Kernel methods in empirical microeconomics with specific application to Engel curve estimation in the analysis of consumer behavior. A set of recently developed tests for the parametric null hypothesis against a nonparametric alternative are discussed and implemented for the Engel curve application. We also consider semi-parametric estimation in partially linear models and the case of endogenous regressors. Gauss-based software is available for each technique implemented in the paper.

I. Introduction

The obsession with linearity in empirical economic analysis clearly does not stem from any strong prior of economic theory. To quote McFadden's 1985 presidential address to the Econometric Society: ''[parametric regression] interposes an untidy veil between econometric analysis and the propositions of economic theory.'' Nonparametric regression analysis seems to provide a compelling alternative to linear regression, allowing the data to determine the ''local'' shape of the conditional mean relationship.

Richard Blundell is a professor of economics at University College London and the Institute for Fiscal Studies and a coeditor of the Journal of Human Resources. *Alan Duncan is a lecturer in economics at the University of York and the Institute for Fiscal Studies.* *The authors are grateful to James Banks, Hidehiko Ichimura, Gauthier Lanot, Arthur Lewbel, Joris Pinske and two anonymous referees for helpful comments. The financial support of the ESRC Centre for the Micro-Economics Analysis of Fiscal Policy at IFS is gratefully acknowledged. Household data from the FES made available by the CSO through the ESRC Data Archive has been used by permission of the HMSO. Neither the CSO nor the ESRC Data Archive bear responsibility for the analysis or the interpretation of the data reported here. The usual disclaimer applies.*

The paper begins with models in which the relationship of interest is given by

$$(1.1) \quad y = g(x) + \varepsilon$$

where ε is defined such that $E(\varepsilon | x) = 0$. For example, in Engel curve analysis y would represent the expenditure or expenditure share on some good or group of goods and x would represent total disposable income or the total budget. Nonparametric regression analysis avoids the imposition of any parametric assumptions on the conditional mean function $g(x)$. Typically x is univariate, as in Engel curve example, although multivariate relationships are relatively easily estimated using standard nonparametric regression techniques.

Engel curve analysis provides a particularly useful illustration of the advantages of nonparametric regression and will be used as a running illustration throughout this paper. It has been at the center of applied microeconomic research on consumer behavior since the early studies of Working (1943) and Leser (1963) uncovered the stability of the expenditure share—log income specification for food expenditures. Recently attention has focused on Engel curves that have more variety of curvature than is permitted by the Working-Leser form underlying the Translog and Almost Ideal models of Jorgenson, Christensen, and Lau (1975a) and Deaton and Muellbauer (1980a) respectively. This reflects growing evidence from a series of empirical studies that suggest quadratic logarithmic income terms are required for certain expenditure share equations (see, Banks, Blundell, and Lewbel 1997; Atkinson, Gomulka, and Stern 1990; Bierens and Pott-Buter 1990; Hausman, Newey, and Powell 1995; Härdle and Jerison 1991; Lewbel 1991; Blundell, Pashardes, and Weber 1993).

Our aim in this paper is to take the reader through the implementation and estimation of various aspects of Kernel regression. The focus will be restricted to kernel regression although nearest neighbor, series and spline techniques are now all commonly available alternative techniques (see Härdle 1990 for a comprehensive review). There will be some discussion of local linear regression techniques introduced recently by Fan (1992). All the techniques discussed are implemented on the expenditure survey data source using the GAUSS-based interactive software NP-REG. This software is available on request.[1]

In general, it may be useful to consider a parametric specification against a nonparametric alternative. We consider and implement a set of recently developed tests for this hypothesis (see Aït-Sahalia, Bickel, and Stoker 1994; Ellison and Ellison 1992; Härdle and Mammon 1993; Zheng 1996). These turn out to give similar and sensible results in our application to the analysis of alternative parametric forms of the Engel curve relationship. We also consider the case of semiparametric estimation in partially linear models. This is important in cases where it is felt that ceratain regressors are likely to enter with a simple linear form. In this situation we can use the Robinson (1988a) (see also Ai and McFadden 1997) approach to semiparametric regression. This turns out to be an effective mechanism for analyzing demographic variables in our Engel curve application. Within the overall semiparametric framework we also consider the case of endogenous x. This extension is developed by adapting the Holly and Sargan (1982) augmented regression approach to the partially linear regression context. In the application using income to instrument total expendi-

1. Requests should be made to Alan Duncan (asd1@york.ac.uk).

ture, it is found to have an important impact on the curvature of the Engel curve relationship. The tests for a parametric specification against a nonparametric alternative are also implemented in this partially linear semiparametric context.

The rest of the paper is organized as follows. In Section II we review briefly the technique of kernel density estimation, before moving in Section III to a discussion of kernel regression methods. We describe the form and properties of the standard Nadaraya-Watson kernel estimator, including a discussion of the sampling distribution of the local regression estimator—and a summary of proposed tests for assessing parametric models against nonparametric alternatives. We present an application of the Engel curve analysis of food expenditure and alcohol expenditure for a large sample of households in the British Family Expenditure Survey. These two commodities display very different nonparametric Engel curve shapes and therefore provide an excellent example of the benefits of implementing nonparametric methods in microeconomic data. In Section IV attention is turned to semiparametric regression and a method to correct for endogeneity in the context of kernel regression. Section V considers the advantages of local polynomial regression and finally Section VI concludes.

II. Kernel Density Estimation

The general form for the kernel density estimator of a P-dimensional variable x is

$$(2.1)\quad \hat{f}_H(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x_i - x)$$

where $K_H(x) = \det(H)^{-1} \cdot K(H^{-1} \cdot x)$ for some multivariate kernel function $K(x)$ and for a given $P \times P$ matrix of bandwidths H. The simplest multivariate kernel is a product of univariate kernels of the form $K(x) = \prod_{p=1}^{P} k(x_p)$, and a typical matrix of bandwidths would be either diagonal or related to the sample covariance S of the variable x, such that $H = h \cdot S^{1/2}$ for some positive scalar h.

When x is univariate, the kernel density estimator (2.1) reduces to

$$(2.2)\quad \hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x)$$

where $K_h(x) = h^{-1} \cdot k(x/h)$ for a scalar bandwidth h.[2] For notational simplicity we concentrate here on univariate estimation techniques, although multivariate extensions are straightforward. From (2.2) we see that the kernel density estimator evaluated at x for a given bandwidth is simply a weighted average of the raw data, with greater weight given to observations close to the point at which the density is estimated. The kernel function itself is symmetric, integrates to unity, and is typically continuously differentiable.

Common choices for the univariate kernel function include the Gaussian in which $k(u) = 1/\sqrt{2\pi}\,\exp(-u^2/2)$ and the Epanechnikov in which $k(u) = 3/4(1 - u^2) \cdot$

2. See Rosenblatt (1956) and Parzen (1962).

$1(|u| \leq 1)$ where $1(\cdot)$ is the unit indicator function. Notice that the Epanechnikov kernel truncates the points local to x when calculating $f(x)$ whereas the normal kernel uses all observations in the calculation of the conditional mean at each point. As a consequence, truncating the normal kernel to reduce the computational burden is common. Useful discussions of other kernel functions can be found in Härdle and Linton (1994) or Silverman (1986).

The choice of bandwidth h is crucial to the appearance and properties of the final density estimate. While the choice can be a purely subjective one, it is common to apply certain rules of thumb. One such rule, as discussed in Silverman (1986), sets bandwidths to minimize the Mean Integrated Squared Error (MISE) of the form

$$(2.3) \quad \text{MISE}(\hat{f}(x)) = \int E[\{\hat{f}(x) - f(x)\}^2] \cdot w(x) \cdot dx$$

where $f(x)$ denotes the "true" density and $w(x)$ denotes some trimming function. If the true density of x is normal then the optimal MISE choice for h in estimating the density $f(x)$ is given by $1.06\ \sigma_x n^{-1/5}$.

An Application

As an illustration of kernel density estimation we can turn to the total budget variable in our Engel curve analysis. Typically the total expenditure variable, log transformed, is often supposed to have a normal cross-section distribution. To see the usefulness of the kernel method Figure 1 presents kernel density estimates of log expenditure for a group of around 1,000 households from a single year of the UK Family Expenditure Survey. These are married couples with no children so as to keep a reasonable degree of homogeneity in the demographic structure (see Banks, Blundell, and Lewbel 1997). A Gaussian kernel is used, with the bandwidth chosen according to Silverman's rule of thumb for a normal density.

The results are interesting, showing that it is relatively difficult to tell apart the nonparametric density from the fitted normal curve which is also shown. The bivariate kernel density plot in Figure 2 is also based on a Gaussian kernel with optimal bandwidth according to Silverman's rule for a normal density. The joint nonparametric density of food expenditure share and log total expenditure again seems close to bivariate normal, with strong negative correlation.

III. Kernel Regression

The aim of kernel regression is to replace $g(x)$ in (1.1) by a local estimator of the conditional mean

$$(3.1) \quad E(y|x) = \int yf(y|x)dy$$

where $f(y|x)$ is the conditional density of y. Noting that $f(y|x) = f(x, y)/f(x)$ and $f(x) = \int f(y, x)dy$, we can rewrite (3.1) as

$$(3.2) \quad E(y|x) = \frac{\int yf(y, x)dy}{\int f(y, x)dy}.$$

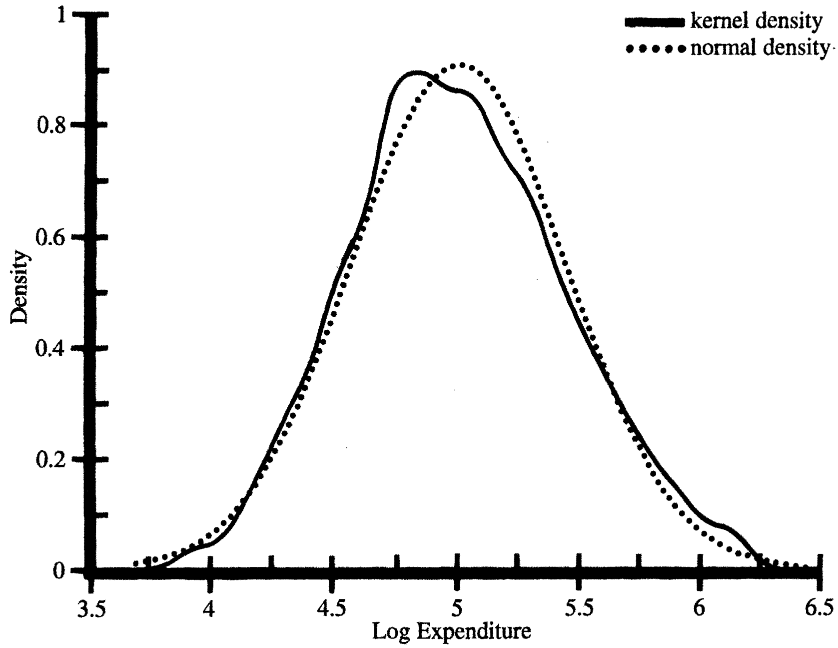


Figure 1
The Density of Log-Expenditure, FES 1980–82
Source: Banks, Blundell, and Lewbel (1997).

The objective of kernel regression is to replace the numerator and denominator in (3.2) with estimators based on locally weighted averages. Specifically, by analogy with (2.2), we can write the Nadaraya-Watson kernel estimator of (3.2) as[3]

$$(3.3)\quad \hat{g}_h(x) = \frac{\sum_{i=1}^{n} y_i K_h(x_i - x)}{n \cdot \hat{f}_h(x)}$$

or equivalently

$$(3.4)\quad \hat{g}_h(x) = \sum_{i=1}^{n} y_i \cdot \pi_{ih}(x)$$

when expressed in terms of a weight function $\pi_{ih}(x)$ of the form

$$(3.5)\quad \pi_{ih}(x) = \frac{K_h(x_i - x)}{\sum_{j=1}^{n} K_h(x_j - x)}.$$

3. See Nadaraya (1964) and Watson (1964).

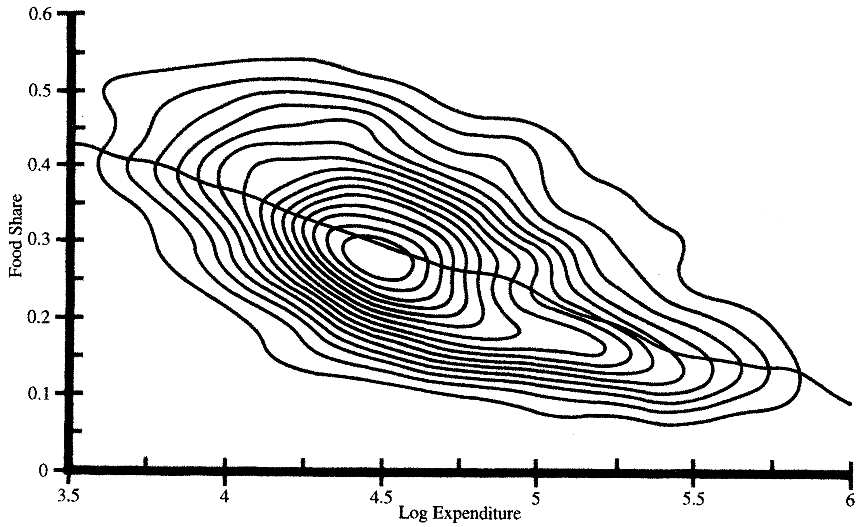


Figure 2
Bivariate Kernel Density: Food Share and Log Expenditure

Figures 3 and 4 present Kernel regressions for the Engel curves for the food share and the alcohol share in the FES data respectively. Each figure includes pointwise 95 percent confidence bands at the decile points in the log expenditure distribution and also plots a quadratic polynomial regression. The estimation of confidence bands is discussed below. These figures are drawn from the Banks, Blundell, and Lewbel (1997) study, where a number of other commodities are considered.

The following two theorems, described more fully in Härdle (1990) and Härdle and Linton (1994) set out conditions for the consistency and asymptotic normality of the estimator (3.3);

Theorem 3.1

For kernel weights $K(\cdot)$ and bandwidth $h = h(n)$ such that

A.1: $\int |K(u)| du < \infty$,

A.2: $\lim_{|u| \to \infty} uK(u) = 0$, and

A.3: $h \to 0$ and $nh \to \infty$ as $n \to \infty$,

then $g_H(x) \to g(x)$ at every point x at which $g(x)$ and $\sigma^2(x)$ are continuous, and $f(x)$ is continuous and positive.

Theorem 3.2

For kernel weights $K(\cdot)$ and bandwidth $h = h(n)$ satisfying A.1 to A.3 and

A.4: $\int |K(u)|^{2+\eta} du < \infty$ for some $\eta > 0$,

A.5: $\lim h^5 n < \infty$,

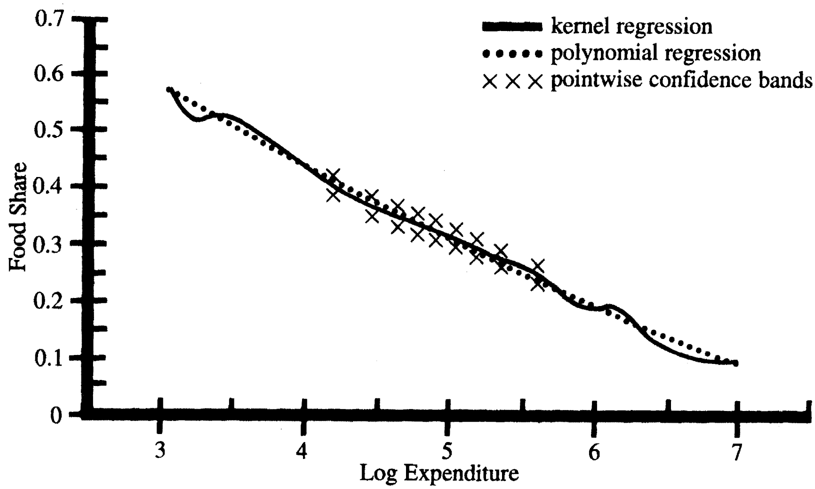


Figure 3
Nonparametric Engel Curve: Food Share
Source: Banks, Blundell, and Lewbel (1997).

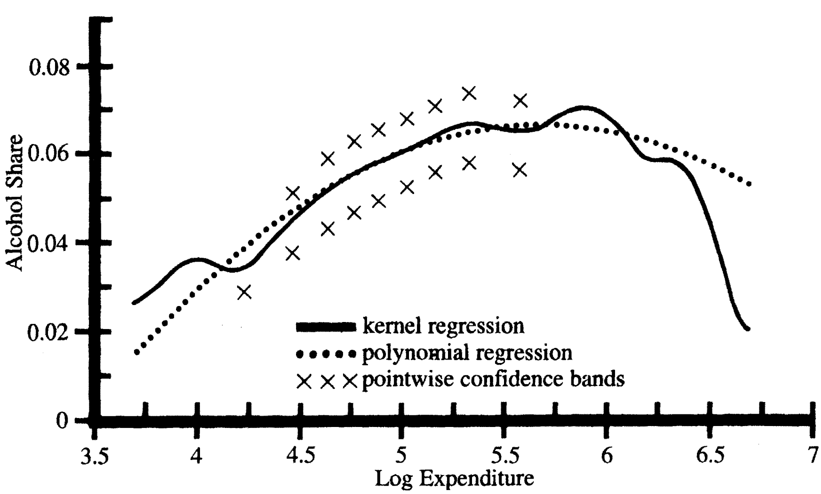


Figure 4
Nonparametric Engel Curve: Alcohol Share
Source: Banks, Blundell, and Lewbel (1997).

then at every point x at which $g(x)$ and $f(x)$ are twice continuously differentiable and positive, we have that

$$(3.6) \quad (nh)^{1/2}\left\{\frac{\hat{g}_h(x) - g(x) - h^2B(x)}{V(x)^{1/2}}\right\} \xrightarrow{L} N(0, 1)$$

where $V(x) = \sigma^2(x)c_K/f(x)$, c_K is a kernel-specific constant $(= \int K^2(u)du)$ and $B(x)$ is a nondisappearing bias.

We need to be able to place a confidence band around the estimated regression curve. A simple way to do this is to choose k fixed points in the x distribution—say the decile points. Under certain regularity conditions (see Härdle 1990), the theorem above allows us to derive pointwise confidence bands of the form

$$(3.7) \quad \left\{(nh)^{1/2}\left[\frac{\hat{g}_h(x_j) - g_h(x_j) - h^2B(x_j)}{V(x_j)^{1/2}}\right]\right\}_{j=1}^{k} \xrightarrow{L} N(0, 1).$$

Although the Working-Leser linear logarithmic formulation appears to provide a reasonable approximation for the food share curve, distinct nonlinear behavior is evident for alcohol. For the alcohol share a quadratic logarithmic share model would seem to fit quite well, confirming the results in Banks, Blundell, and Lewbel (1997). That study also found considerable stability in these overall patterns across the time series of Family Expenditure Surveys.

A. *Average Derivative Estimation*

Economists are often interested in the "slope" of the regression line; that is, the marginal effects on the conditional expectation $g(x) = E(y|x)$ of a change in x. Following Stoker (1991), the kernel regression function (3.4) may be rewritten as

$$(3.8) \quad \hat{g}_h(x) = \hat{c}_h(x)/\hat{f}_h(x)$$

where

$$(3.9) \quad \hat{c}_h(x) = n^{-1} \cdot \sum_{i=1}^{n} y_i \cdot K_h(x_i - x),$$

and, provided that the underlying kernel $K(\cdot)$ is first-order differentiable, the local slope estimator $g'_H(x)$ can be derived as

$$(3.9) \quad \hat{g}'_h(x) = \frac{\hat{c}'_h(x)}{\hat{f}_h(x)} - \hat{g}_h(x) \cdot \frac{\hat{f}'_h(x)}{\hat{f}_h(x)}.$$

For a nonparametric regression, the slopes (3.9) vary over the range of x. Typically, however, interest centers on the nature of the "average" slope $\delta = E[g'(x)]$ of a nonparametric line.

One solution is to estimate directly the average slope $\delta = E[g'_H(x)]$ as a simple (trimmed) sample average, either through the *direct average derivative estimator*

$$(3.10) \quad \hat{\delta}_{DA} = n^{-1} \sum_{i=1}^{n} I_i \cdot \hat{g}'_h(x_i),$$

or the *indirect average derivative estimator*

$$(3.11) \quad \hat{\delta}_{IA} = n^{-1} \sum_{i=1}^{n} I_i \cdot [-\hat{f}'_h(x_i)/\hat{f}_h(x_i)] \cdot y_i,$$

where $I_i = 1(\hat{f}_H(x_i) \geq b)$ for some small b such that $b \to 0$ as $n \to \infty$. Noting that the scales of either $\hat{\delta}_{DA}$ or $\hat{\delta}_{IA}$ are not directly comparable to that of the standard functional derivative, Stoker (1986) discusses an alternative approach to the estimation of average derivatives based on an assumption that the conditional expectation $E(y|x)$ is related to x only through the linear index $x^T\beta$. In particular, if the regression functional can be represented as a *single index* of the form

$$(3.12) \quad E(y|x) = g(x) = M(x^T\beta)$$

then the average slope

$$(3.13) \quad \delta = E[g'(x)] = E[M'(x^T\beta)] \cdot \beta$$

identifies the vector β up to scale. We may absorb the scale of $M(\cdot)$ through the normalization $E(M') = 1$. By defining $m(s) = M(\gamma s)$ for $\gamma = E[M'(x^T\beta)]^{-1}$, the coefficients δ of the scaled index model

$$(3.14) \quad E(y|x) = m(x^T\delta)$$

may be compared directly with those of a linear model.[4]

One can estimate δ as the linear coefficients from an instrumental variables regression of y on x using $\hat{l}_h(x) = -\hat{f}'_h(x)/\hat{f}_h(x)$ as instruments, giving an *indirect slope estimator* δ_{IS} (compare Stoker (1991) p. 63) of the form

$$(3.15) \quad \hat{\delta}_{IS} = \left(\sum_{i=1}^{n} I_i \cdot \hat{l}_h(x_i) \cdot (x_i - \bar{x})\right)^{-1} \cdot \sum_{i=1}^{n} I_i \cdot \hat{l}_h(x_i) \cdot (y_i - \bar{y}).$$

B. Choice of Bandwidth

A central issue in nonparametric estimation by local smoothing is the choice of bandwidth. With a fixed sample the size of bandwidth h determines the degree of smoothing and is therefore of crucial importance for the appearance, interpretation and properties of the final estimate. The choice of bandwidth can be a purely subjective choice, it can relate to some ''rule of thumb,'' or the choice can in some sense be ''automated'' by data-driven cross-validation techniques. The degree of smoothing corresponding to a given bandwidth parameter h is clearly not independent of the

4. See Härdle-Stoker (1989) or Stoker (1991) for a detailed discussion of the asymptotic distribution of various average derivative estimators.

kernel function. The relationship between bandwidths in each case is proportional, however, and a useful list of bandwidth "exchange rates" is given in the Härdle and Linton paper.

The choice of bandwidth involves an implicit tradeoff between the variance and bias of the kernel estimator $\hat{g}_h(x)$ of $g(x)$. As Härdle and Linton (1994) note, the mean squared error (MSE) of $\hat{g}_h(x)$ may be approximated via (3.6) by

$$(3.16)\quad \mathrm{MSE}[\hat{g}_h(x)] = E[(\hat{g}_h(x) - g(x))^2] \approx (nh)^{-1}V(x) + h^4B^2(x).$$

So the contribution to the MSE from the variance (squared bias) term decreases (increases) as h increases, and the bandwidth that minimizes the theoretical MSE may be derived straightforwardly as

$$(3.17)\quad h^*_{MSE} = \left(\frac{V(x)}{4B^2(x)}\right)^{1/5} n^{-1/5}.$$

The general procedure for automated bandwidth selection in practical kernel regression essentially involves choosing that h that minimizes a sample approximation to (3.16). The simplest approach is termed the *resubstitution* method, whereby h is chosen to minimize an objective function of the form

$$\frac{1}{n}\sum_{j=1}^{n} w(x_j)(y_j - \hat{g}_h(x_j))^2$$

for some trimming function $w(x_j)$. For technical reasons, however, the minimizing value of h from resubstitution is downward biased. If, for example, we use a kernel $K(\cdot)$ supported on $[-1, 1]$, then this objective function reduces to zero for any h smaller than the closest two data points in the sample.[5]

To correct this bias, two main methods are typically employed. The first uses a cross-validation (CV) statistic formed from a weighted sum of squared deviations of each y_j from its conditional mean seen as a function of h;

$$(3.18)\quad CV_1(h) = \frac{1}{n}\sum_{j=1}^{n} w(x_j)(y_j - \hat{g}_{h,j}(x_j))^2,$$

where the conditional mean $\hat{g}_{h,j}(x_j)$ in each case is calculated by leaving out the jth observation. That is,

$$\hat{g}_{h,j}(x_j) = \sum_{i\neq j} y_i \cdot \pi_{ih}^{-j}(x_j)$$

where

$$\pi_{ih}^{-j}(x_j) = \frac{K_h(x_i - x_j)}{\sum_{k\neq j}^{n} K_h(x_k - x_j)}.$$

5. We thank an anonymous referee for this observation.

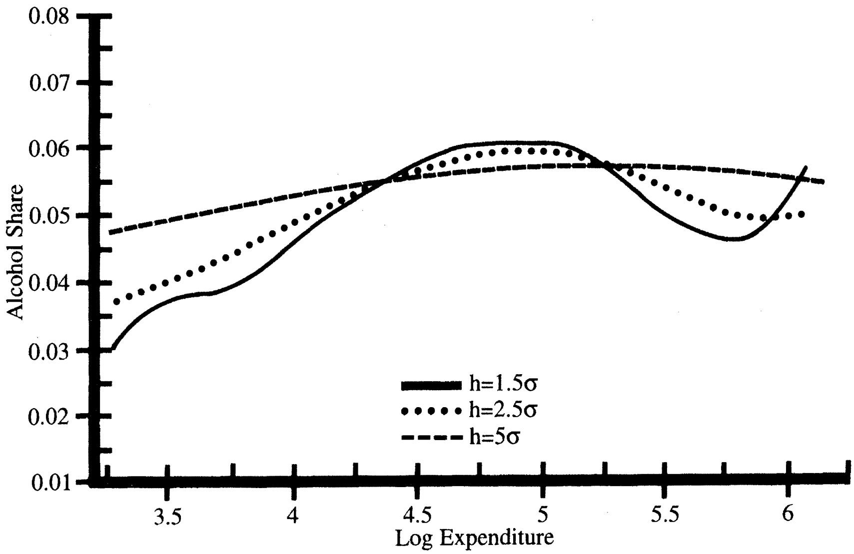


Figure 5
Sensitivity to Bandwidth Selection: Alcohol Engel Curve

By "leaving-out" the jth observation, the cross-validated bandwidth can be demonstrated to be asymptotically optimal with respect to the MSE.

Alternatively, one can employ a penalty function to correct the downward bias associated with the resubstitution method. The objective function for this second approach takes the general form

$$(3.19)\quad CV_2(h) = \frac{1}{n}\sum_{j=1}^{n} w(x_j)(y_j - \hat{g}_h(x_j))^2 \cdot p(\pi_{jh})$$

where $p(\cdot)$ is a penalty function that stops h from becoming too small. Common choices for the form of the penalty function include $p(u) = (1 - u)^{-2}$ leading to a bandwidth selected by Generalized Crossvalidation (see Craven and Wahba 1979), $p(u) = \exp(2u)$ for Akaike's (1970) Information Criterion, or $p(u) = (1 - 2u)^{-1}$ for Rice's (1984) bandwidth selector. Again, it can be demonstrated that bandwidths chosen using objective (3.19) are asymptotically optimal.

Figure 5 investigates the sensitivity of the alcohol Engel curve to variations in the bandwidth. The overall shape of the kernel regression is little affected by variations in the choice of kernel or smoothing parameter at or close to the cross-validated level. When the bandwidth becomes too large, however, the curve flattens.[6] Thus, care must be taken when choosing bandwidths subjectively.

An attractive addition to this method of choosing the bandwidth is to allow the bandwidth to vary with the density of x. In the application below this is shown to

6. In the limit, the Nadaraya-Watson curve becomes horizontal (at the mean) as h becomes arbitrarily large.

smooth out "wiggles" very effectively in areas where the density of x is sparse. If we let

$$\lambda_i = \left[\frac{f_{h_p}(x_i)}{\eta}\right]^{-\rho} \text{ where } 0 \leq \rho \leq 1,$$

for some pilot bandwidth h_P with the normalisation factor η given by

$$\ln \eta = \frac{\sum_j \ln \hat{f}_{h_p}(x_j)}{n},$$

then the *adaptive kernel* estimator takes the form

$$\text{(3.20)} \quad \hat{f}_h^A(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h\lambda_j} K((x_j - x)/\lambda_j h)$$

for the density estimator with a corresponding form for the conditional mean. (See Silverman 1986, for a discussion of the adaptive kernel method).

Figure 6 gives an example of the usefulness of the adaptive kernel. The upper panel plots a nonparametric food Engel curve using the Nadaraya-Watson method, while the lower panel plots the same, but using adaptive kernel techniques. Notice how the "wiggles" in the Engel curves that appear at the edges of the x distribution are essentially ironed out by adaptive estimation.

C. Comparisons with Parametric Specifications

Inference in nonparametric regression can take place in a number of ways. Perhaps the most obvious and the one at the frontier of current research activity is to use the nonparametric regression as an alternative against which to test a parametric null. Recent work by Aït-Sahalia, Bickel, and Stoker (1994), Ellison and Ellison (1992), Härdle and Mammen (1993), and Zheng (1996) is particularly notable.

Aït-Sahalia et al. derive asymptotically normal statistics for the comparison between a nonparametric estimate $\hat{g}_h(x_i)$ and some parametric estimate $\delta(x_i, \hat{\beta})$ of a regression curve based on a simple squared error goodness-of-fit statistic

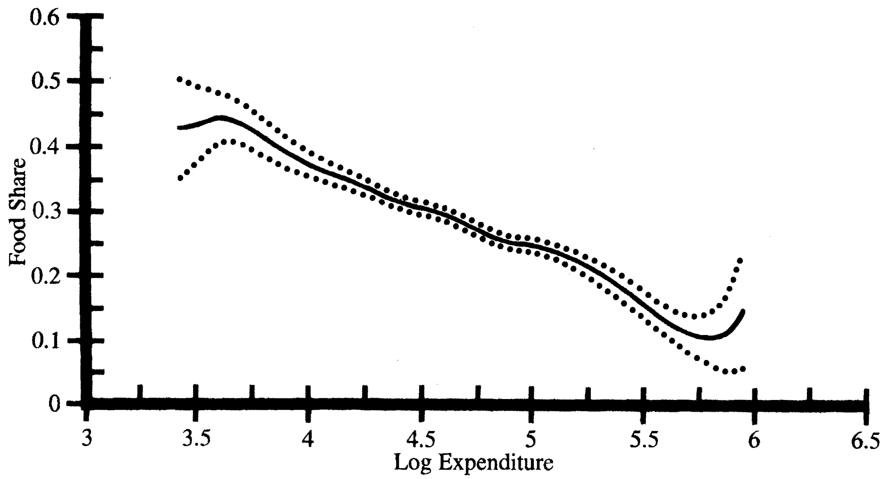
$$\text{(3.21)} \quad \hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} (\hat{g}_h(x_i) - \delta(x_i, \hat{\beta}))^2 w(\hat{f}_h(x_i)),$$

a linear transformation of which is shown to converge at rate $nh^{M/2}$ to a limiting normal distribution with mean zero and estimable variance.

An alternative approach by Zheng (1996) uses the kernel method to construct a moment condition that can be used to distinguish the parametric null from the nonparametric alternative. The test centers around a matrix W_n of kernel weights with typical elements

$$w_{ij} = 1(i \neq j) \cdot K\left(\frac{(x_i - x_j)}{h}\right)$$

where $1(\cdot)$ represents the indicator function. Given a set of parametric residuals $\hat{e}_i = y_i - \delta(x_i, \hat{\beta})$, the statistic



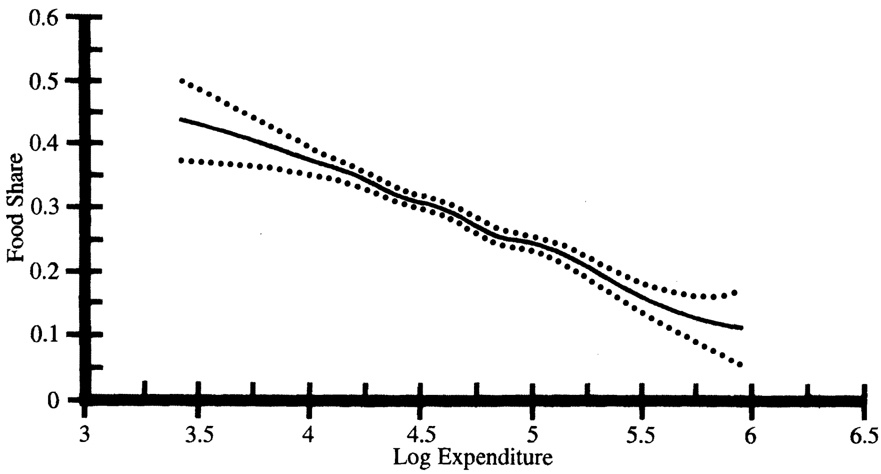


Figure 6
Normal Kernel and Adaptive Kernel Regressions: Food Engel Curves

$$(3.22)\quad T_n = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\hat{e}_i\hat{e}_j}{\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} 2w_{ij}^2(\hat{e}_i\hat{e}_j)^2\right\}^{1/2}}$$

is shown to be asymptotically standard normal under the null and consistent against all deviations from the parametric null, with a convergence rate $nh^{M/2}$.

The test proposed by Ellison and Ellison (1992) has a structure almost identical

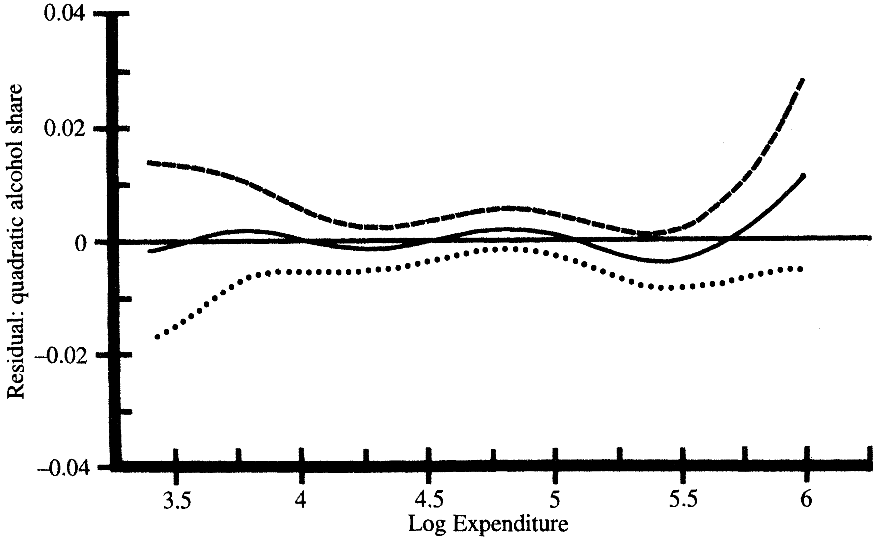


Figure 7
A Residual Based Specification Test: Quadratic Model for Alcohol Share

to that of Zheng (1996), and differs only in the replacement of the residual cross-product term $(\hat{e}_i, \hat{e}_j)^2$ by the variance estimator $\hat{\sigma}^2 = 1/n \sum \hat{e}_i^2$. The authors quote a convergence rate that depends on the properties of the kernel weight matrix.[7]

D. An Application

As an illustration it is interesting to focus on a comparison of the nonparametric Engel curve estimates in Figures 3 and 4 with the simple second order polynomial fit given by the dashed line in the figures. Some guide to the reliability of this approximation can be drawn from the pointwise confidence intervals (evaluated at the mid-points of each decile) also shown in the figures. It is only where the data are sparse and the confidence bands relatively wide that the paths diverge. This can be examined more closely by a residual-based analysis adopted in Banks, Blundell, and Lewbel (1997). In Figure 7 a nonparametric regression curve is presented between the residuals from the parametric quadratic regression curve and the log total expenditure—the X_i variable in the above kernel regressions. A well-specified parametric model should display a line through zero. This is seen to be pretty much the case for our quadratic logarithmic Engel curve for alcohol shares.

For a more formal comparison, we derive Average Derivative estimates and a range of nonparametric specification tests for the food and alcohol Engel curves

7. Other important tests are given in Bierens (1990) and Wooldridge (1992).

Table 1
Nonparametric Estimates: Food and Alcohol Engel Curves

	(1) Food	(2) Alcohol
$\hat{\delta}_{IS}$	−0.1341	0.0004
	(0.0073)	(0.0020)
$\hat{\beta}_{OLS}$	−0.1394	0.0009
	(0.0071)	(0.0034)
H_0: linear parametric form		
χ^2_Z	0.014	5.910
	[0.906]	[0.015]
χ^2_{EE}	0.014	6.973
	[0.906]	[0.008]
χ^2_{ABS}	1.679	4.633
	[0.195]	[0.031]
H_0: quadratic parametric form		
χ^2_Z	0.004	0.404
	[0.948]	[0.525]
χ^2_{EE}	0.004	0.481
	[0.948]	[0.488]
χ^2_{ABS}	0.567	0.526
	[0.451]	[0.468]

Notes: Data are drawn from the 1980–1982 Family Expenditure Surveys. Nonparametric estimates based on a Gaussian kernel with bandwidths chosen by Least Squares crossvalidation [compare equation (3.18)]. Average derivatives $\hat{\delta}_{IS}$ are indirect slope estimates [compare equation (3.15)]. For crossvalidation and ADE calculations, data are trimmed to exclude the smallest 2 percent of estimated densities. All estimates and specification tests are generated using the GAUSS-based software package NP-REG (see Duncan and Jones (1992)).

shown in Figures 3 and 4. These are presented in Columns 1 and 2 of Table 1. The specification tests also reported in the table compare the general nonparametric regressions with both linear and quadratic parametric forms, using the statistics of Zheng (1996) (denoted χ^2_z), Ellison and Ellison (1992) (denoted χ^2_{EE}) and Aït-Sahalia, Bickel, and Stoker (1994) (denoted χ^2_{ABS}). All are distributed as a chi-squared under the null. We are unable to reject linearity for the food share equation, whereas for alcohol share, a quadratic specification enjoys a good degree of empirical support.

IV. Semiparametric Regression and Endogenous Regressors

A. Partially Linear Models

It will often be useful to consider extensions of (1.1) that include a linear parametric part. In Engel curve analysis, for example, it may be useful to add a set of demographic characteristics to the conditional mean specification. The regression specification now has the form

(4.1) $y = g(x) + z'\gamma + \varepsilon$

in which $z'\gamma$ represents a linear index in terms of a finite vector of observable exogenous regressors z and unknown parameters γ. We assume $E(\varepsilon|z, x) = 0$ and $\mathrm{Var}(\varepsilon|z, x) = \sigma^2(z, x)$. This model is typically labeled *semiparametric*. It can be the case that the vector of coefficients γ is the parameter of interest and an estimator of γ is required that is robust to an unknown form for $g(x)$. Following Robinson (1988), a simple transformation of the model can be used to give an estimator for γ. Taking expectations of (4.1) conditional on x, and substracting from (4.1) yields

(4.2) $y - E(y|x) = (z - E(z|x))'\gamma + \varepsilon.$

Replacing $E(y|x)$ and $E(z|x)$ by their nonparametric estimators, denoted $\hat{m}_h^y(x)$ and $\hat{m}_h^z(x)$ respectively, the ordinary least squares estimator for γ is $\sqrt{n}$ consistent and asymptotically normal.

The estimator for $g(x)$ is then simply

(4.3) $\hat{g}_h(x) = \hat{m}_h^y(x) - \hat{m}_h^z(x)'\hat{\gamma}.$

Because $\hat{\gamma}$ converges at $\sqrt{n}$, the asymptotic distribution results for $\hat{g}_h(x)$ remain unaffected by estimation of γ and follows from the distribution of $\hat{m}_h^y(x) - \hat{m}_h^z(x)$.[8]

B. Correcting for Endogeneity

Suppose x is endogenous in the model (1.1) in the sense that

(4.4) $E(\varepsilon|x) \neq 0$ or $E(y|x) \neq g(x)$.

In this case

$$\hat{g}_h(x) \overset{p}{\nrightarrow} g(x)$$

so that the nonparametric estimator will not be consistent.

However, suppose there exists a variable z such that

$$x = \pi z + v \text{ with } E(v|z) = 0.$$

Moreover, assume the following linear conditional model holds

(4.5) $y = g(x) + v\rho + \varepsilon$

with

(4.6) $E(\varepsilon|x) = 0.$

In this case the following semiparametric estimator above can be used to mimic the augmented regression approach developed by Holly and Sargan (1982). Note that

8. Heckman et al. (1995) show this asymptotic distribution result can provide a poor approximation even in moderately sized samples, and implement bootstrap methods which perform well in Monte Carlo comparisons. They also provide an asymptotic variance estimator which accounts for the estimation of γ.

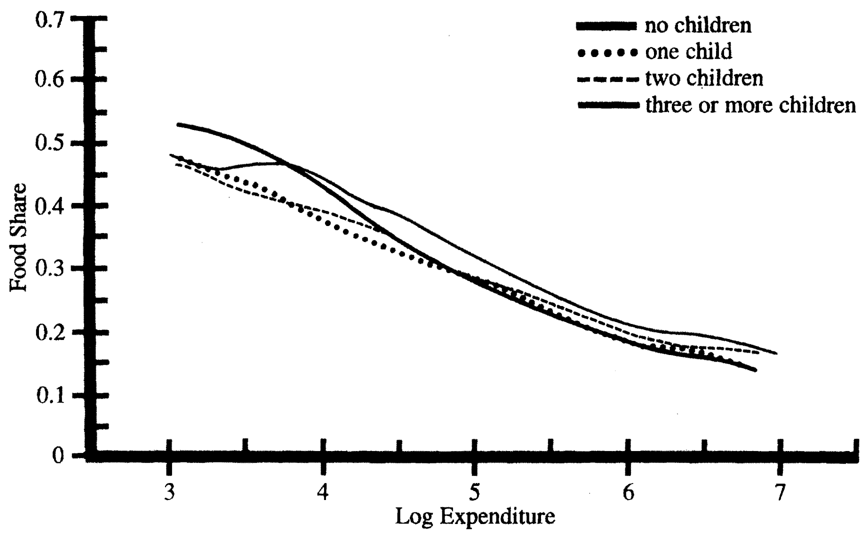


Figure 8
Nonparametric Food Share by Family Size: Adaptive Kernel

(4.7) $y - E(y|x) = (v - E(v|x))\rho + \varepsilon.$

The estimator of $g(x)$ is given by

(4.8) $\hat{g}_h(x) = \hat{m}_h^y(x) - \hat{m}_h^v(x)\hat{\rho}.$

In place of the unobservable error component v we use the first stage residuals

(4.9) $\hat{v} = x - z\hat{\pi}$

where $\hat{\pi}$ is the least squares estimator of π. Because $\hat{\pi}$ and $\hat{\rho}$ converge at $\sqrt{n}$, the asymptotic distribution for $\hat{g}_h(x)$ follows the distribution of $\hat{m}_h^y(x) - \hat{m}_h^v(x)\rho$. Moreover, a test of the exogeneity null $H_0\!:\rho = 0$, can be constructed from this least squares regression.

Newey, Powell, and Vella (1995) have developed a fully nonparametric implementation of this idea for triangular simultaneous equation systems of the type considered here. They adopt a series approach to the estimation of the regression of y on x and v. This generalizes the form of (4.5) and allows an assessment of the additive structure.

C. An Application

To see the importance of a semiparametric transformation of the form described above, we now consider food and alcohol shares among a more heterogeneous group of married women with at least one dependent child. Figure 8 presents adaptive kernel regression lines for the food Engel curve across four different household types

Table 2
Nonparametric and Semiparametric Estimates: Food Engel Curves

	(1)	(2)	(3)	(4)
$\hat{\delta}_{IS}$	−0.1341	−0.1441	−0.1087	−0.1396
	(0.0073)	(0.0091)	(0.0080)	(0.0074)
$\hat{\beta}_{OLS}$	−0.1394	−0.1479	−0.1140	−0.1434
	(0.0071)	(0.0069)	(0.0071)	(0.0069)
$\hat{\gamma}_k$	—	0.0229	—	0.0225
		(0.0032)		(0.0033)
$\hat{\delta}_v$	—	—	−0.0361	−0.0062
			(0.0148)	(0.0151)
H_0: linear parametric form				
χ^2_Z	0.014	0.164	0.003	0.171
	[0.906]	[0.686]	[0.991]	[0.680]
χ^2_{EE}	0.014	0.166	0.003	0.172
	[0.906]	[0.684]	[0.991]	[0.678]
χ^2_{ABS}	1.679	0.772	0.741	0.675
	[0.195]	[0.380]	[0.389]	[0.411]
H_0: quadratic parametric form				
χ^2_Z	0.004	0.142	0.001	0.149
	[0.948]	[0.706]	[0.991]	[0.700]
χ^2_{EE}	0.004	0.144	0.001	0.150
	[0.948]	[0.705]	[0.991]	[0.698]
χ^2_{ABS}	0.567	0.448	0.185	0.375
	[0.451]	[0.503]	[0.667]	[0.540]

Notes: Data: 1980–82 Family Expenditure Surveys sample size 1,025. Nonparametric estimates based on a Gaussian kernel with bandwidths chosen by Least Squares crossvalidation [compare equation (3.18)]. Average derivatives $\hat{\delta}_{IS}$ are indirect slope estimates [compare equation (3.15)]. For crossvalidation and ADE calculations, data are trimmed to exclude the smallest 2 percent of estimated densities. Parameters $\hat{\gamma}_k$ and $\hat{\delta}_v$ are estimated by OLS (compare equation 4.2). All estimates and specification tests are generated using the GAUSS-based software package NP-REG (see Duncan and Jones 1992).

split according to the number of children, and demonstrates a degree of heterogeneity in share equations sufficient to consider semiparametric estimation techniques.

Following Robinson (1988) we examine the extent of misspecification in nonparametric Engel curves among households with at least one dependent child. The most general semiparametric specification controls both for the number of children n_k in the household, and for potential endogeneity in log expenditure via first-stage residual $\hat{v}$ in (4.5). To calculate this residual the log of disposable income is used as the excluded instrumental variable. We write

$$(4.10) \quad y = g(x) + \gamma_k \cdot n_k + \rho_v \cdot \hat{v} + \varepsilon$$

where y represents expenditure share and x denotes log real expenditure.

Tables 2 and 3 present nonparametric average derivative estimates and specifica-

Table 3
Nonparametric and Semiparametric Estimates: Alcohol Engel Curves

	(1)	(2)	(3)	(4)
$\hat{\delta}_{IS}$	0.0004	0.0018	−0.0150	−0.0100
	(0.0020)	(0.0022)	(0.0024)	(0.0024)
$\hat{\beta}_{OLS}$	−0.0009	0.0005	−0.0162	−0.0113
	(0.0034)	(0.0035)	(0.0035)	(0.0035)
$\hat{\gamma}_k$	—	−0.0049	—	−0.0038
		(0.0016)		(0.0017)
$\hat{\delta}_v$	—	—	0.0213	0.0162
			(0.0073)	(0.0076)
H_0: linear parametric form				
χ^2_Z	5.910	6.384	5.677	6.502
	[0.015]	[0.012]	[0.017]	[0.011]
χ^2_{EE}	6.973	7.597	6.695	7.685
	[0.008]	[0.006]	[0.010]	[0.006]
χ^2_{ABS}	4.633	4.297	11.609	8.696
	[0.031]	[0.038]	[0.001]	[0.003]
H_0: quadratic parametric form				
χ^2_Z	0.404	0.370	0.248	0.300
	[0.525]	[0.543]	[0.618]	[0.584]
χ^2_{EE}	0.481	0.443	0.294	0.356
	[0.488]	[0.506]	[0.588]	[0.551]
χ^2_{ABS}	0.526	0.134	1.484	0.736
	[0.468]	[0.715]	[0.223]	[0.391]

Notes: as for Table 2.

tion tests for food and alcohol share equations. The first columns of each table reproduce the numbers in Table 1. Comparing this most general specification [marked (4) in Tables 2 and 3] with more restricted versions. Food expenditure share is estimated to increase with the number of children, with some evidence of endogeneity in log total expenditure. Compared with the most restrictive specification (1), the average slope of the food Engel curve becomes more negative when controlled for heterogeneity (2), and less so once corrected for endogeneity (3). In all specifications, we are unable to reject linearity. For alcohol share, the converse is true; alcohol expenditure decreases slightly as a share of total expenditure as the number of children in the household rises. The endogeneity correction typically forces the slope of Engel curve to be more negative. In line with Banks, Blundell, and Lewbel (1997) we find it important to include quadratic terms in log expenditure in the parametric alcohol share equation.

A graphical comparison of the food and alcohol Engel curves under alternative specifications can be seen in Figures 9 and 10. Note that in all regressions the depen-

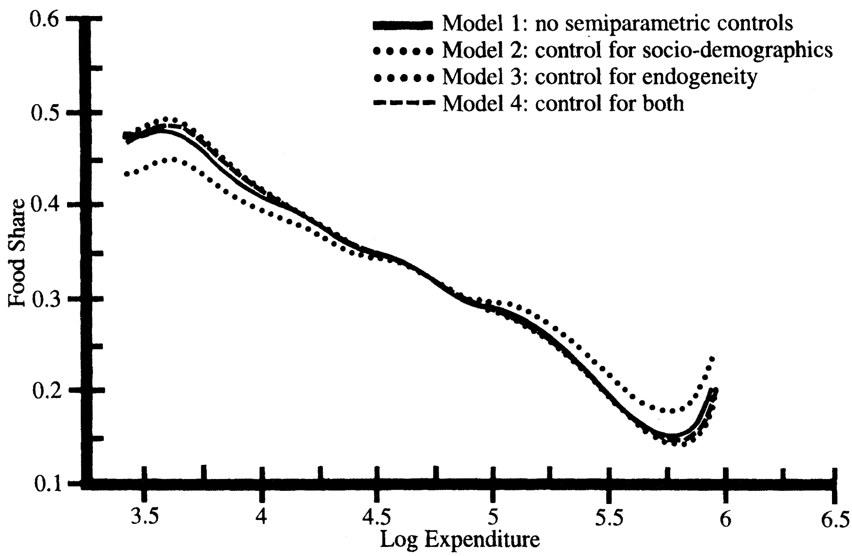


Figure 9
Nonparametric and Semiparametric Engel Curves: Food

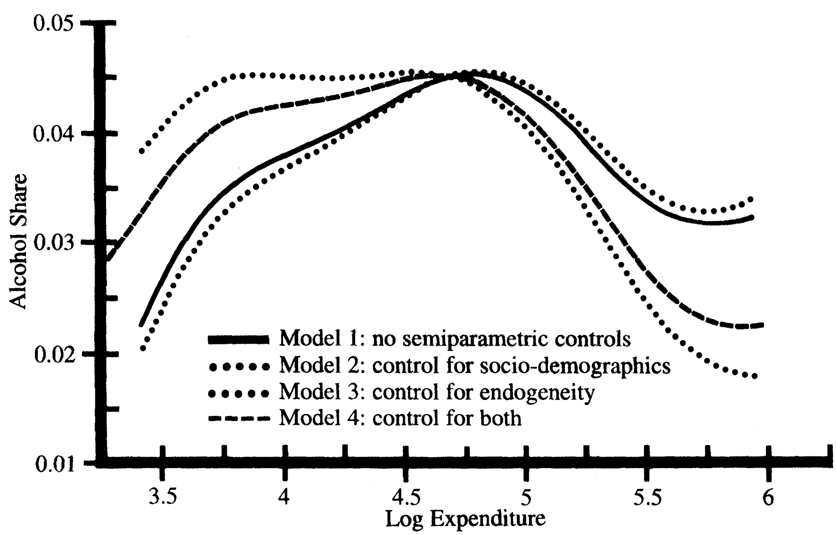


Figure 10
Nonparametric and Semiparametric Engel Curves: Alcohol

dent variable is normalized around $\hat{\gamma}_k \cdot \bar{n}_k$. For the second case, the impact of the semiparametric controls (for endogeneity in particular) is dramatic.

V. Local Polynomial Regression

The kernel regression method analyzed so far can be thought of fitting a sequence of local constants. Although this discussion has not focused on bias, the Nadaraya-Watson kernel estimator is only exactly unbiased when the function being estimated is a constant. Typically in economic applications there is better prior information on the shape of the curve. For example, in the food share Engel curve discussed above, local linearity would clearly be a good approximation. For the alcohol share Engel curve, a local quadratic model would be appropriate. Fan (1992) derived the statistical properties of the local linear estimator which involves fitting a line locally. This is shown to be superior to the Nadaraya-Watson regression. Fan and Gijbels (1992) consider the case of local polynomial estimators. These are discussed in Härdle and Linton (1994). In an interesting paper, Linton and Gozalo (1996) discuss a related kernel nonparametric regression estimator that can be centered at any parametric regression model. It retains all the attractive features of the locally linear estimator but has especially good properties at or near the parametric model. This relates to the centering approaches taken in Ansley, Kohn, and Wong (1993) and Fenton and Gallant (1986) for spline regression and series estimation respectively.

The Gonzalo and Linton estimator can be thought of as the minimizer of the following nonlinear least squares criterion

$$(5.1) \quad S_n(x, \alpha) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - m(x_i, \alpha)\}^2 K_h(x_i - x).$$

for some parametric model $m(x_i, \alpha)$. First S is minimized with respect to α, then for any $\hat{\alpha}_n$

$$(5.2) \quad \hat{g}_h(x) = m\{x, \hat{\alpha}_n(x)\}$$

is the estimator of $g(x)$. Note that, in general, iterative methods are required to find $\hat{g}_h(x)$ and $\hat{\alpha}_n(x)$. However, when the parametric regression function is linear in parameters, then the minimisation problem has an explicit solution. Consider for example the parametric model $m(x_i, \alpha) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \ldots + \alpha_p x_i^p = z_i'\alpha$ where $z_i = (1, x_i, x_i^2, \ldots, x_i^p)'$ and $\alpha = (\alpha_0, \alpha_1, \alpha_2 \ldots, \alpha_p)'$. For any given h, the α that minimizes (5.1) at x can be derived as

$$(5.3) \quad \hat{\alpha}_n(x) = \left(\sum_{i=1}^{n} K_h(x_i - x) \cdot (z_i z_i')\right)^{-1} \left(\sum_{i=1}^{n} K_h(x_i - x) \cdot (z_i y_i)\right),$$

from which the local polynomial estimator $\hat{g}_h(x)$ can be recovered through (5.2). Note that, when $m(x, \alpha) = \alpha_0$, then $\hat{g}_h(x)$ reduces to the Nadaraya-Watson estimator.

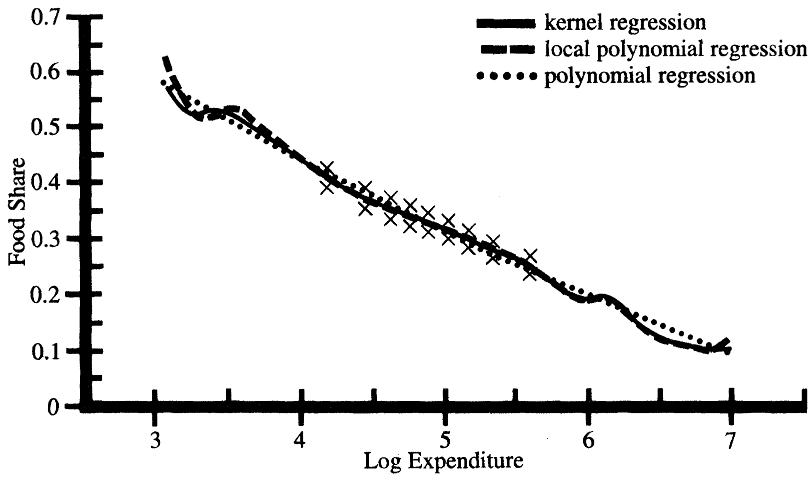


Figure 11
Local Polynomial Regression: Food Shares

Figures 11 and 12 compare Nadaraya-Watson regressions for food and alcohol share with local quadratic and parametric quadratic equivalents.

A major advantage of this approach emerges when the parametric model is true or approximately true. Gonzalo and Linton show that if for some fixed α^0, $g(x) = m(x, \alpha^0)$ for all x, then the highest order of the asymptotic bias of $\hat{g}(x)$ is arbitrarily small. This follows because the derivatives of all orders of $g(x) - m(x, \alpha^0)$ with respect to x equal zero for all x. In fact parametric asymptotic theory applies straightforwardly in this case.

We may use cross-validation techniques based on (3.18) to automate the choice of bandwidths in local polynomial regression. To generate an appropriate leave-one-out estimator of $g(x)$, note first that we may adapt (5.3) to give

$$(5.4) \quad \hat{\alpha}_n^{-j}(x_j) = \left(\sum_{i \neq j}^{n} K_h(x_i - x_j) \cdot (z_i z_i') \right)^{-1} \left(\sum_{i \neq j}^{n} K_h(x_i - x_j) \cdot (z_i y_i) \right).$$

This allows us to form the leave-one-out estimator

$$(5.5) \quad \hat{g}_h^{-j}(x_j) = m\{x, \hat{\alpha}_n^{-j}(x_j)\}$$

at the jth data point x_j. Minimizing (3.18) with respect to h using (5.5) gives a cross-validated bandwidth for local polynomial regression.

Table 4 compares Nadaraya-Watson, local linear and local quadratic regressions in terms of their Mean Squared Errors

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^{n} w(x_j)(y_j - \hat{g}_h(x_j))^2.$$

In all cases bandwidths are chosen by cross-validation. By way of comparison we

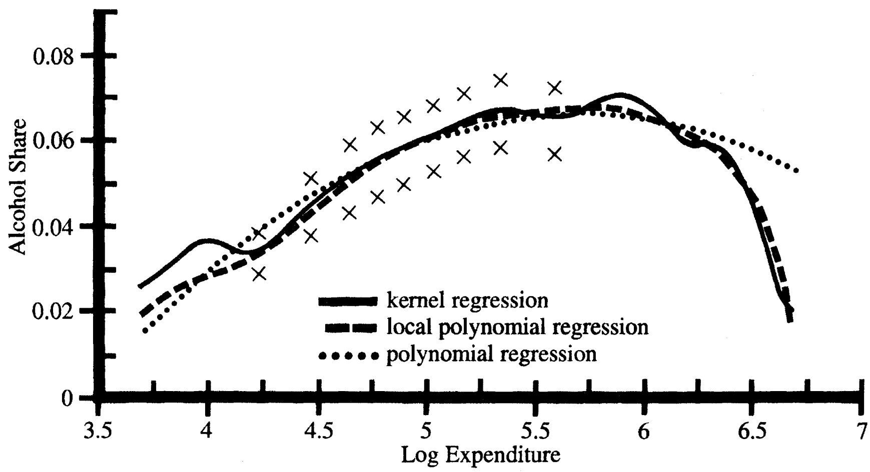


Figure 12
Local Polynomial Regression: Alcohol Shares

also report MSE statistics for fully parametric models. Note how the "optimal" cross-validated bandwidths increase for both the food and the alcohol Engel curve examples as the degree of the local polynomial approximation increases, a result which one might expect if the local polynomial regression estimator reduces bias relative to the Nadaraya-Watson estimator [compare Equation (3.17)].

For the alcohol Engel curve, we see clearly the advantage of a local polynomial regression. Graphical evidence and specification test results from Table 3 suggest that the underlying Engel curve relationship for alcohol is quadratic in log expenditure, and Figure 12 supports this view. The fit of a local quadratic regression for alcohol (as measured by the MSE) dominates the local linear and Nadaraya-Watson curves at their respective cross-validated bandwidths.

VI. Concluding Comments

We have reviewed the techniques of density estimation and nonparametric estimation of the conditional regression function using kernel methods, and described recent developments in the fields of nonparametric specification testing, semiparametric estimation and local polynomial regression. An application to the Engel curve analysis of food expenditure and alcohol expenditure for a large expenditure survey of households in the United Kingdom was presented. These two commodities display very different nonparametric Engel curve shapes and therefore provide an excellent example of the benefits of implementing nonparametric methods in microeconomic data.

One aspect of our review has been on the implementation of a set of recently developed tests for the parametric null hypothesis against a nonparametric alterna-

Table 4
Mean Squared Errors: Food and Alcohol Shares

	h_{CV}	MSE (h_{CV})	MSE_{LS}
		Food Share	
$N - W$	0.1008	7.6367×10^{-3}	1.0472×10^{-2}
LLR	0.2801	7.6248×10^{-3}	7.6401×10^{-3}
LQR	0.5058	7.6288×10^{-3}	7.6425×10^{-3}
		Alcohol Share	
$N - W$	0.1759	1.8443×10^{-3}	1.8610×10^{-3}
LLR	0.2145	1.8440×10^{-3}	1.8636×10^{-3}
LQR	0.2932	1.8433×10^{-3}	1.8456×10^{-3}

Notes: Data are drawn from the 1980–82 Family Expenditure Surveys. Nonparametric estimates based on a Gaussian kernel. $N - W$, *LLR* and *LQR* relate, respectively, to the Nadaraya-Watson, local linear and local quadratic regressions for food and alcohol shares. MSE (h_{CV}) is the Mean Squared Error evaluated at the crossvalidation bandwidth. MSE_{LS} is the Mean Squared Error for a parametric (Least Squares) regression. For crossvalidation and MSE calculations, the trimming function $w(x)$ is set to exclude the smallest 2 percent of estimated densities. Calculated using NP-REG.

tive. These turned out to give similar and sensible results in our application to the analysis of alternative parametric forms of the Engel curve relationship. We considered the case of semiparametric estimation in partially linear models. This is important in cases where it is felt that certain regressors are likely to enter with a simple linear form. We used the Robinson (1988) approach to semiparametric regression which turned out to be a very effective tool for analyzing demographic variables in our Engel curve application. We also described a technique to control the nonparametric conditional mean for endogenous regressors. This was developed by adapting an augmented regression approach to the partially linear regression context. In the application using income to instrument total expenditure was found to have an important impact on the curvature of the Engel curve relationship. The tests for a parametric specification against a nonparametric alternative are also implemented in this partially linear semiparametric framework.

Finally, we have presented an implementation of a local polynomial estimator as an alternative to standard Nadaraya-Watson methods. For our Engel curve examples, the local polynomial estimator supported large cross-validated bandwidths than the Nadaraya-Watson alternative, and typically improved the measured fit (according to an MSE criterion) at these optimal bandwidths.

References

Ai, C., and D. McFadden. 1997. ''Estimation of Some Partially Specified Nonlinear Models.'' *Journal of Econometrics* 76:1–37.

Aït-Sahalia, Y., P. Bickel, and T. Stoker. 1994. ''Goodness-of-Fit Tests for Regression Using Kernel Methods.'' Mimeo. Cambridge: Massachusetts Institute of Technology.

Akaike, H. 1970. ''Statistical Predictor Information.'' *Annals of the Institute of Statistical Mathematics* 22:203–17.

Ansley, C., R. Kohn, and C. Wong. 1993. ''Nonparametric Spline Regression with Prior Information.'' *Biometrika* 80:75–80.

Atkinson, A., J. Gomulka, and N. Stern. 1990. ''Spending on Alcohol. Evidence from the Family Expenditure Survey 1970–1983.'' *Economic Journal* 100:808–27.

Banks, J., R. Blundell, and A. Lewbel. 1997. ''Quadratic Engel Curves and Consumer Demand.'' *Review of Economics and Statistics.*

Bierens, H. 1990. ''A Consistent Conditional Moment Test of Functional Form.'' *Econometrica* 58:1443–58.

Bierens, H., and H. Pott-Buter. 1990. ''Specification of Household Engel Curves by Nonparametric Regression.'' *Econometric Reviews* 9:123–84.

Blundell, R., P. Pashardes, and G. Weber. 1993. ''What Do We Learn about Consumer Demand Patterns from Micro Data?'' *American Economic Review* 83:570–97.

Craven, P., and G. Wahba. 1979. ''Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothness by the Method of Cross-Validation.'' *Numerische Mathematik* 31:377–403.

Deaton, A., and J. Muellbauer. 1980. ''An Almost Ideal Demand System.'' *American Economic Review* 70:312–26.

Duncan, A., and A. Jones. 1992. ''NP-REG: An Interactive Package for Kernel Density Estimation and Nonparametric Regression.'' Discussion Paper W92/07, Institute for Fiscal Studies.

Ellison, G., and S. F. Ellison. 1992. ''A Nonparametric Residual-Based Specification Test: Asymptotic, Finite-Sample and Computational Properties.'' Mimeo. Cambridge: Harvard University.

Fan, J. 1992. ''Design-Adaptive Nonparametric Regression.'' *Journal of the American Statistical Association* 87:998–1004.

Fan, J., and I. Gijbels. 1992. ''Variable Bandwidths and Local Linear Regression Smoothers.'' *Annals of Statistics* 20:2008–2036.

Härdle, W. 1990. *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Härdle, W., and M. Jerison. 1991. ''Cross-Sectional Engel Curves over Time.'' *Recherches Economiques de Louvain* 57:391–431.

Härdle, W., and O. Linton. 1994. ''Applied Nonparametric Methods.'' In *Handbook of Econometrics,* Volume IV, ed. R. Engle, and D. McFadden, 2295–2339. Elsevier Science.

Härdle, W., and T. Stoker. 1989. ''Investigating Smooth Multiple Regression by the Method of Average Derivatives.'' *Journal of the American Statistical Association* 84: 986–95.

Härdle, W., and E. Mammen. 1993. ''Comparing Nonparametric vs. Parametric Regression Fits.'' *Annals of Statistics.* 21:1926–47.

Hausman, J., W. Newey, and J. Powell. 1995. ''Nonlinear Errors in Variables: Estimation of Some Engel Curves.'' *Journal of Econometrics* 65:205–34.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1995. ''Nonparametric Characterization of Selection Bias Using Experimental Data: A Study of Adult Males in JTPA.'' Mimeo, University of Chicago.

Holly, A., and J. Sargan. 1982. ''Testing for Exogeneity in a Limited Information Framework.'' *Cahiers de Recherches Economiques*, No. 8204, Universite de Lausanne.

Jorgenson, D., L. Christensen, and L. Lau. 1975. ''Transcendental Logarithmic Utility Functions.'' *American Economic Review* 65:367–83.

Leser, C. 1963. ''Forms of Engel Functions,'' *Econometrica* 31:694–703.

Lewbel, A. 1991. ''The Rank of Demand Systems: Theory and Nonparametric Estimation.'' *Econometrica* 59:711–30.

Linton, O., and P. Gozalo. 1996. ''Using Parametric Information in Non-Parametric Regression.'' Mimeo. Brown University.

Nadaraya, E. 1964. ''On Estimating Regression.'' *Theory of Probability and its Applications* 9:141–42.

Newey, W. K., J. L. Powell, and F. Vella. 1995. ''Nonparametric Estimation of Triangular Simultaneous Equations Models.'' Mimeo. MIT Department of Economics.

Rice, J. 1984. ''Bandwidth Choice for Nonparametric Regression.'' *Annals of Statistics* 12: 1215–30.

Robinson, P. 1988. ''Root-N-Consistent Semiparametric Regression.'' *Econometrica* 56: 931–54.

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Stoker, T. 1986. ''Consistent Estimation of Scaled Coefficients.'' *Econometrica* 54:1461–81.

———. 1991. *Lectures in Semiparametric Econometrics.* CORE Lecture Series. CORE Foundation.

Watson, G. 1964.''Smooth Regression Analysis.'' *Sankhya* 26:359–72.

Wooldridge, J. 1992. ''A Test of Functional Form against Nonparametric Alternatives.'' *Econometric Theory* 8:452–75.

Working, H. 1943. ''Statistical Laws of Family Expenditure.'' *Journal of the American Statistical Association* 38:43–56.

Zheng, J. 1996. ''A Consistent Test of Functional Form via Nonparametric Estimation Techniques.'' *Journal of Econometrics* 75:263–89.