Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss

by

Jerry A. Hausman and Whitney K. Newey

MIT-CEEPR 93-014WP September 1993

NONPARAMETRIC ESTIMATION OF EXACT CONSUMERS SURPLUS AND DEADWEIGHT LOSS

by

Jerry A. Hausman and Whitney K. Newey

Department of Economics, MIT

October 1990

Revised, December 1992

This research was supported by the NSF. Helpful comments were provided by T. Gorman, I. Jewitt, A. Pakes and participants at several seminars. Sarah Fisher-Ellison and Christine Meyer provided able research assistance.

Abstract

We apply nonparametric regression models to estimation of demand curves of the type most often used in applied research. From the demand curve estimators we derive estimates of exact consumers surplus and deadweight loss, that are the most widely used welfare and economic efficiency measures in areas of economics such as public finance. We also develop tests of the symmetry and downward sloping properties of compensated demand. We work out asymptotic normal sampling theory for kernel and series nonparametric estimators, as well as for the parametric case.

The paper includes an application to gasoline demand. Empirical questions of interest here are the shape of the demand curve and the average magnitude of welfare loss from a tax on gasoline. In this application we compare parametric and nonparametric estimates of the demand curve, calculate exact and approximate measures of consumers surplus and deadweight loss, and give standard error estimates. We also analyze the sensitivity of the welfare measures to components of nonparametric regression estimators such as the number of terms in a series approximation.

Keywords: Consumer surplus, deadweight loss, consumer tests, nonparametric estimation.

Authors:

Professor Jerry A. Hausman
Department of Economics
MIT
Cambridge, MA 02139

Professor Whitney K. Newey
Department of Economics
MIT
Cambridge, MA 02139

1. Introduction

Nonparametric estimation of regression models has gained wide attention in the past few years in econometrics. Nonparametric models are characterized by a very large numbers of parameters. Often they may be difficult to interpret, and their usefulness in applied research has been demonstrated in a limited number of cases. We apply nonparametric regression models to estimation of demand curves of the type most often used in applied research. After estimation of the demand curves, we will then derive estimates of exact consumers surplus and deadweight loss which are the most widely used welfare and economic efficiency measures in areas of economics such as public finance. We also work out asymptotic normal sampling theory for the nonparametric case, as well as the parametric case (where except in certain analytic cases the results are not known).

The paper includes an application to gasoline demand. Empirical questions of interest here are the shape of the demand curve and the average magnitude of welfare loss from a tax on gasoline. In this application we compare parametric and nonparametric estimates of the demand curve, calculate exact and approximate measures of consumers surplus and deadweight loss, and given estimates of standard errors. We also analyze the sensitivity of the welfare measures to components of nonparametric regression estimators such as window width and number of terms in a series approximation.

The definition of exact consumers surplus is based on the expenditure function:

$$CS(p^1,p^0,u^r) = e(p^1,u^r) - e(p^0,u^r),$$

where p^0 are initial prices, p^1 are new prices, and u^r is the reference utility level. The case $r = 1$ corresponds to the case we focus on here, equivalent variation:

$$EV(p^1,p^0,y) = e(p^1,u^1) - e(p^0,u^1) = y - e(p^0,u^1),$$

where y denotes income, fixed over the price change. It is also easy to carry out a similar analysis for compensating variation.

The measure of exact deadweight loss (DWL) used corresponds to the idea of the loss in consumers surplus from imposition of a tax less the compensated tax revenues raised, under the implicit assumption that they are returned to the consumer in a lump sum manner. See Auerbach (1985) for a discussion of the various definitions of deadweight loss (also called the excess burden of taxation). Here we use the Diamond-McFadden (1974) definition of deadweight loss where $p^1 - p^0$ is the vector of taxes. Then, for equivalent variation the definition of exact DWL is:

$$DWL(p^1,p^0,y) = EV(p^1,p^0,y) - (p^1-p^0)'h(p^1,u^1) = y - e(p^0,u^1) - (p^1-p^0)'q(p^1,y),$$

where $h(p,u)$ is the compensated, or Hicksian, demand function and $q(p,y)$ is the Marshallian (market) demand function. Thus, to estimate exact consumers surplus or exact DWL it is necessary to estimate the expenditure function and the compensated function demand which are related by the equation,

$$\partial e(p,u)/\partial p_j = h_j(p,u). \quad \text{(Shephard's Lemma)}.$$

The expenditure function and compensated demand curve are estimated from observable data on the market, or Marshallian, demand curve, $q_j(p,y)$.

Various approximations have been proposed for estimation of the exact welfare measures. Willig (1976) demonstrated that the Marshallian measure of consumers surplus is often close to the exact measures, and he derived bounds as a function of the income share. Various authors have also recommended higher order Taylor-type approximations to the exact welfare measures. Limitations of the approximation approach are that they do not yield measures of precision given that they are based on estimated coefficients in most cases and the approximations may do poorly in some situations which are difficult to specify a priori. Deaton (1986) discusses these problems in more detail. Also, Hausman (1981) demonstrated that while the approximations may often do well for the consumers surplus measure, they often do very poorly in measurement of DWL; Small and Rosen (1981) also demonstrated a similar proposition in the discrete choice situation.

Two approaches have been proposed to estimation of the exact welfare measures from estimation of ordinary demand functions. Hausman (1981) demonstrated that for many widely used single equation demand specifications, the necessarily differential equation could be integrated to derive the expenditure function and also the compensated demand function. He was also able to derive the sampling distribution for the measures of consumers surplus and DWL, given the distribution of the estimated demand coefficients. Vartia (1983), instead of using an analytic solution to the differential equations, proposed a variety of numerical algorithms which estimate consumers surplus and DWL to any desired degree of accuracy. While the Vartia approach can be applied to a wider range of situations, no correct sampling distribution has been derived for the estimated welfare measures, although some approximate results are given in Hayes and Porter-Hudak (1986).

Both the Hausman approach and the Vartia approach can be applied to multiple price changes and are path independent for the true demand function. The various approximation approaches do not share the path independence property which can lead to perplexing computational results and has led to much theoretical misunderstanding in the appropriate literature. Here, we consider nonparametric estimation via an unrestricted estimator that uses some, but not all the implications of path independence. Our approach allows us to test path independence, i.e. symmetry of compensated demands, and impose further implications of path independence to construct more efficient estimators.

The Hausman and Vartia approaches begin with Roy's identity which links the ordinary demand function with the indirect utility function:

$$q_j(p,y) = -[\partial v(p,y)/\partial p_j]/[\partial v(p,y)/\partial y],$$

where v(p,y) is the indirect utility function. The partial differential equation from this equation can be solved along an indifference curve for a unique solution so long as the initial values are differentiable. Let $p(t)$ denote a price path with $p(0) = p^0$ and $p(1) = p^1$, and let $y(t)$ be compensated income, satisfying $V(p(t),y(t)) = V(p^1,y)$. Differentiating with respect to y gives

$$[\partial V(p(t),y(t))/\partial p]' \partial p(t)/\partial t + [\partial V(p(t),y(t))/\partial y]\partial y(t)/\partial t = 0.$$

Hausman notes that this equation can be converted into an ordinary differential equation by use of the implicit function theorem and Roy's identity. Let $S(t,y) = y - e(p(t),u^1)$ denote the equivalent variation for a price change from $p(t)$ to p^1. Then, since compensated income is $y - S(t,y)$,

$$\partial S(t,y)/\partial t = -q(p(t),\ y-S(t,y))' \partial p(t)/\partial t, \quad S(1,y) = 0. \tag{1.1}$$

Alternatively, this equation follows immediately from Shephard's Lemma and the definition of $S(t,y)$. Hausman solves this equation for some widely used demand curves. The solution gives both the expenditure function and indirect utility function, while the right-hand side of this equation gives the compensated demands.

Vartia's numerical solutions to the differential equation also arise from equation (1.1). To solve it Vartia uses a numerical method of Collatz. He orders $t_s = 1 - s/N$, $(s = 0,\dots, N)$, and defines S iteratively by

$$S(t_{s+1}) = S(t_s) - .5[q(p(t_{s+1}),y-S(t_{s+1})) + q(p(t_s),y-S(t_s))]'[p(t_{s+1})-p(t_s)]$$

This algorithm consists of averaging the demand at the last price and the current price and multiply this average by the change in price. By the envelope theorem, the product of the price change times the quantity equals the additional income required to remain on the same indifference curve. Intuitively, $dy = q \cdot dp$, where y is updated at each step of the process, rather than holding y constant which the Marshallian approximation to consumer surplus does. One can also use alternative numerical algorithms to solve equation (1.1), that may lead to faster methods. We use an Buerlisch-Stoer algorithm from Numerical Recipes that does not require solution to the implicit equation in Vartia's algorithm, has a faster (quintic) convergence rate that Vartia's (cubic), and in our empirical example leads to small estimated errors with few demand evaluations.

The possible shortcoming of all applications to date of both the Hausman approach and the Vartia approach is that the parametric form of the demand function is required. In most applied situations, the exact parametric specification of the demand curve (up to unknown parameters) will not be known. Thus, commonly used demand curve specifications may well lead to inconsistent estimates of the welfare measures if the demand curve is misspecified. This problem is potentially quite important, especially in the case of measuring deadweight loss which depends on "second order" properties of the demand curve. See Hausman (1981) for a discussion of the second order properties and their effect on estimates of the deadweight loss.

Varian (1982 a,b,c) has proposed an alternative approach based on the revealed preference ideas of Samuelson (1948) and Afriat (1967, 1973) which is nonparametric, but is only able to estimate upper and lower bounds on the welfare measures. Varian's nonparametric approximation approach is very interesting, but it often yields rather wide bounds, because many price observations per individual are required for tight bounds. Furthermore, use of sampling distributions to measure the precision of the estimates is problematical. Here, we use a nonparametric cross-section demand analysis analysis, imposing enough homogeneity across individuals and smoothness of the demand function that we can estimated it by nonparametric regression. We construct point estimates of exact consumers surplus and DWL as well as precision estimates of the welfare measures. The sensitivity of the welfare measure to the amount of smoothing used in the nonparametric regression can be analyzed straightforwardly.

2. Estimation

Our estimator of consumer surplus is obtained by solving equation (1.1) numerically with $q(p,y)$ replaced by an estimator obtained from nonparametric regression. We will also allow covariates w to enter the demand function $q(p,y,w)$. In our empirical work these covariates are region and time dummies. In other contexts they might be demographic variables. We will try to minimize the dimension of this nonparametric function by restricting w to enter in a parametric way. Let $T(g)$ denote a one-to-one function with range equal to the nonnegative orthant of $\mathbb{R}^k$, such as $T(g) = e^g$, x a one-to-one function of (p,y), and let $\tau(x)$ denote a "trimming" function, to be further discussed below. We will assume that the true value of the demand function is given by

$$(2.1) \qquad q_0(p,y,w) = T(g_0(x,w)), \quad g_0(x,w) = r_0(x) + w'\beta_0,$$

$$T^{-1}(q) = g_0(x,w) + \varepsilon, \quad E[\varepsilon|x] = 0, \quad E[\tau(x)w\varepsilon'] = 0.$$

where q denotes observed quantity and the expectations are taken over the distribution of a single observation from the data. We assume here that a numeraire good has been excluded from q, so that k is the number of goods, minus one.

Thus, the true demand function is assumed to be a function of a partially linear specification for the regression of $T^{-1}(q)$ on p, y, and w.[1] A corresponding estimator of the demand function will be $\hat{q}(p,y,w) = T(\hat{g}(x,w)) = T(\hat{r}(x)+w'\hat{\beta})$, where $\hat{g}$ is an estimator of g_0, such as the kernel or series estimator discussed below. We estimate exact consumer surplus nonparametrically by substituting $\hat{q}(p,y,w)$ for $q(p,y)$

[1]This specification does not restrict the joint distribution of the data, unlike previous work on partially linear models (e.g. Robinson, 1988), where $E[\varepsilon|p,y,w] = 0$ is imposed. A precise description of $g_0(x,w)$, for nonnegative τ, is as the mean-square projection of $E[q|x,w]$ on functions of the form $r(x) + w'\beta$ for the probability distribution $Pr_\tau(\mathcal{A}) = E[\tau(x)1(\mathcal{A})]/Pr(\mathcal{A})$, where $1(\cdot)$ denotes the indicator function.

in equation (1.1) and solving numerically. The empirical results reported in this version of the paper set w equal to its sample mean and $\tau(x) = 1$.

A kernel estimator is a locally weighted average that can be described as follows. Let $\mathcal{K}(v)$ be a kernel function, where v has $k+1$ elements, satisfying $\int\mathcal{K}(v)dv = 1$ and other regularity conditions discussed below, $h > 0$ a bandwidth parameter, and $K_h(v) = h^{-k-1}\mathcal{K}(v/h)$. Also, let the data be denoted by $z_1, \ldots, z_n$, where z includes q, y, w, and possibly other variables. For a matrix function $B(z)$, a kernel estimator of $E[B(z)|x]$ is

$$\hat{E}[B(z)|x] = \sum_{j=1}^{n} B(z_j)K_h(x-x_j)/\sum_{j=1}^{n} K_h(x-x_j).$$

To estimate the partially linear specification in equation (2.1), we "partial out" the coefficients of w in a way analogous to that in Robinson (1988). The estimator of g is

$$\hat{g}(x,w) = \hat{r}(x) + w'\hat{\beta}, \quad \hat{r}(x) = \hat{E}[T^{-1}(q)|x] - \hat{E}[w|x]'\hat{\beta}, \tag{2.2}$$

$$\hat{\beta} = [\sum_{i=1}^{n}\tau_i(w_i-\hat{E}[w|x_i])(w_i-\hat{E}[w|x_i])']^{-1}\sum_{i=1}^{n}\tau_i(w_i-\hat{E}[w|x_i])(T^{-1}(q_i)-\hat{E}[T^{-1}(q)|x_i]),$$

where $\tau_i = \tau(x_i)$. A convenient kernel that we consider in the empirical work is

$$\mathcal{K}(v) = \begin{cases} C(1 - v'v), & v'v \le 1 \\ 0, & \text{otherwise} \end{cases}, \tag{2.3}$$

where the constant C is chosen so $\int\mathcal{K}(v)dv = 1$. This is a multivariate Epanechnikov kernel. The asymptotic theory requires kernels with integrals over $\mathcal{K}(v)$ of certain even powers of v equal to zero, that are often called "higher order," and are useful for reducing asymptotic bias in kernel estimation. Therefore, we also consider in the empirical work a kernel of the form

$$\mathcal{K}(v) = \phi(v_1)\phi(v_2)(12 - 6v_1^2 - 6v_2^2 + 5v_1^4 + 10v_1^2v_2^2 + 5v_2^4), \tag{2.3a}$$

for our one dimensional price ($k = 1$) application.

It is well known that the choice of bandwidth h can have important effects on nonparametric regression. In the empirical work we consider a data based choice of h, equal to a "plug-in" value that minimizes estimated asymptotic mean square error, and also consider the sensitivity of the results to the choice of bandwidth. In the empirical work we also allow the kernel to be data based in that we normalize by the estimated variance of price and income, although the theory does not allow for data-based bandwidth or kernel.

A series estimator is the predicted value from a regression of the log of gasoline consumption on some approximating functions for p and y and on w. For x a one-to-one, smooth transformation as above, let $\phi^K(x) = I_k \otimes (\phi_{1K}(x),\ldots,\phi_{KK}(x))$ denote a matrix of functions of x, the idea being that $r(x)$ is to be approximated by linear combinations of $\phi^K(x)$. Let $\phi^K(x,w) = (\phi^K(x)',w')'$ and $\hat{\gamma} = [\sum_{i=1}^n \phi^K(x_i,w_i)\phi^K(x_i,w_i)']^{-1} \cdot \sum_{i=1}^n \phi^K(x_i,w_i)T^{-1}(q_i)$ be the the coefficients from a regression of $T^{-1}(q_i)$ on $\phi^K(x_i,w_i)$. A series estimator of g with $\tau(x) = 1$ is

$$(2.4) \qquad \hat{g}(x,w) = \phi^K(x,w)'\hat{\gamma} = \phi^K(x)'\hat{\eta} + w'\hat{\beta},$$

where $\gamma = (\eta',\beta')'$ is partitioned conformably with $\phi^K(x)$. This estimator can also be interpreted as "partialling out" w, satisfying equation (2.2) with the kernel conditional expectation estimator replaced by $\hat{E}[B(z)|x] = \phi^K(x)'\hat{\eta}_B$ where $\hat{\eta}_B$ are the coefficients of the least squares regression of $B(z_i)$ on $\phi^K(x_i)$.

Two important types of approximating functions are power series and regression splines. Power series are formed by choosing the elements of $\phi^K(x)$ to be products of powers of the individual components of x. Power series are easy to compute and have good approximation rates for smooth functions, although they are sensitive to outliers and local behavior of the approximation, and can be highly collinear. The collinearity problem can be overcome to some extent by replacing the powers of individual components with orthogonal polynomials, which does not effect the estimator but may lead to easier computation. Regression splines are piecewise polynomials of order Δ with fixed join points. For univariate x in $[0,1]$ with evenly spaced knots (i.e. join points), regression spline approximating functions are $a_{jK}(x) = x^{j-1}$, $(j = 1, \ldots, \Delta+1)$, $(x-(j-\Delta-1)/[K-\Delta])_+^{\Delta}$, $j \geq \Delta + 2$, where $(v)_+^{\Delta} = 1(v\geq 0)v^{\Delta}$. For multivariate x, a regression spline can be formed from all cross-products of univariate splines in the components of x. The spline approximation rate for very smooth functions is not as fast as power series, but they are less sensitive to outliers. Unlike power series, the range of x must be known in order to place the knots. In practice, such a known range could be constructed by dropping from the data any observation where x is not in some known range. Also, the power spline sequence above can be highly collinear, but this problem can be alleviated by replacing them with their corresponding B-splines: e.g. see Powell (1981).

Series estimators are sensitive to the numbers of terms in the approximation. In the empirical work we choose the number of terms by cross-validation, and also try different numbers of terms to see how the results are affected.

Returning now to estimation of consumer surplus, our estimator is constructed by substituting the estimated demand function in equation (1.1) and integrating numerically. As in Section 1, let $p(t)$ be a price path with $p(0) = p^0$ and $p(1) = p^1$. Also, let $S(y,w,g)$ denote the solution to equation (1.1) with demand function $q(p,y) = T(g(x,w))$. Then consumer surplus at particular income and covariate values y_0 and w_0 respectively, with a corresponding estimator, is

$$(2.5) \qquad S_0 = S(y_0,w_0,g_0), \quad \hat{S} = S(y_0,w_0,\hat{g}),$$

where $\hat{g}$ is a kernel, series, or other nonparametric estimator. A corresponding deadweight loss value and associated estimator can be formed by subtracting the "tax receipts," as in

$$(2.6)\qquad L_0 = S_0 - (p^1-p^0)'T(g(p^1,y_0,w_0)),\quad \hat{L} = \hat{S} - (p^1-p^0)'T(\hat{g}(p^1,y_0,w_0)).$$

A summary measure for consumer surplus can be obtained by averaging over income values. In addition, it may sometimes be of interest to average over different prices, to reflect the fact that individuals face different prices. To set up such an average let u index price paths, so that $p(t,u)$ is the price at $t \in [0,1]$ for price path u, with initial and final prices $p(0,u)$ and $p(1,u)$ respectively. Also, let z denote a single data observation that includes u and values y_0 for income and w_0 for covariates. For example, u, y_0, and w_0 might be drawn (simulated) from some distribution, or y_0 and w_0 might be the actual observations. Let $S(z,g)$ be the solution to equation (1.1) for the price path $p(u,t)$ at y_0 and w_0, with demand $q(p,y) = T(g(x,w))$. The average surplus and deadweight loss we consider are weighted means across z, of the form

$$(2.7)\qquad \mu_0 = E[\omega(z)S(z,g_0)]/E[\omega(z)],\quad \hat{\mu} = \bar{\omega}^{-1}\textstyle\sum_{i=1}^{n}S(z_i,\hat{g})/n,\quad \bar{\omega} = \textstyle\sum_{i=1}^{n}\omega(z_i)/n.$$

$$\lambda_0 = \mu_0 - E[\omega(z)\{p(1,u)-p(0,u)\}'T(g_0(p(1,u),y_0,w_0))]/E[\omega(z)],$$

$$\hat{\lambda} = \hat{\mu} - \bar{\omega}^{-1}\textstyle\sum_{i=1}^{n}\omega(z_i)\{p(1,u_i)-p(0,u_i)\}'T(\hat{g}(p(1,u_i),y_{0i},w_{0i}))/n.$$

It is interesting to note that the consumer surplus estimator may converge faster than the deadweight loss estimator. In particular, $\hat{S}$ is like an integral over one dimension (the variable t in $p(t)$), and so has a faster convergence rate than $\hat{L}$, which depends on the value of $\hat{g}$ at a particular point.[2] Similarly, average consumer surplus and deadweight loss may have different convergence rates. One important case where their convergence rates will be the same, both being the parametric $1/\sqrt{n}$ rate, is

[2]It is known from the work on semiparametric estimation that integrals or averages converge faster than pointwise values, e.g. Powell, Stock, and Stoker (1989). The fact that one-dimensional integrals converge faster than pointwise values has recently been shown in Newey (1992a) for kernel estimators.

when the initial and final prices and income have sufficient variation. An example would be that where the initial price for each individual is the price they actually faced, and the tax rate is the same across individuals. In this case averaging will take place over all the arguments of the nonparametric estimates, which is known from the semiparametric estimation literature to result in $\sqrt{n}$-consistency (under appropriate regularity conditions).

So far, our nonparametric consumer surplus estimators have ignored the residual $\varepsilon = T^{-1}(q) - g_0(p,y,w)$. This approach is consistent with current practice in applied econometrics, and is difficult to improve on without more information about the residual. One can ignore the residual if it is all measurement error and not if it contains individual heterogeneity. Hausman (1985) shows that it is possible to separate out measurement error and heterogeneity in parametric, nonlinear models, but the amount of heterogeneity is not identified in our nonparametric, linear in residual specification. Even when ε is all heterogeneity, it may be possible to interpret the demand function as corresponding to a particular consumer type. Suppose that $\varepsilon = \varepsilon(p,y,w,v)$ for some function $\varepsilon(p,y,w,v)$ of prices, income, covariates and a taste variable v, where v is independent of price and income. In general $\varepsilon(p,y,w,v)$ will depend on p and y, as shown by Brown and Walker (1990).[3] Nevertheless, if $\varepsilon(p,y,w,v)$ is identically zero for some value of v then $g(p,y,w)$ can be interpreted as the demand function for that value of v (e.g. for $\varepsilon = \sigma(p,y,w)v$). In the rest of the paper we stay with the specification of demand as $T(g_0(p,y,w))$, corresponding to an interpretation of ε as measurement error or to evaluation at a particular consumer type.[4]

[3]They showed that for $T(q) = q$ the residual ε must be functionally dependent on p and y. Also, Brown has indicated to us in private communication that the same result is true for $T(q) = e^q$.

[4]An alternative approach recently suggested by Brown and Newey (1992) is to average consumer surplus over different consumer types, when $\varepsilon(p,y,w,v)$ is an estimable, one-to-one function of v.

3. Asymptotic Variance Estimation

Under regularity conditions given in Section 6, all of the estimators will be asymptotically normal. To be specific, let $\hat{\theta}$ denote any one of the estimators previously presented. For a kernel estimator there will be V_0 and $\alpha \geq 0$ such that

$$(3.1) \qquad \sqrt{n}\sigma^{\alpha}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_0).$$

The "full-average," $\sqrt{n}$-consistent case corresponds to $\alpha = 0$, while in other cases the convergence rate for $\hat{\theta}$ will be $1/(\sqrt{n}\sigma^{\alpha})$, which is slower than $1/\sqrt{n}$ by $\sigma \longrightarrow 0$. For a series estimator there will be V_n such that

$$(3.2) \qquad \sqrt{n}V_n^{-1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1).$$

In the $\sqrt{n}$-consistent case the series and kernel estimators will have the same asymptotic variance, with V_n converging to V_0 from equation (3.1). In other cases the two estimators generally will not have the same asymptotic variance, and the series estimator will satisfy the weaker property of equation (3.2), that does not specify a rate of convergence. Exact convergence rates for series estimators are not yet known, except for $\sqrt{n}$-consistency, although it is possible to bound the convergence rate. Despite this lack of a convergence rate, equation (3.2) can still be used for asymptotic inference.

For large sample inference, suppose that there is an estimator $\hat{V}_n$ of V_n in equation (3.2). If $(V_n/\hat{V}_n)^{1/2} \xrightarrow{p} 1$ then it follows by equation (3.2) (and the Slutzky theorem) that

$$(3.3) \qquad \sqrt{n}\hat{V}_n^{-1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1).$$

Consequently, a $1-\alpha$ large sample confidence interval will be

$$(3.4) \qquad \hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{V}_n/n}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. One could also use equation (3.4) to form large sample hypothesis tests.

To form large sample confidence intervals, as in equation (3.4), an estimator $\hat{V}_n$ of V_n is needed. For kernel estimators, one method of forming $\hat{V}_n$ would be to derive a formula for V_0 and then substitute estimators for unknown quantities to form $\sigma^{-2\alpha}\hat{V}_0$. This procedure is not very feasible, because the asymptotic variances are quite complicated, as described in Section 6. Instead we use an alternative method, from Newey (1992a), that only requires knowing the form of $\hat{\theta}$. For series estimators we also use a method that just uses the form of $\hat{\theta}$.

The asymptotic variance estimators for kernel and series estimators have some common features. In each case,

$$(3.5) \qquad \hat{V}_n = \textstyle\sum_{i=1}^n \hat{\psi}_{ni}^2/n,$$

where the estimates $\hat{\psi}_{ni}$ are constructed in the following way. Also, in each case, the variance will be based on the form of the surplus estimator, as in

$$(3.6) \qquad \hat{\theta} = \bar{\omega}^{-1}\textstyle\sum_{i=1}^n a(z_i,\hat{g})/n, \quad \bar{\omega} = \textstyle\sum_{i=1}^n \omega(y_i)/n,$$

where $a(z,g)$ is a function of a single observation z and the partially linear specification $g = g(x,w) = r(x) + w'\beta$, $\hat{g}$ is the kernel or series estimator described earlier, and $\omega(y)$ is a weight function. The specification of $a(z,g)$ and $\omega(y)$ corresponding to each case is

$$(3.7) \qquad \hat{\theta} = \hat{S}: \quad a(z_i,g) = S(y_0,w_0,g); \quad \omega(y) = 1;$$

$$\hat{\theta} = \hat{L}: \quad a(z_i,g) = S(y_0,w_0,g) - (p^1-p^0)'T(g(p^1,y_0,w_0));$$

$$\hat{\theta} = \hat{\mu}: \quad a(z_i,g) = \omega(y_i)S(z_i,g);$$

$$\hat{\theta} = \hat{\lambda}: \quad a(z_i,g) = \omega(y_i)\{S(z_i,g) - [p(1,u_i)-p(0,u_i)]'T(g(p(1,u_i),y_i,w_i))\},$$

where $g(p,y,w)$ denotes $g(x,w) = r(x) + w'\beta$ evaluated at the x that is the image of (p,y). For both kernel and series estimation, $\hat{\psi}_{ni}$ will have two components, one of which accounts for the variability of z_i in $a(z_i,g)$ and $\omega(y_i)$, and the other for the variability of $\hat{g}$. The first component is the same for both kernel and series, so that we can specify

$$(3.8) \qquad \hat{\psi}_{ni} = \hat{\psi}_i^z + \hat{\psi}_i^g, \quad \hat{\psi}_i^z = \bar{\omega}^{-1}a(z_i,\hat{g}) - \hat{\theta} - \bar{\omega}^{-1}\hat{\theta}[\omega(y_i) - \bar{\omega}],$$

where an n subscript on $\hat{\psi}_i^z$ and $\hat{\psi}_i^g$ is suppressed for notational convenience. The $\hat{\psi}_i^z$ term is an asymptotic approximation to the influence on $\hat{\theta}$ of the i^{th} observation in $\bar{\omega}^{-1}\sum_{i=1}^n a(z_i,g)$. The term $\hat{\psi}_i^g$ will account for the estimation of g. It can be constructed from an asymptotic approximation to the influence on $\hat{\theta}$ of the i^{th} observation in $\hat{g}$, taking a different form for kernel and series estimators.

For kernel estimators the idea for forming $\hat{\psi}_{ni}^g$, developed in Newey (1992a), is to differentiate with respect to the i^{th} observation in the kernel estimator. This calculation amounts to a "delta-method" for kernels, and leads to an estimator that is robust to heteroskedasticity and has the same form, no matter what the convergence rate of $\hat{\theta}$ is. To describe the estimator, let δ denote a scalar and

$$(3.9) \qquad \hat{h}^r(x) = \textstyle\sum_{j=1}^n (T^{-1}(q_j) - w_j'\hat{\beta})K_h(x-x_j)/n, \quad \hat{f}(x) = \sum_{j=1}^n K_h(x-x_j)/n,$$

$$\hat{A}_i = \partial[n^{-1}\textstyle\sum_{j=1}^n a(z_j, \{\hat{f} + \delta K(\cdot - x_i)\}^{-1}\{\hat{h}^r + \delta(T^{-1}(q_i) - w_i'\hat{\beta})K_h(\cdot - x_i) + w_i'\hat{\beta})/n]/\partial\delta|_{\delta=0},$$

$$\hat{G}^\beta = \partial[\textstyle\sum_{j=1}^n a(z_j, \hat{g}_\beta)/n]/\partial\beta|_{\beta=\hat{\beta}}, \quad \hat{g}_\beta(x,w) = \hat{E}[T^{-1}(q) - w'\beta|x] + w'\beta$$

$$\hat{M} = \textstyle\sum_{i=1}^n \tau_i(w_i - \hat{E}[w|x_i])(w_i - \hat{E}[w|x_i])/n,$$

$$\hat{\psi}_i^g = \bar{\omega}^{-1}\{\hat{A}_i - \textstyle\sum_{j=1}^n \hat{A}_j/n + \hat{G}^\beta\hat{M}^{-1}\tau_i(w_i - \hat{E}[w|x_i])(T^{-1}(q_i) - \hat{g}(x_i,w_i))\}.$$

The term $\hat{A}_i$ in an asymptotic approximation to the influence of the i^{th} observation in $\hat{r}(x)$ on the average of $a(z_j,\hat{g})$, while the second term in $\hat{\psi}_i^g$ is a fairly standard delta-method term for estimation of β.

For series estimators the idea is to apply the "delta-method" as if the series approximation were exact. This results in correct asymptotic inference because it accounts properly for the variance, while the bias is small under appropriate regularity conditions. To describe the estimator, let

$$(3.10)\qquad \hat{G}^{\gamma} = \partial[\textstyle\sum_{j=1}^{n}a(z_j,g_\gamma)/n]/\partial\gamma|_{\gamma=\hat{\gamma}}, \quad g_\gamma(x,w) = \phi^K(x,w)'\gamma,$$

$$\hat{\psi}_i^g = \bar{\omega}^{-1}\hat{G}^{\gamma}\hat{\Sigma}^{-1}\phi^K(x_i,w_i)[T^{-1}(q_i) - \hat{g}(x_i,w_i)], \quad \hat{\Sigma} = \textstyle\sum_{i=1}^{n}\phi^K(x_i,w_i)\phi^K(x_i,w_i)'/n.$$

Here $\hat{\psi}_i^g$ is a standard "delta-method" term for ordinary least squares estimation of γ.

For either kernels or series, the main difficulty in computing $\hat{\psi}_i^g$ is calculating the derivatives $\hat{A}_i$ and $\hat{G}^{\beta}$ or $\hat{G}^{\gamma}$. For each of the estimators described in Section 2, it is possible to derive analytical expressions for these derivatives, but the expressions are so complicated as to make them almost useless for calculation. Instead, these derivatives can be calculated by numerical differentiation. This calculation only requires evaluation of $\sum_{i=1}^{n}a(z_i,g)/n$ for many different values of g, which is quite feasible, particularly for series estimators.

A procedure analogous to that for the series estimator can be used to construct a consistent asymptotic variance estimator for exact consumer surplus for any parametric specification of the demand function. For a parametric specification, $a(z,\gamma)$ will depend on the parameters γ of the demand function. In this case $\hat{G}^{\gamma} = \partial[\sum_{i=1}^{n}a(z_i,\gamma)/n]/\partial\gamma|_{\gamma=\hat{\gamma}}$ can be calculated by numerical differentiation. Then, supposing that $\sqrt{n}(\hat{\gamma} - \gamma_0) = \sum_{i=1}^{n}\Psi^{\gamma}(z_i)/\sqrt{n} + o_p(1)$ and that $\hat{\Psi}_i^{\gamma}$ is an estimator of $\Psi^{\gamma}(z_i)$, we can form

$$(3.11)\qquad \hat{\psi}_i = \hat{\psi}_i^z + \bar{\omega}^{-1}\hat{G}^{\gamma}\hat{\Psi}_i^{\gamma}.$$

Asymptotic inference for parametrically estimated exact consumer surplus could then be carried out as described above for $\hat{V}_n$ as in equation (3.4).

4. Testing Consumer Demand Conditions

Tests of the downward slope and symmetry of Hicksian (compensated) demands provide useful specification checks for consumer surplus estimates, and are of interest in their own right as tests of consumer theory. Here we consider tests that are natural by-products of consumer surplus estimation. An implication of symmetry is that consumer surplus is independent of of the price path, which can be tested by comparing estimates based on different price paths. The downward sloping property can be tested by comparing the demand at the new price with compensated demand at the initial price, which is easily computed from the consumer surplus estimate.

Path independence can be tested by comparing consumer surplus for the same income and covariates but different price paths. To describe this test, let j index a price path $p^j(t)$, with $p^j(0) = p^0$ and $p^j(1) = p^1$. Let $\hat{S}_j$ denote the equivalent variation estimator described above for the price path $p^j(t)$, income y_0, and covariates w_0. An implication of symmetry of compensated demand is that all $\hat{S}_j$, $(j = 1, \ldots, J)$ should converge to the same limit. This implication can be tested by comparing the different estimators. A simple way to construct this test is by minimum chi-square. Let $\hat{\psi}_i^j$ denote the corresponding estimator from equation (3.7), and

$$(4.1) \qquad \hat{\Pi} = (\hat{S}_1,\ldots,\hat{S}_J)', \quad \hat{\Psi}_i = (\hat{\psi}_i^1,\ldots,\hat{\psi}_i^J)', \quad \hat{\Omega} = \textstyle\sum_{i=1}^n \hat{\Psi}_i\hat{\Psi}_i'/n.$$

Here $\hat{\Omega}$ is an estimator of the joint asymptotic variance of $\hat{\Pi}$. Let e denote a $J \times 1$ vector of 1's. Then the test statistic is given by

$$(4.2) \qquad T = n(\hat{\Pi} - \tilde{S}\cdot e)'\hat{\Omega}^{-1}(\hat{S} - \tilde{\mu}\cdot e), \quad \tilde{\mu} = (e'\hat{\Omega}^{-1}e)^{-1}e'\hat{\Omega}^{-1}\hat{\Pi}.$$

Under the symmetry hypothesis the asymptotic distribution of this test statistic will be $\chi^2(J-1)$.

The estimator $\tilde{S}$ may be of interest in its own right. By the usual minimum chi-square estimation theory, under symmetry of compensated demands it will be at least as asymptotically efficient as any of the estimators $\hat{S}_j$, with an estimated asymptotic variance matrix $(e'\hat{\Omega}^{-1}e)^{-1}$. This efficiency improvement leads to the question of an efficiency bound. This question could be answered, in part, by deriving the optimal way to average over all different price paths, a question outside the scope of this paper.

The downward slope of compensated demands can be tested by testing for nonnegativity of $(p^1-p^0)'[T(g_0(p^1,y_0,w_0)) - T(g_0(p^0,y_0-S(y_0,w_0,g_0),w_0)]$ over several different prices, incomes, and covariates. To describe this test, let p_j^0, p_j^1, y_{0j}, w_{0j}, $(j = 1, \ldots, J)$ denote different values, let $\hat{S}_j$ denote the corresponding equivalent variation estimates, and

$$(4.3) \qquad \hat{\theta}_j = (p_j^1 - p_j^0)'[T(\hat{g}(p_j^1,y_{j0},w_{j0})) - T(\hat{g}(p_j^0,y_{j0}-\hat{S}_j,w_{j0}))], \quad (j = 1, \ldots, J).$$

An implication of convexity of the expenditure function is that each of these estimators should have a nonnegative limit. This hypothesis can be tested using an estimator of the asymptotic variance matrix of $\tilde{\theta} = (\hat{\theta}_1,\ldots,\hat{\theta}_J)'$, that can be constructed via the approach of Section 4. Let $\hat{\psi}_i^j$ be as described in equation (3.7), for $a(z,g) = (p_j^1-p_j^0)'[T(g(p_j^1,y_{j0},w_{j0}))-T(g(p_j^0,y_{j0}-S_j(y_{j0},w_{j0},g),w_{j0})]$, $w(z) = 1$ and let $\hat{\Omega} = \sum_{i=1}^n (\hat{\psi}_i^1,\ldots,\hat{\psi}_i^J)'(\hat{\psi}_i^1,\ldots,\hat{\psi}_i^J)/n$. Alternatively, if the income values y_{j0} are mutually distinct then the asymptotic covariances between the $\hat{\theta}_j$ will be zero, so that $\hat{\Omega} = \sum_{i=1}^n \mathrm{diag}((\hat{\psi}_i^1)^2,\ldots,(\hat{\psi}_i^J)^2)/n$ will suffice. The asymptotic approximation to the distribution of $\tilde{\theta}$ is then that $\tilde{\theta}$ is normal with variance $\hat{\Omega}/n$. Thus, the hypothesis that the limit of $\tilde{\theta}$ is a nonnegative vector can be tested by applying multivariate tests of inequality restrictions developed in the statistics literature. A particularly simple test would be to reject if $\min_j\{\sqrt{n}\hat{\theta}_j/\hat{\Omega}_{jj}\} \le k$ for some k. The size of this test could be calculated by simulating the distribution of the minimum of vector of mean zero normals with variance matrix equal to the correlation matrix implied by $\hat{\Omega}$.

It is possible to combine these two types of tests, of symmetry and of downward sloping compensated demand, into a single test of consumer demand theory, by "stacking" $\tilde{S}$ and $\tilde{\theta}$ into a single vector $(\tilde{S}',\tilde{\theta}')$. The $\hat{\psi}$ values could be stacked in the corresponding way, and an estimator of the variance of $(\tilde{S}',\tilde{\theta}')'$ formed as the average outer product of the stacked vector of $\hat{\psi}$ values. Also, it would be possible to consider versions of these tests based on the consumer surplus averages of equation (2.7). Furthermore, it should be possible to give these tests some asymptotic power against any alternative to demand theory by letting J grow with the sample size in such a way that different price paths and income and covariate "cover" all values in their support. These extensions are beyond the scope of this paper.

5. An Application to Gasoline Demand

To estimate the nonparametric and parametric demand functions for gasoline, we use data from the U.S. Department of Energy. The first three waves of the data were collected in the Residential Energy Consumption Survey conducted for the Energy Information Agency of the Department of Energy. Surveys were conducted in 1979, 1980, and 1981 at the household level. Gasoline consumption is kept by diary for each month; in our analysis we use average household gallons consumed per month. The gasoline price is the weighted average of purchase price over a month. Note that gasoline prices were quite high during most of this period in the U.S. because of the second (Iranian) oil shock. Gasoline prices averaged between \$1.34-1.46 for these 3 years where we use 1983 \$. Income is divided into 12 categories with the highest category being over \$50,000 (in 1983 \$). Here we used the conditional median for national household income above \$50,000. Lastly geographical information is given by 8 census regions. Average driving patterns differ significantly across regions of the U.S.. The last three waves of data were collected by the Energy Information Agency in the Residential Transportation Energy

Consumption Survey for the years 1983, 1985, and 1988. Price and income were collected in the same manner as the earlier surveys. (The upper limit on income changed in the surveys; however the technique used to estimate income in the top category remained the same) Note that the (real) gasoline price in the U.S. fell throughout this period so that by 1988 it had decreased to levels, about \$0.83, approximately equal to prices before the first oil shock in 1974. In the latter surveys, 9, rather than 8, census divisions were used. Since we are unable to map the earlier 8 region breakdown into 9 regions, or vice-versa, in the empirical specifications we use different sets of indicator variables depending on the survey year.

Overall, we have 18,109 observations which should provide sufficient observations to do nonparametric estimation and achieve fairly precise results. Our empirical approach is to do both the nonparametric and parametric estimation with indicator variables both for survey year and for regions. Thus, we have 20 indicator variables in our specifications. In the parametric specifications, we allow for interactions, most of which are found to be statistically significant which is to be expected given our very large sample. However, we decided to use the same set of indicator variables in both the nonparametric and parametric specifications to make for easier comparisons.

We give four types of nonparametric estimates, using the Epanechnikov kernel from equation (2.3), the normal higher order kernel from equation (2.3a), a cubic regression spline with evenly spaced knots, and power series. We use a log-linear demand specification, where $T(g) = e^g$. Also, the covariates w are 20 indicator variables that allow for different region and survey year effects. In the results we present we evaluate demand and consumer surplus at a fixed value for w equal to its sample mean.

We tried several different bandwidths for the kernel estimators, based on the smoothness of the graphs discussed below. The Epanechnikov kernel estimates used a bandwidth of .82 for the coefficients $\hat{\beta}$ of w, and $\tau(x) = 1$. The bandwidths were used to form $\hat{r}(x)$, equal to $h = .82$, $h = .55$, and $h = .45$. For the normal, higher order kernel in equation (2.3a), we used bandwidths of $h = .55$ and $h = .45$ for both the nonparametric part and the coefficients of w.

For the series estimates, we used cross-validation to help choose the number of terms. Cross-validation is the sum of squares of prediction errors from predicting one observation using coefficients estimated from all other observations. Minimizing a cross-validation criteria is known to lead to minimum asymptotic mean-square error estimates. Table 5.1 reports the criteria for splines and power series that are additive in (log) price and income. No interactions were included because they were never found to be significant, either in a statistical sense or in terms of lowering the cross-validation criteria.

Table 5.1: Cross-validation for Series Estimators

Cubic spline, evenly spaced knots		Power series	
Knots	CV	Order	CV
1	832	1	922
2	826	2	903
3	842	3	834
4	857	4	824
5	801	5	825
6	797	6	900
7	784	7	823
8	801	8	791
9	782	9	804
10	801		

The criteria are minimized at 7 or 9 knots and 8^{th} order power series. The theory suggests that one should choose a number of knots that is greater than the mean-square error minimizing one, so the bias goes to zero faster than the variance. For this reason we prefer 8 or 9 knots and an 8^{th} or 9^{th} order polynomial. The results for 9 knots were similar to those for 8, except that the estimated standard errors were very large, so we do not report them. Instead we report results for 6, 7, and 8 knots.

For purposes of comparison we also report results for standard parametric forms for the demand function. The specification of the demand function is:

$$(5.1) \qquad \ln q_i = \eta_0 + \eta_1 \ln p + \eta_2 \ln y + w'\beta + \varepsilon,$$

where y is household income, p is gasoline price, and w are the 20 region and time dummies discussed above. The estimated income elasticity is .37 with standard error .01 while the estimated price elasticity is −0.80 with standard error .09. To check on the sensitivity of our estimates, we also estimated a "translog" type of parametric specification where we allow for quadratic terms in log income and log price as well as an interaction term between log income and log price. The income-price interaction term has no estimated effect, but the quadratic terms do have an effect on the estimates. The sum of squared errors decreases from 8900.1 to 8877.0 which is a decrease of 0.26%, but a traditional F statistic is calculated to be 16.1 with three degrees of freedom due to the large sample size. The estimated elasticities for the log-linear and log-quadratic model are quite similar at the median gasoline prices, \$1.23: the log-linear price elasticity is −0.81 at all income levels while the log-quadratic model price elasticity is approximately −0.87 with very little variation across income levels. The two specifications do have different elasticities at lower and higher gasoline prices. The log linear model price elasticity, since it is estimated as a single parameter, remains at −0.81 across all gasoline prices while the log-quadratic model, which has a variable price elasticity, has an estimated elasticity of approximately −0.64 for gasoline price of \$1.08 (the first quartile price) while the log-quadratic model has an estimated elasticity of −1.14 at a gasoline price of \$1.43 (the third quartile). Since the results for log quadratic translog specification are only approximately similar to the simpler liner specification, we present results for both of the parametric specifications in what follows.

Figures 1 - 3 show the estimated nonparametric log demand with respect to log price, evaluated at mean income, for the parametric, Epanechnikov kernel, and spline specifications. We do not give graphs for other income values because the graphs have similar shape and the spline results discussed below strongly suggest that the log demand function is additive in log price and log income, in which case the shape of log demand will not depend on income. There are interesting differences between the parametric and nonparametric estimates, with the nonparametric estimates having a much more complicated shape than the parametric ones. In Figure 2 we find kernel demand curves that are generally downward sloping over the range of the data. In our opinion the demand curve for $h = .82$ looks "too smooth," and the one for $h = .45$ looks "too rough" so in the subsequent analysis we will use $h = .55$ as our preferred bandwidth. The shape of the spline demand function is qualitatively similar to that the kernel estimate, although the demand function does slope up very slightly at some points on the graph. We tested whether this upward slope indicates a failure of demand theory using the test for of downward sloping compensated demand described above. For 8 knots (our preferred number) a price change from $1.39 - 1.46, which is the range over which the demand curve slopes up, our N(0,1) statistic .90. This value is not statistically positive at any conventional (one-sided) critical value.[5]

We next estimate the exact consumers surplus, here the equivalent variation, across our different estimated demand curves. We consider two sets of price changes for gasoline: an increase from $1.00 to $1.30 (in 1983 $) per gallon and an increase from $1.00 to $1.50 per gallon. The starting price of $1.00 corresponds roughly to 1992 gasoline prices and a 50 cent increase is well within the range of the data. We estimate the equivalent variation for these prices changes at the median of income. We present the estimates in Table 5.2. Estimated standard errors are given in parentheses, and were calculated using the formulas given earlier.

[5]The conventional significance levels may not be appropriate here, because we have chosen the interval based on the estimated demand function. However, the conventional critical values should provide a bound when the test statistic is maximized over choice of interval, with our test being an approximate maximum.

Table 5.2: Yearly Equivalent Variation Estimates

Parametric Estimates	\$1.00-1.30	\$1.00-1.50
1. Log linear model	282.34 (2.07)	442.00 (2.31)
2. Translog quadratic model	285.44 (3.03)	444.16 (3.90)
Epanechnikov Kernel Estimates		
3. h = .45	\$281.31 (4.36)	\$445.09 (4.46)
4. h = .55	\$284.00 (4.11)	\$447.41 (4.11)
5. h = .82	279.87 (3.58)	445.42 (3.59)
Normal, Higher Order Kernel Estimates		
6. h = .45	286.44 (6.56)	450.14 (7.07)
7. h = .55	284.36 (5.35)	445.35 (6.34)
Cubic Spline Estimates		
8. 6 knots	284.58 (4.93)	444.63 (6.31)
9. 7 knots	282.30 (4.68)	441.72 (5.82)
10. 8 knots	287.12 (4.77)	447.80 (6.04)
Power Series Estimates		
11. 8th order	287.31 (4.76)	448.64 (5.91)
12. 9th order	287.27 (4.75)	448.55 (5.88)

Note that all the nonparametric estimates are quite close. A choice of different nonparametric estimator assumptions leads to virtually the same welfare estimates. Surprisingly, although the graphs show quite different shapes for parametric and nonparametric estimates, the welfare estimates are quite similar. A Hausman test statistic based on the difference of the kernel estimate and the difference of their estimated variances, which is valid if the disturbance is homoskedastic, is 1.3, which is not significant for a two-tailed tests based on the normal distribution.

We also estimate the deadweight loss from a rise in the gasoline tax of either $0.30 or $0.50 which would induce the corresponding rise in gasoline prices. We base our estimate of deadweight loss on the equivalent variation measure of the compensating variation. The results are given in Table 5.3:

Table 5.3: Yearly Average Deadweight Loss Estimates

Parametric Estimates	$1.00-1.30	$1.00-1.50
1. Log linear model	26.92 (3.13)	62.50 (6.77)
2. Translog quadratic model	27.07 (3.15)	75.30 (7.24)
Epanechnikov kernel estimates		
3. h = .45	$27.70 (5.65)	$33.41 (8.68)
4. h = .55	27.94 (5.01)	36.90 (7.97)
5. h = .82	22.89 (3.88)	37.10 (6.80)
Normal, higher order kernel estimates		
6. h = .45	$36.38 (6.47)	$33.09 (14.06)
7. h = .55	$31.62 (5.72)	47.61 (8.92)
Cubic Spline Estimates		
8. 6 knots.	28.60 (5.03)	51.05 (8.65)
9. 7 knots	38.68 (5.37)	46.84 (8.77)
10. 8 knots	35.72 (5.27)	46.95 (8.94)
Power Series Estimates		
11. 8th order	34.92 (5.21)	48.66 (8.80)
12. 9th order	34.86 (5.38)	48.74 (8.86)

The estimates of DWL are very similar across bandwidths for the kernel estimator, but are sensitive to the use of a higher order kernel. We give more credence to the higher order kernel estimates, because of their theoretically better property of having smaller bias relative to mean-square error, and their similarity to the spline estimates. The spline results are sensitive to the number of knots. We prefer the results for the larger number of knots, because they are theoretically preferred and the results seem to be less sensitive to choice between 7 and 8 than to 6 and 7.

We do find rather large differences between the nonparametric and parametric estimates of deadweight loss. The estimated differences between the nonparametric estimates and the log linear parametric estimates are in the range of 40-50%. In particular, the nonparametric estimates seem to be somewhat smaller for the larger price change and larger for the smaller price change. These are economically significant differences, with the ratio of DWL to tax revenue varying widely between parametric and nonparametric specifications. The differences are also statistically significant: A Hausman test of based on the log-linear and normal kernel specification is -10.14 for the larger price change and bandwidth .55 and is 5.18 for the smaller price change and bandwidth .45. Thus, efficiency decisions on which commodities to tax and the size of the tax might well depend on the rather sizable differences in the estimated efficiency cost changes from the increased taxes.

6. Asymptotic Distribution Theory

In this section we give results for the kernel estimator of consumer surplus at particular income and covariate values, and for power series. The conclusions for other cases will be described, without full sets of regularity conditions. These results show that the standard error formulas given earlier are correct, and describe the asymptotic properties of the estimators, without overburdening the reader.

Precise conditions are useful for stating the results. Let $p_t(t,u) = \partial p(t,u)/\partial t$ and $g(p,y,w)$ denote $r(x(p,y)) + w'\beta$. Let $\mathcal{U}$ be the support of u, $\mathcal{P} = p([0,1]\times\mathcal{U})$ be the image of $p(t,u)$ on $[0,1]\times\mathcal{U}$, and $\mathcal{W}$ be the support of w. Also, let $\mathcal{Y} = [\underline{y},\bar{y}]$ be a set of y values that includes those where the demand function is evaluated in solving equation (1.1). The conditions will involve supremum norms for the demand function over the set $\mathcal{Z} = \mathcal{P}\times\mathcal{Y}\times\mathcal{W}$. Let $\|g\|_j = \sup_{z\in\mathcal{Z},\ell\le j}\|\partial^\ell g(p,y,w)/\partial(p,y,w)^\ell\|$, where for a matrix function $B(z)$, $\partial^\ell B(z)/\partial z^\ell$ denotes any vector of all distinct ℓ^{th} order partial derivatives of all elements of $B(z)$. Let "a.e." denote almost everywhere with respect to Lebesgue measure.

Assumption 1: $\mathcal{W}$, $\mathcal{U}$, and $\mathcal{Y}$ are compact, $\mathcal{Z}$ is contained in the support of z_i, $\|g_0\|_2 < \infty$, $T(g)$ and $x(p,y)$ are one-to-one and three times continuously differentiable with nonsingular Jacobians on their respective domains, $p(t,u)$ is twice continuously differentiable on $[0,1]\times\mathcal{U}$, and $\mathcal{P}$ does not include zero. Also, $\omega(y)$ is bounded and continuous a.e. and for $\bar{C} = \sup_{[0,1]\times\mathcal{U}\times\mathcal{Y}\times\mathcal{W}}|T(g_0(p(t,u),y,w))'p_t(t,u)|$, $\omega(y) = 0$ for y outside $\mathcal{Y}_\varepsilon = [\underline{y}+\bar{C}+\varepsilon,\bar{y}-\bar{C}-\varepsilon]$, for some $\varepsilon > 0$.

We will derive the results under an i.i.d. assumption and certain moment conditions specified in the following result. Let $\tau(x)$ be a trimming function that will be identically equal to one for series estimators.

Assumption 2: z_i is i.i.d., $E[\|T^{-1}(q)\|^4] < \infty$, x is continuously distributed with bounded density $f_0(x)$, and $E[\tau(x)(w-E[w|x])(w-E[w|x])']$ is nonsingular.

Assumptions 1 and 2 are useful for both the kernel and series results. Following the earlier format, we will first give results for kernel estimators. The next three assumptions are more or less standard conditions for kernel estimators.

Assumption 3: $E[\|T^{-1}(q)\|^4|x]f_0(x)$ is bounded, $\tau(x)$ is bounded, bounded away from zero on $z \in \mathcal{Z}$, and zero outside a compact set on which $f_0(x)$ is positive.

Assumption 4: There is a positive integer s such that $\mathcal{K}(u)$ is twice continuously differentiable, with Lipschitz derivatives, $\mathcal{K}(u)$ is zero outside a bounded set, $\int\mathcal{K}(u)du = 1$, and for all $j < s$, $\int\mathcal{K}(u)[\otimes_{\ell=1}^{j}u]du = 0$.

Assumption 5: There is a nonnegative integer d and extensions of $E[T^{-1}(q)|x]$ and $E[w|x]$ to all of $\mathbb{R}^{k+1}$ such that $f_0(x)$, $f_0(x)r_0(x)$, and $f_0(x)E[w|x]$ are continuously differentiable to order $d \geq s + 2$ on $\mathbb{R}^{k+1}$.

Assumption 4 requires that the kernel be a higher order, bias-reducing type. An example of such a kernel is the Gaussian one used in the application.

To describe the result for the kernel estimator of consumer surplus at a particular income and covariate value, it is necessary to introduce a little more notation. Let $S(t)$ denote the solution to equation (1.1) at the truth, i.e. the solution to $dS/dt = -T(g_0(p(t),y_0-S(t),w_0))'p_t(t)$, $S(1) = 0$, where $p_t(t) = \partial p(t)/\partial t$. Let $x(t) = x(p(t),y_0-S(t))$, and partition $x(t) = (x_1(t),x_2(t)')'$, where $x_1(t)$ is a scalar. Let

$$\zeta(t) = \xi(t)\cdot\exp\{-\int_0^t\xi(v)'[\partial g_0(p(v),y_0-S(v),w_0))/\partial y]dv\},$$

$$\xi(t) = T_g(g_0(p(t),y_0-S(t),w_0))' p_t(t).$$

Let $t(\tau)$ be the inverse function of $x_1(t)$ (which will exist by Assumption 6 below), $x(\tau) = x(t(\tau)) = (\tau, x_2(\tau))'$, and $\zeta(\tau) = \zeta(t(\tau))$.

Assumption 6: i) There is $\eta > 0$ such that $p(t)$ can be extended to a function with domain $\mathcal{D} = [-\eta, 1+\eta]$ such that $x_1(t)$ is one-to-one with derivative bounded away from zero; ii) $E[\varepsilon\varepsilon'|x]$ is continuous a.e. and for some $\eta > 0$, and $(0,\gamma')$ partitioned conformably with $x(\tau) = (\tau, x_2(\tau)')$, $\int_{x_1(\mathcal{D})} \{\sup_{\|\gamma\| \le \eta}(1+E[\|\varepsilon\|^4|x=x(\tau)+(0,\gamma))f_0(x(\tau)+(0,\gamma))\}d\tau < \infty$. iii) f_0 is bounded away from zero on $x(p(\mathcal{D}),[\underline{y},\bar{y}])$, for $\underline{y}$ and $\bar{y}$ from Assumption 1.

Note that $x_2(\tau)$ is differentiable by the inverse function theorem and the chain rule, and let $\tilde{\mathcal{K}}(\tau) = \int[\int\mathcal{K}(v,u+[\partial x_2(\tau)/\partial\tau]v)dv]^2du$. The asymptotic variance of consumer surplus at a point will be

$$V_0 = \int 1(0 \le t(\tau) \le 1)\tilde{K}(\tau)f_0(x(\tau))^{-1}|\partial t(\tau)/\partial t|^2 E[\{\zeta(\tau)'\varepsilon\}^2|x=x(\tau)]d\tau.$$

Let $1(A)$ denote the indicator function for the set A. The following result gives the asymptotic distribution of the kernel estimator of consumer surplus evaluated at a particular income and covariate value.

Theorem 1: If Assumptions 1 - 6 are satisfied, for $\omega(y) = 1(y = y_0)$, $\sigma = \sigma(n)$ *with* $n\sigma^{2k+4}/[\ln(n)]^2 \longrightarrow \infty$, *and* $n\sigma^{2s} \longrightarrow 0$, *then* $\sqrt{n}\sigma^{k/2}(\hat{S}-S_0) \xrightarrow{d} N(0, V_0)$. *If in addition* $n\sigma^{3k+7}/\ln(n) \longrightarrow \infty$, *then for* $\hat{V}$ *in equation (3.5),* $\sigma^k\hat{V} \xrightarrow{p} V_0$.

For our application, where $k = 1$, the conditions on the bandwidth σ are that $n\sigma^{10}/\ln(n) \longrightarrow \infty$ and $n\sigma^{2s} \longrightarrow 0$. These conditions require that $s > 5$, i.e. that the kernel be at least sixth order. The normal kernel used in the application is such a sixth order kernel.

The asymptotic distribution of a kernel estimator of deadweight loss at a particular income and covariate value is straightforward to derive. Because the "tax receipts" term $(p^1-p^0)'T(\hat{g}(p^1,y_0,w_0))$ depends on the demand function evaluated at a particular point, it will have a slower convergence rate than the consumer surplus "integral," and hence will dominate the asymptotic distribution. Then standard results on pointwise convergence of kernel regression estimators can be applied to obtain, for $x^1 = x(p^1,y_0)$,

$$n^{(k+1)/2}(\hat{L} - L_0) \xrightarrow{d} N(0, V_0),$$

$$V_0 = [\int\mathcal{K}(u)^2du]f_0(x^1)^{-1}\cdot E[\{(p^1-p^0)'T_g(g_0(x^1,w_0))\varepsilon\}^2|x=x^1].$$

Also, it is straightforward to show consistency of the asymptotic variance estimator, by means like those used to prove Theorem 1.

As previously noted, average consumer surplus and deadweight loss will be $\sqrt{n}$-consistent if initial and final prices are allowed to vary. In this case the asymptotic distribution will be the same for both kernel and series estimators. This asymptotic distribution will be described below.

Some of the conditions need to be modified for series estimators.

Assumption 7: $E[\varepsilon_i\varepsilon_i'|x_i,w_i]$ is bounded and has smallest eigenvalue that is bounded away from zero.

Let $\|g\|_j$ be as defined above except that the supremum is taken over the support of (x,w), and let $\mathcal{X}$ denote the support of x.

Assumption 8: $\phi_{kK}(x)$ consists of products of powers of the elements of x that are nondecreasing in order as K increases, with all terms of a given order included before the order is increased, $\mathcal{X}$ is a compact rectangle and the density of x is bounded away from zero on $\mathcal{X}$.

The condition that the density is bounded away from zero is useful for controlling the variance of a series estimator. The next condition is useful for controlling the bias.

Assumption 9: $r_0(x)$ and $E[w|x]$ are continuously differentiable of all orders on $\mathcal{X}$ and there is a constant C such that for all integers d the partial derivatives of order d are bounded in absolute value by C^d on $\mathcal{X}$.

This smoothness condition is undoubtedly stronger than necessary. It is used in order to apply second order Sobolev norm approximation rates (i.e. approximation of the function and derivatives up to order 2), where a literature search has not yet revealed such approximation rates for power series except under this hypothesis. Also, results for regression splines are not given here because multivariate Sobolev approximation rates do not seem to be readily available for them.

Average equivalent variation and deadweight loss will be $\sqrt{n}$-consistent under certain conditions that we now describe. Let t_i denote a random variable that is uniformly distributed on $[0,1]$ and independent of z_i, $\tilde{p}_i = p(t_i,u_i)$, $\tilde{y}_i = y_i-S(t_i,z_i,g_0)$, and $\tilde{x}_i = x(\tilde{p}_i,\tilde{y}_i)$. Let $S(t,z) = S(t,z,g_0)$ and

$$\zeta(t,z) = \xi(t,z)\cdot\exp(-\int_0^t \xi(v,z)' g_{0y}(p(v,u),y-S(v,z),w))dv),$$

$$\xi(t,z) = T_g(g_0(p(t,u),y-S(t,z),w))' p_t(t,u).$$

Assumption 10: Conditional on w, $\tilde{x}$ is continuously distributed with bounded density $f^{\mu}(x|w)$ and for the density $f(x|w)$ of x_i given w, $a(x,w) = E[\omega(y_i)\zeta(t_i,z_i)|\tilde{x}_i=x,w_i=w]$ is zero outside the support of $f(x|w)$, and $f(x|w)^{-1}a(x,w)$ is bounded.

Let $\bar{\omega} = E[\omega(y_i)]$ and

$$\tilde{\zeta}^{\mu}(x,w) = f(x|w)^{-1}\tilde{f}^{\mu}(x|w)E[\omega(y_i)\zeta(t_i,z_i)|\tilde{x}_i=x,w_i=w].$$

$$\bar{\zeta}^{\mu}(x,w) = E[\tilde{\zeta}^{\mu}(x,w)|x] + (w-E[w|x])'M^{-1}E[(w-E[w|x])'\tilde{\zeta}^{\mu}(x,w)].$$

$$\psi_i^{\mu} = \bar{\omega}^{-1}\omega(y_i)S(0,z_i,g_0) - \mu_0 + \bar{\omega}^{-1}\mu_0[\omega(y_i) - \bar{\omega}] + \bar{\omega}^{-1}\bar{\zeta}^{\mu}(x_i,w_i)'\varepsilon_i.$$

Then the asymptotic variance of $\hat{\mu}$ will be $E[\psi_i^{\mu}\psi_i^{\mu}{}']$.

Under an additional condition we can also derive the asymptotic variance of deadweight loss. We consider here only the case where there is sufficient variation in the final price to achieve $\sqrt{n}$-consistency. The next assumptions embodies this requirement, by the condition that the final price is continuously distributed.

Assumption 11: Conditional on w, $\bar{x}_i = x(p^1(u_i),y_i)$ is continuously distributed with bounded density $f^{\lambda}(x|w)$, and $\omega(y)f^{\lambda}(x|w)$ is zero outside the support of $f(x|w)$.

Let $T_i = \bar{\omega}^{-1}\omega(y_i)(p^1-p^0)'T(g(p^1(u_i),y_i)))$ and

$$\tilde{\zeta}^{\lambda}(x,w) = f(x|w)^{-1}\tilde{f}^{\lambda}(x|w)E[\omega(y_i)(p^1-p^0)'T_g(g_0(x_i,w_i))|\bar{x}_i=x,w_i=w].$$

$$\bar{\zeta}^{\lambda}(x,w) = E[\tilde{\zeta}^{\lambda}(x,w)|x] + (w-E[w|x])'M^{-1}E[(w-E[w|x])'\tilde{\zeta}^{\lambda}(x,w)].$$

$$\psi_i^{\lambda} = \psi_i^{\mu} - \{T_i - E[T_i] - E[T_i]\bar{\omega}^{-1}[\omega(y_i) - \bar{\omega}] + \bar{\omega}^{-1}\bar{\zeta}^{\lambda}(x_i,w_i)'\varepsilon_i\}.$$

Then the asymptotic variance of $\hat{L}$ will be $E[(\psi_i^L)^2]$.

The next result shows asymptotic normality of the power series estimators.

Theorem 2: Suppose that Assumptions 1 - 2, 7 - 9 are satisfied, with $\tau(x) = 1$, and $K = K(n)$ satisfies $K^{11}/n \longrightarrow 0$ and $Kn^{-\gamma} \longrightarrow \infty$ for some $\gamma > 0$. Then for $\hat{\theta} = \hat{S}$ or $\hat{\theta}$ $\hat{L}$, equation (3.3) will be satisfied and $\hat{\theta} = \theta_0 + O_p(K/\sqrt{n})$. If Assumption 10 is also satisfied and $V_0 = E[(\psi_i^{\mu})^2] > 0$, then $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0,V_0)$ and $\hat{V} \xrightarrow{p} V_0$. If Assumption 11 is also satisfied and $V_0 = E[(\psi_i^{\lambda})^2] > 0$, then $\sqrt{n}(\hat{\lambda} - \lambda_0) \xrightarrow{d} N(0,V_0)$ and $\hat{V} \xrightarrow{p} V_0$.

Appendix: Proofs of Theorems

Throughout the Appendix C will denote a generic positive constant, that may be different in different uses.

One intermediate result that is needed for both kernel and series estimators is a linearization of the solution to equation (1.1) around the true demand function. Some additional notation is needed to set up this linearization. Let z be a data observation that includes u and $S(t,z,g)$ denote the corresponding solution to equation (1.1), i.e. the solution to $\partial S/\partial t = -T(g(p(t,u),y-S,w))$. Let $g_y(p,y,w)$ denote the derivative of $g(p,y,w)$ with respect to y, and

$$(6.1) \qquad \zeta(t,z,g) = \xi(t,z,g)\cdot\exp\{-\int_0^t \xi(v,z,g)' g_{0y}(p(v,u),y-S(v,z,g),w))dv\},$$

$$\xi(t,z,g) = T_g(g(p(t,u),y-S(t,z,g),w))' p_t(t,u).$$

Under conditions specified below, when g is near the truth g_0, equivalent variation can be approximated by the linear functional

$$(6.2) \qquad \Delta(z,g;\tilde{g}) = \int_0^1 g(p(t,u),y-S(t,z,\tilde{g}),w)'\zeta(t,z,\tilde{g})dt,$$

evaluated at $\tilde{g} = g_0$.

Lemma A1: If Assumption 1 is satisfied then there is an $\varepsilon > 0$ *and a constant* C *such that for all* $z \in \mathcal{Z}_\nu$, $\|g-g_0\|_2 < \varepsilon$, *and* $\|\tilde{g}-g_0\|_2 < \varepsilon$, *it is the case that* $|\omega(y)S(0,z,\tilde{g})-\omega(y)S(0,z,g)-\omega(y)\Delta(z,\tilde{g}-g;g)| \le C\|\tilde{g}-g\|_0\|\tilde{g}-g\|_1$, $|\omega(y)\Delta(z,g;g_0)| \le C\|g\|_0$, $|\omega(y)S(0,z,g)-\omega(y)S(0,z,g_0)| \le C\|g-g_0\|_0$, *and for any* $\bar{g}$ *with* $\|\bar{g}\|_1 < \infty$, $|\omega(y)\Delta(z,\bar{g};g)-\omega(y)\Delta(z,\bar{g};g_0)| \le C(\|\bar{g}\|_0\|g-g_0\|_1 + \|\bar{g}\|_1\|g-g_0\|_0)$.

Proof of Lemma A1: By Assumption 1, it suffices to prove the result when $\omega(y) = 1$ for all $y \in \mathcal{Y}_\varepsilon$. We use standard results on existence and continuity of solutions of differential equations, e.g. in Finney and Ostberg (1976). Let $q(t,y,z,g) = T(g(p(t,u),y,w))'p_t(t,u)$. By $T(g)$ thrice continuously differentiable, its derivative up to order 2 are bounded and Lipschitz on any bounded set. Then by $p_t(t,u)$ bounded, we can choose ε so that for $\|\tilde{g}-g_0\|_2 < \delta$, $\|g-g_0\|_2 < \delta$, and δ small enough,

$$\text{(A.1)} \qquad \sup_{[0,1]\times\mathcal{Y}\times\mathcal{Z}}|\partial^j q(t,y,z,\tilde{g})/\partial y^j - \partial^j q(t,y,z,g)/\partial y^j| \le C\|\tilde{g}-g\|_j. \quad (j = 0,\ 1,\ 2).$$

In particular, for δ small enough, $\sup_{[0,1]\times\mathcal{Y}\times\mathcal{Z}}|q(t,y,z,g)| \le \bar{C}+\varepsilon$. Also, by construction of $\mathcal{Y}_\varepsilon$, $(p(t,u),y-s) \in \mathcal{P}\times\mathcal{Y}$ for $(t,u,y,s) \in [0,1]\times\mathcal{U}\times\mathcal{Y}_\varepsilon\times[-\bar{C},\bar{C}]$. Then by Theorem 12-6 of Finney and Ostberg (1976), there exists a solution $S(t,z,g)$ to $\partial S(t,z,g)/\partial t = -q(t,y-S(t,z,g),z,g)$, $S(1,z,g) = 0$, for $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$, and y now included in z. Furthermore, by integration of equation (1.1) on $t \in [0,1]$, $|S(t,z,g)| < \bar{C}+\varepsilon$, so that $y-S(t,z,g) \in \mathcal{Y}$ for $z \in \mathcal{Z}_\varepsilon$. Also, the same existence and boundedness properties hold for $\tilde{g}$ replacing g in q and S.

Next, $\|g\|_2 < \infty$ by $\|g_0\|_2 < \infty$ and $\|g-g_0\|_2 < \delta$. Also, by $p(t,u) \in \mathcal{P}$ and $y-S(t,z,g) \in \mathcal{Y}$ for $t \in [0,1]$, and $z \in \mathcal{Z}_\varepsilon$, it follows that $g(p(t,u),y-S(t,z,g),w)$ and $g_y(p(t,u),y-S(t,z,g),w)$ are bounded on this set. Then boundedness of $\zeta(t,z,g)$ follows by by $T_g(g)$ bounded on any compact set and boundedness of $p_t(t,u)$, giving the second conclusion.

Next, it follows by Theorem 12-9 (equation 12-22) of Finney and Ostberg (1976), $T(g)$ Lipschitz on any bounded set, $p_t(t,u)$ bounded, and eq. (A.1) that

$$\text{(A.2)} \qquad \sup_{t\in[0,1],z\in\mathcal{Z}_\varepsilon}|S(t,z,\tilde{g})-S(t,z,g)| \le C\|\tilde{g}-g\|_0,$$

which implies the third conclusion.

Next, for all $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$, by $|S(t,z,g)| \le \bar{C}+\varepsilon$ and $|S(t,z,g_0)| < \bar{C}$, it follows that $\|\bar{g}(p(t,u),y-S(t,z,g),w)\| \le \|\bar{g}\|_0$, and by a mean-value expansion and eq. (A.2) that $\|\bar{g}(p(t,u),y-S(t,z,g),w) - \bar{g}(p(t,u),y-S(t,z,g_0),w)\| \le \|\bar{g}_y(p(t,u),y-\bar{S}(t,z,g,g_0),w)\||S(t,z,g)-S(t,z,g_0)| \le \|\bar{g}\|_1\|g-g_0\|_0$ for an intermediate value

$\bar{g}$. Then by boundedness of $\zeta(t,z,g_0)$,

$$\text{(A.3)} \qquad |\Delta(z,\bar{g};g)-\Delta(z,\bar{g};g_0)| \le \|\bar{g}\|_0 \sup_{t\in[0,1],z\in\mathcal{Z}_\varepsilon} \|\zeta(t,z,g)-\zeta(t,z,g_0)\| + C\|\bar{g}\|_1\|g-g_0\|_0.$$

Also, $\tilde{g}(p(t,u),y-S(t,z,g),w))$, $\tilde{g}_y(p(t,u),y-S(t,z,g),w))$, and $\tilde{g}_{yy}(p(t,u),y-S(t,z,g),w))$ are all bounded for uniformly in $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$, $\|\tilde{g}-g_0\|_2 < \varepsilon$ and $\|g-g_0\|_2 < \varepsilon$, as are the same expressions with $S(t,z,g)$ replaced by a value in between $S(t,z,g)$ and $S(t,z,g_0)$. Also, it then follows by mean-value expansion arguments like those above, including expansions of in $S(t,z,g)$ around $S(t,z,g_0)$, that uniformly in $\|g-g_0\|_2 < \varepsilon$,

$$\text{(A.4)} \qquad \sup_{t\in[0,1],z\in\mathcal{Z}_\varepsilon} \|\zeta(t,z,g)-\zeta(t,z,g_0)\| \le C\|g-g_0\|_1.$$

For example, for a value $\bar{S}(t,z,g,g_0)$ in between $S(t,z,g)$ and $S(t,z,g_0)$, $\|g_y(p(t,u),y-S(t,z,g),w)) - g_{0y}(p(t,u),y-S(t,z,g_0),w))\| \le \|g_y(p(t,u),y-S(t,z,g),w)) - g_{0y}(p(t,u),y-S(t,z,g),w))\| + \|g_{0y}(p(t,u),y-S(t,z,g),w)) - g_{0y}(p(t,u),y-S(t,z,g_0),w))\| \le \|g-g_0\|_1 + \|g_{0yy}(p(t,u),y-\bar{S}(t,z,g,g_0),w))\|\cdot \|S(t,z,g)-S(t,z,g_0)\| \le C\|g-g_0\|_1$. The fourth conclusion then follows by eq. (A.3).

Finally, to show the first conclusion, let $D(t,z,\tilde{g},g) = S(t,z,\tilde{g})-S(t,z,g)$. For notational convenience, suppress the t and z arguments, and let $\tilde{S} = S(t,z,\tilde{g})$ and $S = S(t,z,g)$. Differencing the differential equation gives

$$\text{(A.5)} \qquad \partial D/\partial t = -q(y-\tilde{S},\tilde{g}) + q(y-S,g) = -[q(y-S,\tilde{g}) - q(y-S,g)] - [q(y-\tilde{S},g) - q(y-S,g)]$$

$$- \{q(y-\tilde{S},\tilde{g})-q(y-\tilde{S},g) - [q(y-S,\tilde{g}) - q(y-S,g)]\}$$

$$= -\xi(g)'\{\tilde{g}(y-S)-g(y-S)\} - q_y(y-S,g)D - R(g,\tilde{g}),$$

$$R(g,\tilde{g}) = [q(y-S,\tilde{g})-q(y-S,g)-\xi(g)'\{\tilde{g}(y-S)-g(y-S)\}] + [q(y-\tilde{S},g)-q(y-S,g)-q_y(y-S,g)D]$$

$$+ [q(y-\tilde{S},\tilde{g})-q(y-\tilde{S},g)-q(y-S,\tilde{g})+q(y-S,g)] = R_1(g,\tilde{g}) + R_2(g,\tilde{g}) + R_3(g,\tilde{g}).$$

The first equation here is an inhomogeneous linear differential equation, with final condition $D|_{t=1} = 0$, nonconstant coefficient $-q_y(y-S,g)$, and nonconstant shift $-\xi(g)'\{\tilde{g}(y-S)-g(y-S)\} + R(g,\tilde{g})$. Let $\nu(t,z,g) = \exp[-\int_0^t q_y(r,y-S(r,z,g),z,g)'\xi dr]$??. Then the solution to this linear equation at $t = 0$ is

(A.6) $$D|_{t=0} = \int_0^1[\xi(g)'\{\tilde{g}(y-S)-g(y-S)\} + R(g,\tilde{g})]\nu(t,z,g)dt = \Delta(z,\tilde{g}-g) + \int_0^1 R(g,\tilde{g})\xi(t,z,g)dt.$$

By $\tilde{g}(y-S)$ and $g(y-S)$ bounded and T_g twice continuously differentiable, the elements of $\partial^2T(\bar{g}(y-S))/\partial g^2$ will be bounded on $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$, for any $\bar{g}$ on a line joining $\tilde{g}$ and g (that may differ from element to element of $\partial^2T(g)/\partial g^2$). Then by a mean-value expansion, for all $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$,

(A.7) $$|R_1(g,\tilde{g})| \le C\|p_t\|\|\partial^2T(\bar{g})/\partial g^2\|\|\tilde{g}(y-S)-g(y-S)\|^2 \le C\|\tilde{g}-g\|_0^2.$$

By $|\tilde{S}| < \bar{C}+\varepsilon$ and $|S| < \bar{C}+\varepsilon$, $q(y-s,\tilde{g}) - q(y-s,g)$ is differentiable in an open interval containing $\tilde{S}$ and S. Let $\bar{S} = \bar{S}(t,z,\tilde{g},g)$ be the mean value for an expansion of $q(y-\tilde{S},g)$ around S, with $\bar{S}$ between S and $\tilde{S}$, so that $y-\bar{S} \in \mathcal{Y}$ and $|\bar{S}-S| \le |D|$. A similar statement holds for the mean value S^* of an expansion of $q_y(y-\bar{S},g)$ around S. Then for all $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$,

(A.8) $$|R_2(g,\tilde{g})| \le |q_y(y-\bar{S},g)-q_y(y-S,g)||D| \le |q_{yy}(y-S^*,g)||D|^2 \le C\|\tilde{g}-g\|_0^2.$$

Similarly, for a mean-value expansion of $q(y-\tilde{S},\tilde{g}) - q(y-\tilde{S},g)$ around S, for all $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$,

(A.9) $$|R_3(g,\tilde{g})| \le |q_y(y-\bar{S},\tilde{g})-q_y(y-\bar{S},g)||\tilde{S}-S| \le C\|\tilde{g}-g\|_1\|\tilde{g}-g\|_0.$$

where $\bar{g}$, $\bar{S}$, and $\tilde{S}$ denote mean values. Then combining eqs. (A.7) - (A.9) and noting that $\xi(t,z,g)$ is bounded uniformly in $t \in [0,1]$, $z \in \mathcal{Z}_\varepsilon$, and $\|g-g_0\| < \varepsilon$, we have $\int_0^1 R(g,\tilde{g})\xi(t,z,g)dt \le C\|\tilde{g}-g\|_1\|\tilde{g}-g\|_0$, so the first conclusion follows by eq. (A.6). QED.

Proof of Theorem 1: We first consider $\hat{S}$, and proceed by using the Lemmas of Section 5 of

Newey (1992a) (N henceforth). Let f, h^q, and h^w denote possible values for the functions $f(x)$, $f(x)E[T^{-1}(q)|x]$ and $f(x)E[w|x]$ respectively, and $h = (f,h^q,h^w)$. Let $g(z;\beta,h) = f(x)^{-1}[h^q(x)-\beta' h^w(x)] + \beta' w$ and for any $\tilde{h}$ let $L(x,h;\beta,\tilde{h}) = \tilde{f}(x)^{-1}[h^q(x)-\beta' h^w(x)] - \tilde{f}(x)^{-2}[\tilde{h}^q(x)-\beta'\tilde{h}^w(x)]f(x)$. Let β and h in N equal $(\theta,\beta')'$ and h here. Let $m(z,\beta,h) = (m_1(z,\beta,h),m_2(z,\beta,h)')'$ there be $m_1(z,\beta,h) = S(0,y_0,w_0,g(\cdot;\beta,h)) - \theta$ and $m_2(z,\beta,h) = \tau(x)[q-g(z;\beta,h)]\otimes[w-f^{-1}(x)h^w(x)]$. For $\Delta(z,g;\tilde{g})$ from equation (A.2) let $D_1(z,h;\tilde{h},\beta) = \Delta(z,L(\cdot,h;\beta,\tilde{h});g(\cdot;\beta,\tilde{h}))$ and $D_2(z,h;\tilde{h},\beta) = -\tau(x)\tilde{f}(x)^{-1}[q-g(z;\beta,h)]\otimes[h^w(x)-\tilde{f}^{-1}(x)\tilde{h}^w(x)f(x)] - \tau(x)L(x,h;\beta,\tilde{h})\otimes[w-\tilde{f}^{-1}(x)\tilde{h}^w(x)]$. By Assumption 1 with $\omega(y) = 1(y = y_0)$ it follows that $y_0 \in \mathcal{Y}_\varepsilon$. Let $\|g\|_j$ be as defined preceding Assumption 1, and $\|h\|_j = \sup_{x\in x(\mathcal{P}x[\underline{y},\bar{y}]),\ell\le j}\|\partial^\ell h(x)/\partial x^\ell\|$. Then by the hypothesis that the density of x is bounded away from zero on $x(\mathcal{P}x[\underline{y},\bar{y}])$, it follows by a straightforward application of the quotient rule for derivatives that if $\|h-h_0\|_j$, $\|\tilde{h}-h_0\|_j$, and $\|\beta-\beta_0\|$ are small enough then $\|g-\tilde{g}\|_j \le C\|h-\tilde{h}\|_j$ for $g = g(\cdot;h,\beta)$ and $\tilde{g} = g(\cdot;\tilde{h},\beta)$. Also, it follows by the usual mean value expansion for ratios that for such h and $\tilde{h}$, $\|g-\tilde{g}-L(\cdot,h-\tilde{h};\beta,\tilde{h})\|_0 \le \|h-\tilde{h}\|_0^2$. Then by the conclusion of Lemma 1 and by $\|h-\tilde{h}\|_0 \le \|h-\tilde{h}\|_1$, for $\|\tilde{h}-h_0\|_2$ and $\|h-h_0\|_2$ small enough,

$$\text{(A.10)}\qquad |m_1(z,\beta,h)-m_1(z,\beta,\tilde{h})-D_1(z,h-\tilde{h};\beta,\tilde{h})|$$

$$\le |\Delta(z,g-\tilde{g};\tilde{g})-\Delta(z,L(\cdot,h-\tilde{h};\beta,\tilde{h}),\tilde{g})| + C\|g-\tilde{g}\|_0\|g-\tilde{g}\|_1$$

$$\le C\|g-\tilde{g}-L(\cdot,h-\tilde{h};\beta,\tilde{h})\|_0 + C\|h-\tilde{h}\|_0\|h-\tilde{h}\|_1 \le C\|h-\tilde{h}\|_0\|h-\tilde{h}\|_1.$$

It also follows by similar reasoning that for $\|\beta-\beta_0\|$ and $\|\tilde{h}-h_0\|_2$ small enough,

$$\text{(A.11)}\qquad |D_1(z,\bar{h};\beta,\tilde{h})| = |\Delta(z,L(\cdot,\bar{h};\beta,\tilde{h});\tilde{g})| \le C\|L(\cdot,\bar{h};\beta,\tilde{h})\|_0 \le C\|\bar{h}\|_0,$$

$$|D_1(z,\bar{h};\beta,\tilde{h})-D_1(z,\bar{h};\beta_0,h_0)| = C\|\bar{h}\|_0\|\tilde{h}-h_0\|_1 + C\|\bar{h}\|_1(\|\tilde{h}-h_0\|_0+\|\beta-\beta_0\|),$$

$$|m_1(z,\beta,h)-m_1(z,\beta_0,h_0)| \le C\|h-h_0\| + C\|\beta-\beta_0\|.$$

It also follows by straightforward algebra that for $\|\beta-\beta_0\|$, $\|h-h_0\|_0$, and $\|\tilde{h}-h_0\|_2$ small enough, there is $b(z) = C(1+\|q\|)$ such that

$$\text{(A.12)} \qquad \|m_2(z,\beta,h)-m_2(z,\beta,\tilde{h})-D_2(z,h-\tilde{h};\beta,\tilde{h})\| \le b(z)(\|h-h_0\|_0)^2, \quad |D_2(z,\bar{h};\beta,\tilde{h})| \le b(z)\|\bar{h}\|_0,$$

$$\|D_2(z,\bar{h};\beta,\tilde{h})-D_2(z,\bar{h};\beta_0,h_0)\| \le b(z)\|\bar{h}\|_0\|\tilde{h}-h_0\|_0,$$

$$\|m_2(z,\beta,h)-m_2(z,\beta_0,h_0)\| \le b(z)(\|h-h_0\|_0 + \|\beta-\beta_0\|).$$

Furthermore, $E[b(z)^4] < \infty$ and for $\eta_n^j = [\ln(n)/(n\sigma^{k+2j})]^{1/2}+\sigma^s$ and $\alpha = m/2$, we have $\eta \longrightarrow 0$, $\sqrt{n}(\eta_n^0)^2 \longrightarrow 0$, $\sqrt{n}\sigma^{\alpha}\eta_n^0\eta_n^1 \longrightarrow 0$, implying that $1/(\sqrt{n}\sigma^k) \le \ln(n)/(\sqrt{n}\sigma^k) \longrightarrow 0$. Therefore, the hypotheses of Lemma 5.4 of N are satisfied, giving

$$\text{(A.13)} \qquad \sqrt{n}\sigma^{\alpha_\ell}\textstyle\sum_{i=1}^n[m_\ell(z_i,\hat{h},\beta_0)-m_\ell(z_i,h_0,\beta_0)]/n = \sqrt{n}\sigma^{\alpha_\ell}[m_\ell(\hat{h})-m_\ell(h_0)] + o_p(1), \quad (\ell = 1,\ 2),$$

for $\alpha_1 = \alpha$ and $\alpha_2 = 0$, and $m_\ell(h) = \int D_\ell(z,h;h_0,\beta_0)dF(z)$.

Next, by hypothesis, $m_2(h) = 0$. Let $\zeta(\tau) = \zeta(t(\tau),y_0,w_0,g_0)$ and $f_\tau(\tau) = 1(0\le t(\tau)\le 1)|\partial t(\tau)/\partial\tau|$ denote the density of τ when t is uniformly distributed on $[0,1]$. By the inverse function theorem, $f_\tau(\tau)$ is bounded and continuous a.e. with compact support. Then by the definition of Δ and L,

$$\text{(A.14)} \qquad m_1(h) = \int\Delta(z,L(\cdot,h;\beta_0,h_0);g_0)dF(z)$$

$$= \int\omega(t)h(x(t))dt, \quad \omega(\tau) = f_0(x(\tau))^{-1}\zeta(\tau)'[-r_0(x(\tau)),I,-\beta_0']f_\tau(\tau).$$

By Assumptions 5 and 6, $\omega(\tau)$ is bounded, continuous almost everywhere, and zero outside $\tau([0,1])$. Also, by the inverse function theorem and the chain rule, $x_2(\tau)$ is continuously differentiable with bounded derivatives on $\tau(\mathcal{D})$, a compact convex set containing $\tau([0,1])$ in its interior. By the above shown conditions, $n\sigma^{2k} \longrightarrow \infty$ and $n\sigma^{2s} \longrightarrow 0$. Then it follows by Lemma 5.4 of N that $\sqrt{n}\sigma^{\alpha}[m_1(\hat{h})-m_1(h_0)] \xrightarrow{d} N(0,V_0)$. It then follows by equation (A.13), $\sigma \longrightarrow 0$, and the triangle inequality that $\sqrt{n}\sigma^{\alpha}\sum_i m_1(z_i,\beta_0,\hat{h})/n \xrightarrow{d} N(0,V_0)$. Furthermore, by equation (A.13), $m_2(\hat{h}) = 0$, and $m_2(h_0) = 0$ it follows that $\sqrt{n}\sigma^{\alpha}\sum_i m_2(z_i,\beta_0,\hat{h})/n \xrightarrow{p} 0$.

Next, note that by Lemma A1, for $\|\beta-\beta_0\|$ and $\|h-h_0\|_2$ small enough, $m_1(z,\beta,h)$ is differentiable in β with derivative that is bounded uniformly in β and h, and derivative with respect to θ equal to -1. Also, $m_2(z,\beta,h)$ is linear in β, and it is straightforward to show that $n^{-1}\sum_{i=1}^{n}\partial m_2(z_i,\beta,\hat{h})/\partial\beta$ converges in probability to $E[\partial m_2(z,\beta_0,h_0)/\partial\beta]$, with a first column of zeros, and the remaining columns being nonsingular by Assumption 2. The first conclusion then follows by a Taylor expansion.

To show the second conclusion it is useful to verify Assumption 5.2 of N for both m_1 and m_2. For m_1, parts i) - iii) of Assumption 5.2 follow by eqs. (A.10) - (A.11), with $\Delta = 0$, $\Delta_1 = \Delta_2 = 1$, and $\Delta_3 = 0$. Also, part iv) is satisfied with $\delta_{\beta n} = 1/2$, by the conditions that $n\sigma^{3k+4}/\ln(n) \longrightarrow \infty$ and $n\sigma^{2s} \longrightarrow 0$. It then follows by Lemma 5.5 of N that $\sigma^{2\alpha}\sum_{i=1}^{n}\hat{U}_i^2/n \xrightarrow{p} V_0$ for $\hat{U}_i = \hat{\psi}_i^z + \hat{A}_j - \sum_{j=1}^{n}\hat{A}_j/n$. Also, it is straightforward to check that Assumption 5.2 of N is satisfied for m_2, so that for $\bar{\psi}_i = \tau_i(w_i-\hat{E}[w|x_i])(T^{-1}(q_i)-\hat{g}(x_i,w_i)))$, $\sigma^{2\alpha}\sum_{i=1}^{n}\|\bar{\psi}_i\|^2/n \xrightarrow{p} 0$. Also, by arguments similar to those for asymptotic normality, $\hat{G}^{\beta}$ is bounded in probability and $\hat{M} \xrightarrow{p} M$, so the second conclusion follows by the triangle inequality. QED.

Proof of Theorem 2: The proof proceeds by verifying the hypotheses of Theorem A.1 of Newey (1992b), that will henceforth be referred to as N. Assumption A.1 of N follows by Assumption 7 and boundedness of $a(z,g_0)$ in each case. Assumption A.2 of N follows by Assumption 8 here and Lemma 8.4 of Newey (1991), with $\|\tilde{\phi}^K\|_j \le CK^{1+2j}$, $j = 0, 1, 2$. Assumption A.3 of N follows by Assumption 9 here and Lemma 8.2 of Newey (1991), with α_d equal to any positive number. Assumption A.4 of N, with $\Delta = 2$, $\Delta_1 = 0$, $\Delta_2 = \Delta_3 = 1$, follows by Lemma A1 for each case. Furthermore, by $Kn^{-\gamma} \longrightarrow \infty$ for some $\gamma > 0$, it follows that for any Γ_1, Γ_2 >, there are α_d such that $n^{\Gamma_1}K^{\Gamma_2}K^{-\alpha_d} \longrightarrow 0$. Therefore,

$$\|\tilde{\phi}^K\|_2[K^{1/2}/\sqrt{n} + K^{-\alpha_2}] \le C\cdot K^5\cdot K^{1/2}/\sqrt{n} + o(1) = o(1),$$

$$\sqrt{n}\|\tilde{\phi}^K\|_0\|\tilde{\phi}^K\|_1(K^{1/2}/\sqrt{n} + K^{-\alpha_0})(K^{1/2}/\sqrt{n} + K^{-\alpha_1}) \le CK\cdot K^3\cdot K/\sqrt{n} + o(1) = o(1),$$

$$K^{1/2}\|\tilde{\phi}^K\|_0^2/\sqrt{n} \le CK^{1/2}K^2/\sqrt{n} = o(1), \quad K^{1/2}\|\tilde{\phi}^K\|_1/\sqrt{n} \le CK^{1/2}K^3/\sqrt{n} = o(1), \quad \sqrt{n}K^{-\alpha_0} = o(1),$$

so that the rate hypotheses of Theorem A.1 of N are satisfied. Therefore, it suffices to show that for $\hat{\theta} = \hat{S}$ or $\hat{\theta} = \hat{L}$, Assumption A.6 or A.7 of N is satisfied, while for $\hat{\theta} = \hat{\mu}$ or $\hat{\theta} = \hat{\lambda}$, Assumption A.5 of N is satisfied.

For $\hat{S}$, the $A(z,g;g_0)$ from N is $\int\zeta(\tau)'[r(x(p(\tau),y_0-S(\tau))+w_0'\beta]d\tau$. Consider g values with $\beta = 0$. It follows as in the proof of Lemma A.4 of N that Assumption A.6 of N is satisfied. For $\hat{L}$, $x_0 = x(p^1,y_0)$ and $T_{g0} = T_g(g_0(x_0,w_0))$,

$$E[A(z,r(x);g_0)] = \int\zeta(\tau)'r(x(\tau))f_\tau(\tau)d\tau - (p^1-p^0)'T_{g0}r(x_0).$$

By T_{g0} nonsingular and $p^1 \ne p^0$ there exists $\bar{r}$ such that $(p^1-p^0)'T_{g0}\bar{r} \ne 0$. Also by τ continuously distributed and $p(t)$ one-to-one on $(0,1)$, $\text{Prob}(x(\tau)=x_0) = 0$. Therefore, by reasoning similar to that in the proof of Lemma A.4 of N, there exists $r_j(x)$ with $r_j(x_0) = \bar{r}$, that is everywhere continuous, bounded uniformly in j, converges to zero for all $x \ne x_0$, as $j \longrightarrow \infty$, and hence $\int\zeta(\tau)'r(x(\tau))f_\tau(\tau)d\tau \longrightarrow 0$, so there exists $r_K(x) = \Phi^K(x)'\eta_K$ such that $E[\|\Phi^K(x)'\eta_K\|^2] \longrightarrow 0$ and $E[A(z,r_K(x);g_0)] \longrightarrow -(p^1-p^0)'T_{g0}\bar{r} \ne 0$, so that Assumption A.6 of N is satisfied.

For $\hat{\mu}$, $A(z,g;g_0)$ in N is $\bar{\omega}^{-1}\omega(y)\Delta(z,g;g_0)$ here, so that

$$\bar{\omega}E[A(z,g;g_0)] = E[\omega(y_i)g(\tilde{x}_i,w_i)'\zeta(t_i,z_i)] = E[E[\omega(y_i)\zeta(t_i,z_i)'\,|\,\tilde{x}_i,w_i]g(\tilde{x}_i,w_i)]$$

$$= E[\int E[\omega(y_i)\zeta(t_i,z_i)'\,|\,\tilde{x}_i=x,w_i=w]g(x,w)\tilde{f}(x|w)dx]$$

$$= E[\int\tilde{\zeta}(x,w)'g(x,w)f(x|w)dx] = E[\tilde{\zeta}(x_i,w_i)'g(x_i,w_i)] = E[\bar{\zeta}(x_i,w_i)'g(x_i,w_i)],$$

where the last equality follows by straightforward calculation. It also follows similarly that Assumption A.5 of N is satisfied for $\hat{\lambda}$. Then the conclusion follows by Theorem A.1 of N. QED.

References

Afriat, S. (1967): "The Construction of a Utility Function from Expenditure Data", *International Economic Review*, 8, 67-77.

Afriat, S. (1973): "On a System of Inequalities on Demand Analysis: An Extension of the Classical Method", *International Economic Review*, 14, 460-472.

Auerbach, A. (1985): "The Theory of Excess Burden and Optimal Taxation", in A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, vol. I, Amsterdam: North Holland.

Bickel, P.J. (1992): "Consistent Bandwidth Selection in Nonparametric Regression," *Journal of the American Statistical Association*.

Bierens, H.J. (1987): "Kernel Estimators of Regression Functions", in T. Bewley, *Advances in Econometrics*, Fifth World Congress, vol I, Cambridge: Cambridge University Press.

Brown, B.W. and Walker (1990): "Individual Heterogeneity and Heteroskedasticity in Demand Equations," *Econometrica*.

Brown, B.W. and W.K. Newey (1992): "Efficient Semiparametric Estimation of Expectations," working paper, Department of Economics, Rice University.

Deaton, A. (1986): "Demand Analysis", in Z. Griliches and M. Intriligator, *Handbook of Econometrics*, vol. III, Amsterdam: North Holland.

Diamond, P. and D. McFadden (1974): "Some Uses of the Expenditure Function in Public Finance," *Journal of Public Economics*, 3, 3-21.

Finney, R.L. and D.R. Ostberg (1976): *Elementary Differential Equations with Linear Algebra*, Reading: Addison-Wesley.

Hardle, W. (1990): *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

Hausman, J. (1981): "Exact Consumer's Surplus and Deadweight Loss", *American* Economic *Review*, 71, 662-676.

Hausman, J. (1985): "The Econometrics of Nonlinear Budget Sets," *Econometrica*, 53, 1255-1282.

Newey, W.K. (1991): "Consistency and Asymptotic Normality of Nonparametric Projection Estimators," MIT Department of Economics Working Paper No. 584, June 1991.

Newey, W.K. (1992a): "Kernel Estimation of Partial Means and a General Asymptotic Variance Estimator," MIT Department of Economics Working Paper, January.

Newey, W.K. (1992b): "Asymptotic Normality of Nonlinear Functionals of Series Estimators," Department of Economics Working Paper, January.

Porter-Hudak, S. and K. Hayes (1986): "The Statistical Precision of a Numerical Methods Estimator as Applied to Welfare Loss," *Economics Letters*, 20, 255-257.

Powell, M.J.D. (1981): *Approximation Theory and Methods*, Cambridge, England: Cambridge

University Press.

Powell, J.L., J.H. Stock, and T.M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica,* 57, 1403-1430.

Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica,* 56, 931-954.

Samuelson, P.A. (1948): "Consumption Theory in Terms of Revealed Preference", *Economica*, 5, 61-71

Small, K. and H. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models", *Econometrica*, 49, 105-130.

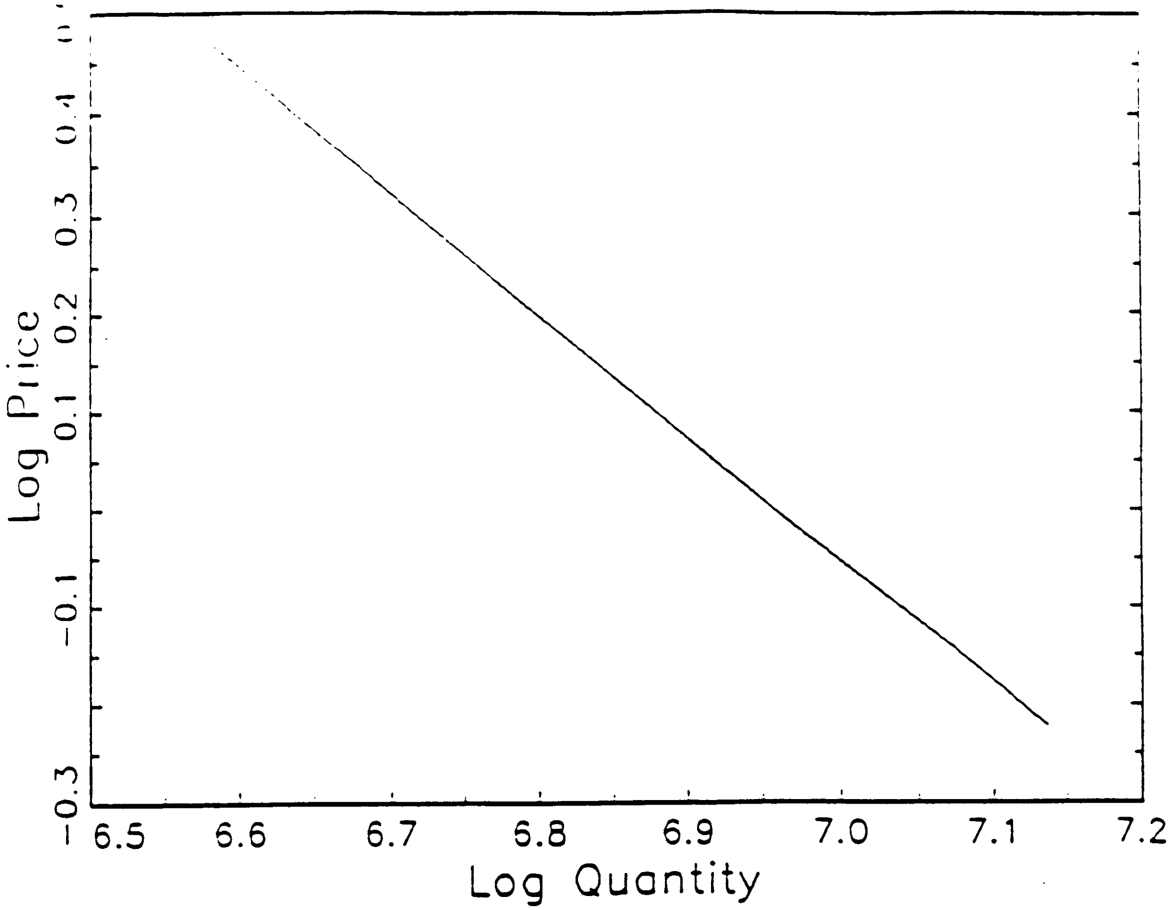
Varian, H. (1982a): "The Nonparametric Approach to Demand Analysis", *Econometrica*, 50, 945-974.

Varian, H. (1982b): "Nonparametric Tests of Models of Consumer Behavior," *Review* of Economic *Studies*, 50, 99-110.

Varian, H. (1982c): "Trois Evaluations de l'impact 'Social' d'un Changement de Prix", *Cahiers du Seminar d'Econometrie* 24, 13-30.

Vartia, Y. (1983): "Efficient Methods of Measuring Welfare Change and Compensated Income in terms of Ordinary Demand Functions", *Econometrica*, 51, 79-98.

Willig, R. (1976): "Consumer's Surplus without Apology," *American Economic Review,* 66, 589-597.



Translog Demand

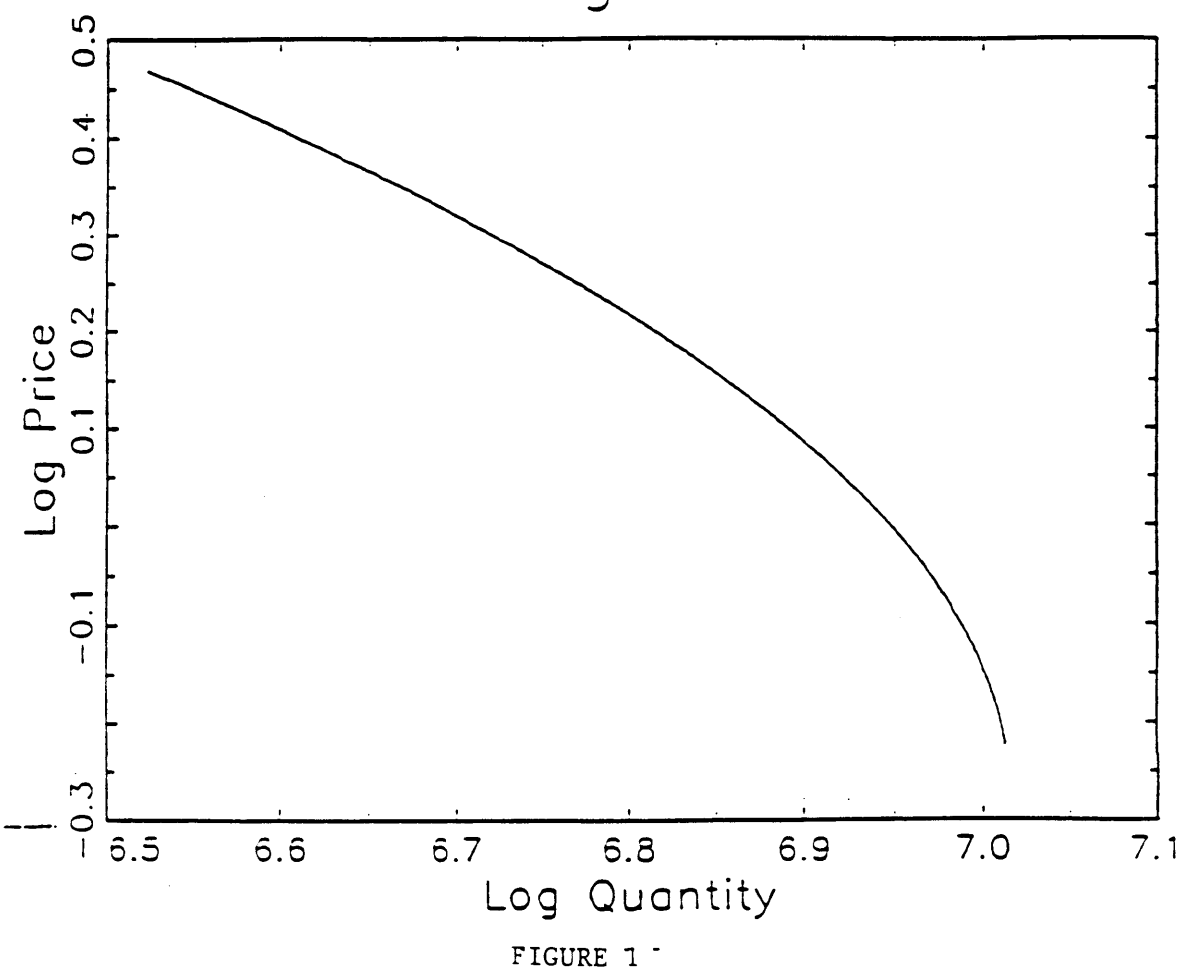
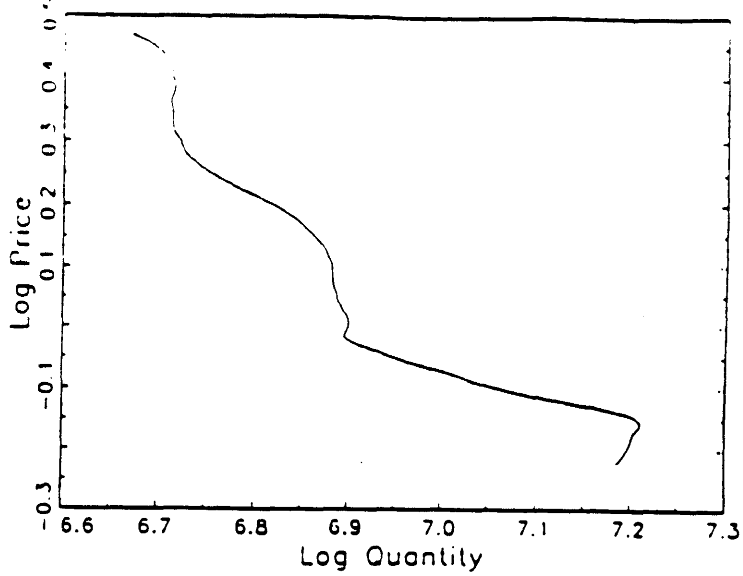
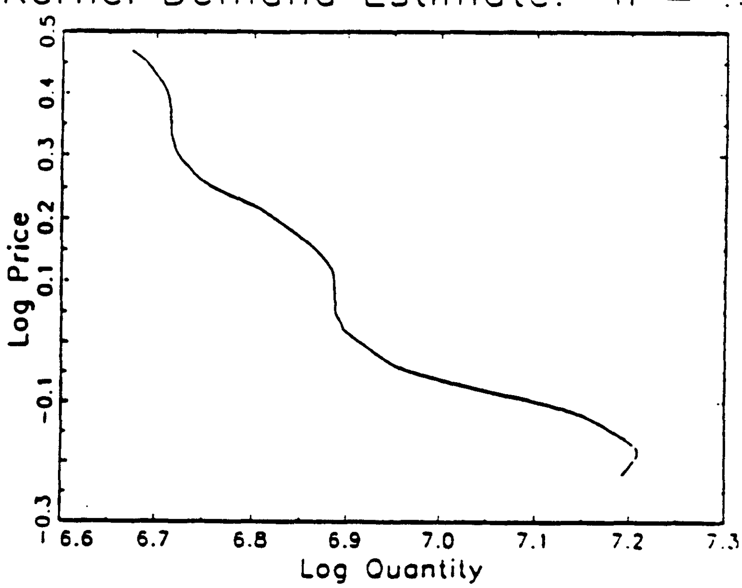


FIGURE 1



Kernel Demand Estimate: h = .55



Kernel Demand Estimate: h = .82

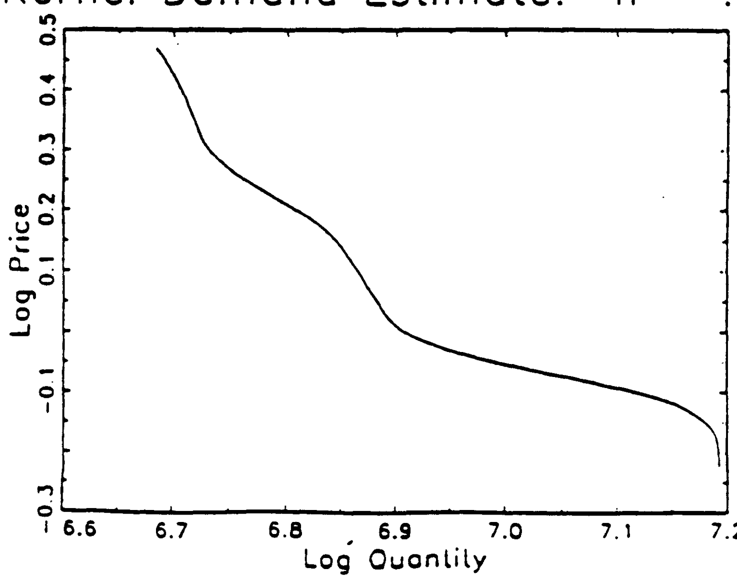
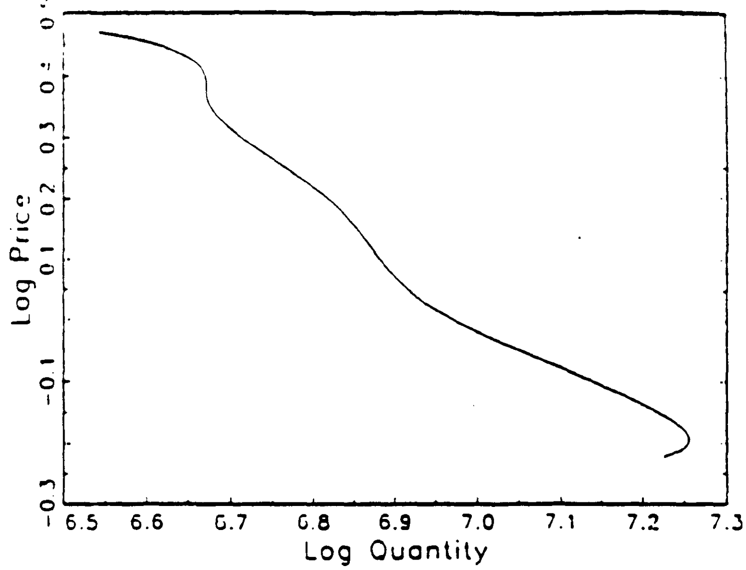
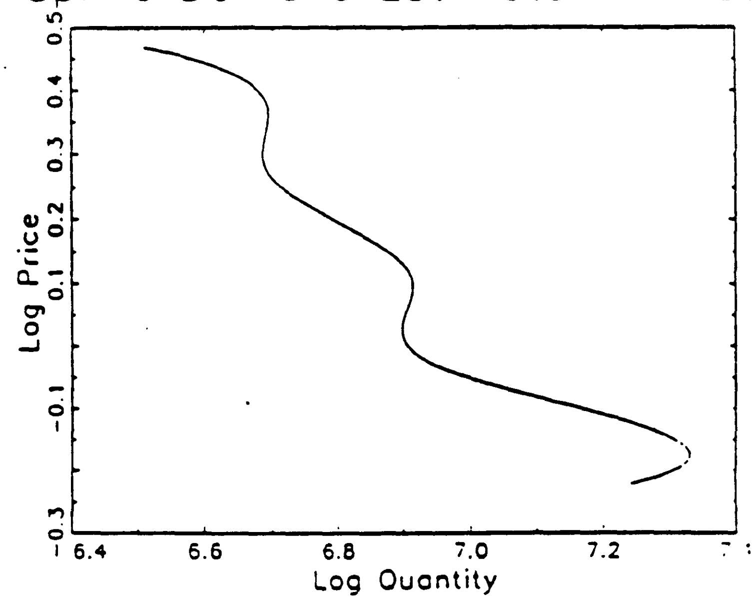


FIGURE 2

Spline Demand Estimate: 6 Knots



Spline Demand Estimate: 7 Knots



Spline Demand Estimate: 8 Knots

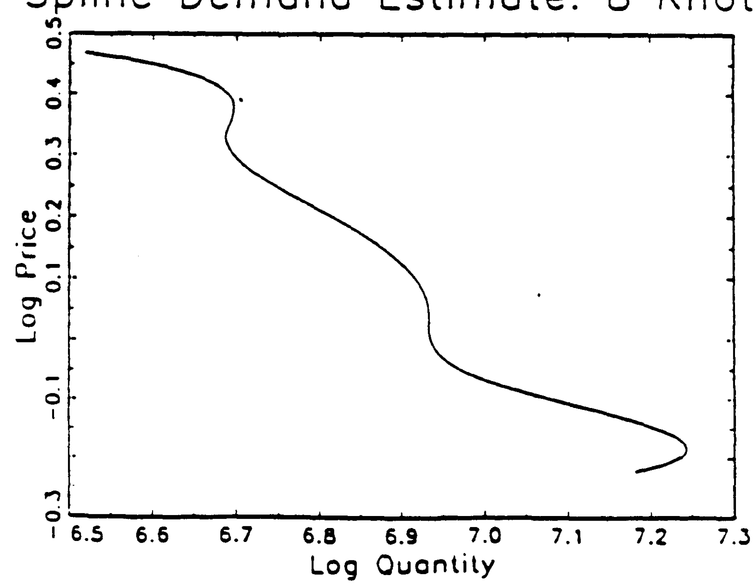


FIGURE 3