ESTIMATING FEATURES OF A DISTRIBUTION FROM BINOMIAL DATA*

Arthur Lewbel[†]
Boston College

Oliver Linton[‡]
London School of Economics

Daniel McFadden
University of California, Berkeley[§]

Revised September 2002

Abstract

A statistical problem that arises in several fields is that of estimating the features of an unknown distribution, which may be conditioned on covariates, using a sample of binomial observations on whether draws from this distribution exceed threshold levels set by experimental design. Applications include bioassay and destructive duration analysis. Another application is referendum contingent valuation in resource economics, where one is interested in features of the distribution of values placed by consumers on a public good such as endangered species. Sample consumers are asked whether they favor a referendum that would provide the good at a cost specified by experimental design. This paper provides estimators for moments and quantiles of the unknown distribution in this problem under a variety of nonparametric and semiparametric specifications.

*This research was supported in part by the National Science Foundation through grants SES-9905010 and SBR-9730282, by the E. Morris Cox Endowment, and by the ESRC.

[†]Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Phone: (617) 552-3678. E-mail address: lewbel@bc.edu

[‡]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk

[§]Department of Economics, University of California, Berkeley, CA 94720-3880, USA. E-mail address: mcfadden@econ.berkeley.edu

1 Introduction

A statistical problem that arises in several fields is that of estimating the features of an unknown distribution, which may be conditioned on covariates, using a sample of binomial observations on whether draws from this distribution exceed threshold levels set by experimental design. Consider estimating features of the distribution of some household economic variable W such as wealth, or the willingness to pay (WTP) for a good or resource such as a change in environmental quality. To minimize response bias, each subject i is asked if their W_i exceeds a test value V_i chosen by experimental design.[1] An observation consists of the test value or bid V_i that is posed to subject i, covariates X_i (such as the subject's age, income level, geographic location, or political party affiliation) and a binary indicator Y_i which equals one in the event that W_i exceeds V_i, and zero otherwise, so $Y_i = I(W_i > V_i)$, where $I(\cdot)$ is the indicator function. Objects of interest might include the moments of the distribution of wealth among individuals with certain observable characteristics such as demographics and education level, or the mean, variance and (for median voter models) median willingness-to-pay for a resource among individuals with characteristics like income level, party affiliation, and geographic location, that make them likely voters. Other statistical problems that have the same structure include bioassay[2] and destructive testing.[3]

Many parametric and semiparametric estimators of the distribution of W exist. See, e.g., Kanninen (1993) and Crooker and Herriges (2000) for comparisons of various, mostly parametric, WTP

[1] In many studies, follow up queries are used to gain more information about W, however, we will not consider the use of follow up data, because follow up responses may be shadowed by the framing effect of the first bid. This shadowing effect is common in unfolding bracket survey questions on economic variables, and on stated willingness to pay (WTP) for economic goods. McFadden (1994) provides references and experimental evidence that responses to follow up test values can be biased. There are additional issues of the impact of framing of questions on survey responses, particularly anchoring to test values, including the initial test value; see Green et al. (1998) and Hurd et al. (1998). The data generation process may then be a convolution of the target distribution and a distribution of psychometric errors. This paper will ignore these issues and treat the data generation process as if it is the target distribution. The difficult problem of deconvoluting a target distribution in the presence of psychometric errors is left for future research.

[2] In bioassay the goal is estimation of features of the distribution of survival time W until the onset of an abnormality in laboratory animals exposed to an environmental hazard. The animals are sacrificed at times determined by experimental design, and tested for the abnormality. An observation consists of a vector of covariates X such as attributes of the animal and the exposure, a time V at which the animal is sacrificed for testing, and an indicator Y for whether the test reveals the presence of the abnormality at time V.

[3] An example of destructive testing would be estimation of features of the distribution of speeds W at which car safety device fails. At speeds selected by experimental design, drive cars into a barrier and determine whether a dummy occupant is injured. An observation consists of covariates X (attributes of the car, device, and dummy) a speed V at which the car is tested, and an indicator Y for injury to the test dummy.

estimators. Semiparametric estimators include Chen and Randall (1997), Creel and Loomis (1997), and An (2000). We propose nonparametric consistent estimators for conditional (on covariates) moments of the unknown distribution. We also provide root n consistent estimators for the case where the unknown distribution depends on covariates through a single index location shift, and we provide estimators of conditional quantiles of the unknown distribution.

A common estimation method is to completely parameterize W, e.g., to assume W equals $X^T\theta_0-\varepsilon$ with $\varepsilon \sim N(\alpha, \sigma^2)$. The model then takes the form of a standard probit $Y = I[X^T\theta_0 - V > \varepsilon]$ and can then be estimated using maximum likelihood. However, estimation of the features of the distribution of W differs from ordinary binomial response model estimation in a variety of ways, especially when the model is not fully parameterized.

One difference is that the primary goal of estimation is moments or quantiles of W, rather than the response or choice probabilities of Y. So, for example, in the above parameterized model $E(W|X = x) = X^T\theta_0 - \alpha$, and therefore any semiparametric binomial response model estimator that fails to estimate the location term α, such as Klein and Spady (1993), is insufficient for estimation of WTP. Another important difference is the presence of a covariate V that is determined by experimental design. We exploit this feature of the data in the construction of our estimators.

For identification and consistency we assume the experimental design is randomized with a strictly positive test value density. A drawback of virtually all existing WTP data sets is that they draw bids from discrete distributions. We show in an appendix that, without additional (strong) identifying assumptions, features of the WTP cannot be identified if the bid distribution is discrete.[4] Still, in Monte Carlos and an empirical application we find that the estimators we propose perform reasonably with discrete bid distributions, as long as the number of mass points is not too small.

We consider estimation for a variety of information sets. In the most general case, the distribution of $W|X$ is completely unspecified apart from smoothness, and is nonparametrically estimated. This includes as a special case, and is strictly weaker than, the location model $W = g(X) - \varepsilon$, where ε is independent of X and the function g and the error distribution are unknown. The second case we look at is the semiparametric model $W = g(X, \theta_0) - \varepsilon$ for a known function g, where the parameters θ_0 and the distribution of ε are unknown.

For each of these two cases regarding W (i.e., nonparametric and semiparametric), the design

[4] Alternative identifying assumptions might include homogeneity as in Matzkin's (1992) threshold crossing model, or An's (2000) model which assumes W is an unknown monotonic transformation of $X^T\theta_0 + \varepsilon$ with the distribution of ε known. Chen and Randall (1997) and Creel and Loomis (1997) assume identification in their SNP models, but that may not be reasonable given our nonidentification result and their discrete bid data. See also Manski and Tamer (2002) and Das (2002) for related results, since V can be interpreted as providing (unbounded) interval observations of W.

distribution of the test value V may either be known or unknown to the researcher, which yields a total of four different estimation scenarios. We provide separate estimators for each of these four situations, since each may be relevant depending on the size and content of the data.

2 Model Specification

The goal is estimation of conditional moments or quantiles of a latent, unobserved random scalar W, conditioned on a vector of observed covariates X. The conditional cumulative distribution function of W, denoted $G(w \mid x)$, is unknown but assumed to be smooth.

A test value v is set by a randomized experimental design or natural experiment. The value v is a realization of a random variable V, drawn from either a known or unknown conditional density $h(v \mid x)$ (we consider both cases). It is assumed that W is conditionally independent of V, conditioning on X (consistent with experimental design).

Define Y to equal one in the event that W exceeds V, and zero otherwise, so $Y = I(W > V)$ where $I(\cdot)$ is the indicator function. The observed data consist of a random sample of realizations of covariates X, test values V, and outcomes Y. The framework is similar to random censored regressions (with censoring point V), except that for random censoring we would observe W for observations having $W > V$, whereas in the present context we only observe $Y = I(W > V)$.

Given a function $r(w, x)$, the goal is estimation of the conditional moment $\mu_r(x) = E[r(W, X) \mid X = x]$ for any chosen x in the support of X. Of particular interest are the moments based on $r(W, X) = W^k$ for integers k. In addition to moments we may also be interested in quantiles. Let $w_q(x)$ denote the q'th quantile of W given x.

If the conditional distribution of W given $X = x$ is finitely parameterized, then those parameters can generally be efficiently estimated by maximum likelihood (corresponding to ordinary binary choice model estimation, e.g., logit or probit models), thereby yielding efficient estimates for conditional moments $\mu_r(x)$ and quantiles $w_q(x)$ defined in terms of those parameters.

Assuming that the conditional distribution of W given X is not finitely parameterized, we propose semiparametric and nonparametric estimators for these moments and quantiles. The semiparametric estimators assume that the conditional mean of W is finitely parameterized. The nonparametric estimators only require smoothness assumptions, but suffer from the usual curse of dimensionality. We provide limit normal distributions for these estimators. The semiparametric estimators all converge at the rate that would be obtained if draws w were observed.

The fundamental source of identification is $G(v|x) = E[Y|V = v, X = x]$, i.e., the observable conditional probability mass function of the binary Y provides the conditional distribution function of W, evaluated at v. Conditional moments of W could therefore be recovered from estimates of

$G(v|x)$, e.g., $E(W|X=x) = \int_{\text{supp}(W)} v[dG(v|x)/dv]dv$. We provide more direct estimators below. It follows that nonparametric or semiparametric identification of moments of W requires that the support of V contain the support of W, since moments depend on the function $G(v|x)$ evaluated at every $v \in \text{supp}(W)$.

This illustrates a flaw in the design of most WTP experiments. The typical design draws V from a discrete distribution with a small number of mass points. It follows that G can only be recovered at these points, and hence moments of W cannot be estimated unless the distribution of W is finitely parameterized. So, for example, the semiparametric estimators of WTP proposed by Chen and Randall (1997) and by Creel and Loomis (1997) actually require continuous V, though their applications are discrete.

As required for identification, our estimators assume that V is drawn from a continuous distribution. In our simulation studies, we will examine the size of bias that results when our estimators are applied both with discrete V and continuous V. The next section provides results that will form the basis for the proposed estimators. Later sections provide limiting distributions.

3 Identification

Make the following assumptions.

Assumption A.1. The covariate vector X has compact support $\mathcal{X} \subseteq \mathbb{R}^d$. The latent scalar W has an unknown, twice continuously differentiable, strictly monotonic, conditional c.d.f. $G(w \mid x)$ and probability density function $g(w|x)$, with a compact support $[\alpha_0(x), \alpha_1(x)]$. The test variable V is continuously distributed with a known or unknown positive probability density function $h(v \mid x)$ having compact support $[\delta_0(x), \delta_1(x)]$ such that $\delta_0(x) \leq \alpha_0(x)$ and $\delta_1(x) \geq \alpha_1(x)$. The variables W and V are conditionally independent, given X. Let $Z = (X, V, Y)$.

Define $m(v, x)$ by

$$m(v, x) = E[Y|V = v, X = x],$$

and let m^{-1} be the inverse of the function m with respect to its first element (which exists on the support of W given assumption A.1), so if $t = m(v, x)$ then $v = m^{-1}(t \mid x)$ for $v \in [\alpha_0(x), \alpha_1(x)]$.

Assumption A.2. The function $r(w, x)$, chosen by the researcher, is regular, meaning that it is continuous in (w, x) for all w and x on their supports, and for each x is twice continuously differentiable in w. Let κ be a known constant that is in the support of W. The moment $\mu_r(x)$ exists, where $\mu_r(x)$ is defined by

$$\mu_r(x) = E[r(W, X) \mid X = x].$$

Define $r'(w,x) = \partial r(w,x)/\partial w$ and $s_r(z)$ by

$$s_r(z) = r(\kappa, x) + \frac{r'(v,x)[y - 1(v < \kappa)]}{h(v \mid x)}.$$

For any regular function r, Theorem 1 below provides an expression for the conditional mean $\mu_r(x)$. Also provided is the q'th conditional quantile of W given $X = x$, denoted $w_q(x)$.

THEOREM 1. *Let Assumptions A.1 and A.2 hold. Then*

$$\begin{aligned} \mu_r(x) &= E[s_r(Z) \mid X = x] \\ w_q(x) &= m^{-1}(1 - q \mid x). \end{aligned}$$

PROOF OF THEOREM 1. First observe that, given the conditional independence of W and V,

$$m(v,x) = E[Y|V = v, X = x] = 1 - G(v|x).$$

Next, by definition, $\mu_r(x) = \int_{\alpha_0(x)}^{\alpha_1(x)} r(v,x)g(v|x)dv$. Integration by parts yields

$$\mu_r(x) = r(v,x)[G(v|x) - 1(v \geq \kappa)]|_{v=\alpha_0(x)}^{\alpha_1(x)} - \int_{\alpha_0(x)}^{\alpha_1(x)} r'(v,x)[G(v|x) - 1(v \geq \kappa)]dv.$$

Therefore, collecting terms we find that

$$\begin{aligned} \mu_r(x) &= r(\kappa, x) + \int_{\alpha_0(x)}^{\alpha_1(x)} r'(v,x)[E(Y|V = v, X = x) - 1(v < \kappa)]dv && (1) \\ &= \int_{\alpha_0(x)}^{\alpha_1(x)} E[s_r(Z)|V = v, X = x)h(v \mid x)dv \\ &= E[s_r(Z) \mid X = x], && (2) \end{aligned}$$

where the last equality uses the assumptions regarding the supports of W and V, and the law of iterated expectation. The conditional quantile expression follows from $G(v|x) = 1 - m(v,x)$. ∎

Theorem 1 provides the basis for the nonparametric moment estimators described in the next section, and for some semiparametric and quantile estimators. Essentially, based on Theorem 1, $\mu_r(x)$ may be estimated as the fitted values of either a parametric or nonparametric regression of $s_r(Z_i)$ on

X_i. Note that if the nonparametric location model $W = g(X) - \varepsilon$ holds, then $g(x) = \mu_r(x) + E(\varepsilon)$ with $r(w, x) = w$.[5]

Theorem 1 shows nonparametric identification of moments assuming the distribution of V is continuous. Theorem 5 in the appendix shows that this continuity assumption (or some other alternative strong restriction on the model) is necessary, in that identification fails when V is discrete.

Corollary 1 below will be used to obtain faster converging moment and quantile estimators, based on stronger assumptions.

ASSUMPTION A.3. The latent W satisfies $W = g(X, \theta_0) - \varepsilon$, where g is a known function, $\theta_0 \in \Theta$ is a vector of parameters, and ε is a disturbance that is distributed independently of V, X, with unknown, twice continuously differentiable c.d.f. $G^*(\varepsilon)$ and compact support $[a_0, a_1]$ that contains zero. Define $U = g(X, \theta_0) - V$. Let $\psi(U)$ denote the unconditional probability density function of U. The support of U contains the interval $[a_0, a_1]$.

Define $s_r^*(x, u, y)$ by

$$s_r^*(x, u, y) = r[g(x, \theta_0), x] + \frac{r'[g(x, \theta_0) - u, x][y - 1(u > 0)]}{\psi(u)}.$$

COROLLARY 1. *Let Assumptions A.1, A.2, and A.3 hold. Then*

$$\begin{aligned}
\psi(u) &= E[h_{V|X}(g(X, \theta_0) - u)] \\
G^*(u) &= E(Y \mid U = u) \\
\mu_r(x) &= E[s_r^*(x, U, Y)] \\
w_q(x) &= g(x, \theta_0) - G^{*-1}(1 - q).
\end{aligned}$$

PROOF OF COROLLARY 1. Having $\psi(u) = E[h_{V|X}(g(X, \theta_0) - u)]$ follows from the definitions of U, ψ, and $h_{V|X}$ and the law of iterated expectation. Also from definitions, $Y = I(\varepsilon < U)$ which implies that $G^*(u) = E(Y \mid U = u)$. Next, following the same steps as in Theorem 1 we have

$$\begin{aligned}
\mu_r(x) &= \int_{a_0}^{a_1} r[g(x, \theta_0) - u, x][\partial G^*(u)/\partial u]du \\
&= r[g(x, \theta_0), x] + \int_{a_0}^{a_1} r'[g(x, \theta_0) - u, x][G^*(u) - 1(u > 0)]du \\
&= \int_{a_0}^{a_1} E[s_r^*(x, U, Y) \mid U = u]\psi(u)du = E[s_r^*(x, U, Y)].
\end{aligned}$$

[5] In that case, if you take $r(w, x) = w^2$, then $\mu_r(x) = g^2(x) - 2g(x)E(\varepsilon) + \sigma_\varepsilon^2$ for some constant σ_ε^2. In this special case of a location model, many functions r provide information about g.

Finally, the quantile expression follows from $G(w \mid X = x) = 1 - G^*[g(x, \theta_0) - w]$. ■

The advantage of Corollary 1 over Theorem 1 for estimation is that in Corollary 1, $\mu_r(x)$ and $\psi(u)$ are expressed as unconditional expectations and so can be estimated using ordinary sample averages (given an estimate of θ). Similarly, using Corollary 1 estimation of the quantiles $w_q(x)$ given θ only requires estimation of the one dimensional regression $G^*(u) = E(Y \mid U = u)$, instead of the high dimensional $m(v, x)$.

4 Estimators

We suppose that a random sample $Z_i = (X_i, V_i, Y_i)$ for $i = 1, \ldots, n$ is observed, where V_i is a realization of V, Y_i is a realization of Y, and X_i is a realization of X. Using this data, based on Theorem 1 and Corollary 1 we provide five different estimators for $\mu_r(x)$, denoted $\widehat{\mu}_{r\lambda}(x)$ for $\lambda = 1, 2, 3, 4, 5$.

The estimator $\widehat{\mu}_{r1}(x)$ is for nonparametric estimation when the experimental design density $h(v \mid x)$ is known, and $\widehat{\mu}_{r2}(x)$ is for nonparametric estimation when $h(v \mid x)$ is unknown. Similarly, $\widehat{\mu}_{r3}(x)$ and $\widehat{\mu}_{r4}(x)$ cover the cases of semiparametric estimation with $h(v \mid x)$ known and unknown, respectively. An additional semiparametric estimator $\widehat{\mu}_{r5}(x)$ is provided that is simpler than $\widehat{\mu}_{r3}$ or $\widehat{\mu}_{r4}$, but may only be used for certain simple choices of r.

To describe these estimators, let $\widehat{E}$ denote an estimated expectation. An unconditional estimated expectation just denotes the sample average, while a conditional estimated expectation denotes a nonparametric regression.

4.1 Nonparametric Estimators

Suppose that Assumptions A.1 and A.2 hold, but A.3 does not hold, so moments of W must be nonparametrically estimated. If the experimental design, and hence the density function h, is known, then $s_r(Z_i)$ can be constructed for each observation i, and $\mu_r(x)$ may then be consistently estimating by nonparametrically regressing $s_r(Z_i)$ on X_i. This first estimator is

$$\widehat{\mu}_{1r}(x) = \widehat{E}[s_r(Z) \mid X = x],$$

where in this case $\widehat{E}$ is any regression smoother [see below].

Note that $\widehat{\mu}_{1r}(x)$ depends on the design density h.[6] An estimator that does not entail knowing

[6] If h is unknown, then based on $\widehat{\mu}_{1r}$ an estimator of $\mu_r(x)$ could be constructed by first estimating h. Specifically, one could replace $h(v \mid x)$ with an estimate $\widehat{h}(v \mid x)$ (using, e.g., kernel density estimation) in the definition of $s_r(z)$. Call the result $\widehat{s}_r(z)$. The estimator of $\mu_r(x)$ would then be $\widehat{\mu}^*_{1r}(x) = \widehat{E}[\widehat{s}_r(Z) \mid X = x]$

or estimating the density h is the following. Recall that $m(v,x) = E[Y|V=v, X=x]$. Let $\widehat{m}(v,x)$ be a consistent estimator of m, that is, a nonparametric regression of y on x, v, so

$$\widehat{m}(v,x) = \widehat{E}[Y|V=v, X=x].$$

Let a_0 and a_1 be known or estimated constants such that $a_0 \leq \alpha_0(x)$ and $a_1 \geq \alpha_1(x)$. Then, based on the proof of Theorem 1, a consistent estimator of $\mu_r(x)$ is given by

$$\widehat{\mu}_{2r}(x) = r(\kappa, x) + \int_{a_0}^{a_1} r'(v,x)[\widehat{m}(v,x) - 1(v < \kappa)]dv.$$

We give some more details later about the construction of these estimators.

4.2 Semiparametric Estimators

This section discusses rate root n estimation of arbitrary conditional moments based on Corollary 1. For these estimators we let Assumption A.3 hold, in addition to Assumptions A.1 and A.2, i.e., the specification is semiparametric. It will be convenient to first consider the case where θ_0 in Assumption 3 is known, implying that the conditional mean of W is known up to an arbitrary location (since ε is not required to have mean zero). A special case of known θ_0 is when x is empty, i.e., estimation of unconditional moments of W, since in that case we can without loss of generality take g to equal zero.

4.2.1 Estimation With Known θ

Suppose that θ_0 is known. Considering first the case where the design density h is also known, for a given u define the sample average $\widehat{\psi}(u)$ by

$$\widehat{\psi}(u) = \frac{1}{n}\sum_{i=1}^{n} h_{V_i|X_i}[g(X_i,\theta_0) - u].$$

Then, based on Corollary 1, we have the estimator

$$\widehat{\mu}^*_{3r}(x) = r[g(x,\theta_0), x] + \frac{1}{n}\sum_{i=1}^{n} \frac{r'[g(x,\theta_0) - U_i, x][Y_i - 1(U_i > 0)]}{\widehat{\psi}(U_i)}.$$

This estimator is computationally extremely simple, since it entails only sample averages. Special cases of the estimator $\widehat{\mu}_{3r}(x)$ were proposed by McFadden (1994) and by Lewbel (1997).

Let $\widetilde{\psi}(u)$ be an estimator of $\psi(u)$ that does not depend on knowledge of h. For example $\widetilde{\psi}(u)$ could be a (one dimensional) kernel density estimator of the density of U, based on the data $\widehat{U}_i$ and evaluated at u. We then have the estimator

$$\widehat{\mu}^*_{4r}(x) = r[g(x,\theta_0), x] + \frac{1}{n}\sum_{i=1}^{n} \frac{r'[g(x,\theta_0) - U_i, x][Y_i - 1(U_i > 0)]}{\widetilde{\psi}(U_i)},$$

which may be used when h is unknown.

4.2.2 Estimation with Unknown θ

First consider estimation of θ. Assumptions A.1, A.2, and A.3 imply that

$$E(W \mid X = x) = \alpha + g(x, \theta)$$

for some arbitrary location constant α (this constant is unknown since no location constraint is imposed upon ε). It follows from Theorem 1 that $E\left[s_w(Z) \mid X = x\right] = E(W \mid X = x)$, where $s_w(Z)$ denotes $s_r(Z)$ with $r(w, x) = w$. Therefore, we may obtain a simple root n consistent, asymptotically normal estimate $\widetilde{\theta}$ of θ by minimizing the least squares criterion as follows

$$(\widehat{\theta}, \widehat{\alpha}) = \arg\min_{\theta,\alpha} \frac{1}{n} \sum_{i=1}^{n} \left(s_w(Z_i) - \alpha - g(X_i, \theta)\right)^2. \tag{3}$$

If g is linear in parameters, then a closed form expression is possible for both parameter estimates. An estimator $\widehat{h}$ may be used in place of h inside $s_w(Z_i)$ if h is unknown. The estimator $(\widehat{\theta}, \widehat{\alpha})$ is an ordinary nonlinear weighted least squares, and so is root n consistent and asymptotically normal with a standard limiting distribution, under standard regularity conditions. If h is not known, one could replace $h(V \mid X)$ with an estimate $\widehat{h}(V \mid X)$ in the definition of $\widehat{\beta}$. The resulting estimator would then take the form of an ordinary two step estimator with a nonparametric first step (the estimation of h) which, with regularity, will be root n consistent and asymptotically normal. This estimator of θ and α is equivalent to the estimator for general binary choice models proposed by Lewbel (2000), though Lewbel provides other extensions, such as to estimation with endogenous regressors.

Assumptions A.1, A.2, and A.3 make the latent error ε independent of X, and therefore the binary choice estimator of Klein and Spady (1993) will provide a semiparametrically efficient estimator of θ (note that the Klein and Spady estimator does not identify a location constant α, but that is not required, since no location constraint is imposed upon ε).

Let $\widehat{\theta}$ denote the chosen root N consistent, asymptotically normal estimator for θ_0. Replacing θ_0 with any $\theta \in \Theta$ we may rewrite the estimators of the previous section as $\widehat{\mu}^*_{\lambda r}(x; \theta)$ for $\lambda =$ 3, or 4. Note that in addition to directly appearing in the equations for $\widehat{\mu}^*_{\lambda r}$, θ also appears in the definition of $U_i = g(X_i, \theta) - V_i$. We later derive the root N consistent, asymptotically normal limiting distribution for each estimator $\widehat{\mu}_{\lambda r}(x) = \widehat{\mu}^*_{\lambda r}(x; \widehat{\theta})$, where we suppress the dependence on $\widehat{\theta}$ for simplicity. The estimators are not differentiable in U_i, which complicates the derivation of their limiting distribution, e.g., Theorem 6.1 of Newey and McFadden (1994) is not directly applicable due to this nondifferentiability.

4.2.3 A Special Case

First suppose that $r(w, x) = w^k$, which is the most common choice of function r in applications. Note that for any k,

$$E[W^k|X = x] = E\left[(g(X, \theta_0) - \varepsilon)^k \,|X = x\right] = \sum_{\ell=0}^{k} g(x, \theta_0)^\ell (-1)^{k-\ell} \binom{k}{\ell} E(\varepsilon^{k-\ell})$$

by the binomial expansion. Therefore, we can write

$$E[W^k|X = x] = \sum_{\ell=0}^{k} g(x, \theta_0)^\ell \xi_\ell,$$

where ξ_ℓ, $\ell = 0, \ldots, k$ are unknown parameters depending on the moments of the error distribution and on the binomial coefficients. We can estimate the nuisance parameters ξ_ℓ by solving the least squares problem

$$(\widehat{\xi}_0, \ldots, \widehat{\xi}_k) = \arg \min_{\xi_0, \ldots, \xi_k} \frac{1}{n} \sum_{i=1}^{n} \left(s_{W^k}(Z_i) - \sum_{\ell=0}^{k} g(X_i, \widehat{\theta})^\ell \xi_\ell \right)^2,$$

where $\widehat{\theta}$ is any root-N consistent estimator such as defined in (3). Then let

$$\widehat{\mu}_{5w^k}(x) = \sum_{\ell=0}^{k} g(X_i, \widehat{\theta})^\ell \widehat{\xi}_\ell.$$

to estimate $\mu_{w^k}(x)$.[7]

This estimator should work well when k is small, but otherwise a large number of auxiliary parameters have to be estimated and this may increase the variance of the estimate of $\mu_r(x)$ considerably. It is also sensitive to the existence of moments.

4.3 Quantile Estimators

In addition to moments, one may also desire estimates of conditional quantiles of W. Let Assumptions A.1 and A.2 hold and define $\widehat{G}(v \mid x) = 1 - \widehat{m}(v, x)$. Then, based on Theorem 1, an estimate of the q'th quantile of W given $X = x$ is just

$$\widehat{w}_q(x) = \widehat{G}^{-1}(q \mid x).$$

[7] This method can be extended to a more general class of r functions than just polynomials. Suppose $r(w, x) = \sum_{j=1}^{\infty} \psi_j(x) w^j$ for some known coefficients $\{\psi_j(x)\}_{j=1}^{\infty}$. This is true for a large class of r functions of interest like the exponential and logarithm. Then in practice, we approximate $r(w, x)$ by $\sum_{j=1}^{\tau} \psi_j(x) w^j$, where $\tau = \tau(n)$ is some truncation parameter, and then apply our method for estimating $\mu_{w^k}(x)$.

The rate of convergence of this estimate will be slow, because of the high dimension of $\widehat{m}$.

If Assumption A.3 holds in addition to A.1 and A.2, then faster convergence is possible. Given Corollary 1 we have $U = g(X, \theta_0) - V$, $G^*(u) = E(Y \mid U = u)$, and $w_q(x) = g(x, \theta_0) - G^{*-1}(1-q)$. Therefore, let $\widehat{U}_i = g(X_i, \widehat{\theta}) - V_i$ and estimate the conditional quantile $w_q(x)$ by

$$\begin{aligned} \widehat{G}^*(u) &= \widehat{E}(Y \mid \widehat{U} = u) \\ \widetilde{w}_q(x) &= g(x, \widehat{\theta}) - \widehat{G}^{*-1}(1-q), \end{aligned}$$

where the function $\widehat{G}^*$ is obtained by nonparametrically regressing Y on $\widehat{U}$, and is then numerically inverted to obtain $\widehat{G}^{*-1}$. This estimator $\widetilde{w}_q(x)$ will converge at a faster rate than the nonparametric quantile estimator $\widehat{w}_q(x)$. With sufficient regularity, $\widetilde{w}_q(x)$ is asymptotically normal and converges at the same rate as a one dimensional nonparametric regression estimator, i.e., the same as the best rate that could be obtained if realizations of the latent W were observed.

4.4 Further Comments

The basic location model can be generalized in a number of ways that are popular in applications. For example, we may prefer the specification for $\log W$ instead of W itself. In this case, one can just replace V_i by $\log V_i$ and κ by $\log \kappa$ and compute everything as before except that now the function $r(w, x)$ is replaced by $r(\exp(w), x)$. More generally, this works with any known monotonic transformation τ of W. In the specific case where $E(W|X = x)$ is of interest, one estimates the parameters θ, α in the transformed scale by (3) and then computes the inverse transformed moment.

Estimating the fully nonparametric model may be unattractive when there are many covariates, and we may then seek some compromise model that reduces the dimensionality, like index models, partial linear, or additive models.

5 Estimation Details and Distribution Theory

In this section we provide more detail about the computation of the estimators $\widehat{\mu}_{jr}(x)$, $j = 1, 2, 3, 4$ and their distribution theory. The theory for $\widehat{\mu}_{5w}(x)$ follows from standard results on nonlinear regression, and we do not have anything to say about it, so do not present it here.

5.1 Nonparametric Estimators

We will base our estimation of conditional expectations on the local linear kernel method. This method has been extensively analyzed and has some attractive properties like being design adaptive, and best linear minimax; see Fan and Gijbels (1996) for further discussion. In the estimator $\widehat{\mu}_{1r}(x)$, we smooth $s_r(Z_i)$ against X_i so that we let $(\widehat{\delta}_0, \widehat{\delta}_x)$ minimize the following localized least squares criterion

$$\sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right)\left[s_r(Z_i) - \delta_0 - \delta_x'(X_i - x)\right]^2,$$

where $K(t) = \prod_{j=1}^{d} k(t_j)$ and k is a univariate kernel function, while $b = b(n)$ is a bandwidth. Then let

$$\widehat{\mu}_{1r}(x) = \widehat{\delta}_0. \tag{4}$$

This estimator is linear in the dependent variable, i.e., is of the form $\sum_{i=1}^{n} w_{ni}(x) s_r(Z_i)$, where the smoothing weights $w_{ni}(x)$ depend only on $\{X_1, \ldots, X_n\}$. For simplicity we shall suppose that all X variables are continuous, and have more or less the same scale - in practice it is desirable to rescale the covariates so they have a common scale.

In computing the estimator $\widehat{\mu}_{2r}(x)$ we require an estimator of $m(v, x)$, which is given by the local linear smooth of Y_i on X_i, V_i. Specifically, let $(\widehat{\delta}_0, \widehat{\delta}_v, \widehat{\delta}_x)$ minimize the following localized least squares criterion

$$\sum_{i=1}^{n} k\left(\frac{v - V_i}{b}\right) K\left(\frac{x - X_i}{b}\right)\left[Y_i - \delta_0 - \delta_v(V_i - v) - \delta_x'(X_i - x)\right]^2,$$

and let $\widehat{m}(v, x) = \widehat{\delta}_0 = \sum_{i=1}^{n} w_{ni}(v, x) Y_i$. Then $w_{ni}(v, x)$ are smoothing weights that depend only on $\{(X_1, V_1), \ldots, (X_n, V_n)\}$. Then let

$$\widehat{\mu}_{2r}(x) = r(\kappa, x) + \int_{\alpha_0(x)}^{\alpha_1(x)} r'(v, x)[\widehat{m}(v, x) - 1(v < \kappa)]dv, \tag{5}$$

where the univariate integral is computed numerically, say by Gaussian Quadrature. This estimator is in the class of marginal integration/partial mean estimators sometimes used for estimating additive nonparametric regression models, see Linton and Nielsen (1995), Newey (1994), and Tjøstheim and Auestad (1994), except that the integrating measure λ, where $d\lambda(v) = r'(v, x) 1(\alpha_0(x) \leq v \leq \alpha_1(x))dv$, is not necessarily a probability measure, i.e., it may not be positive or integrate to one. The distribution theory for the class of marginal integration estimators has already been worked out for a number of specific smoothing methods, see the above references.

Define the error term $\varepsilon_i = Y_i - m(V_i, X_i)$ that is independent across i and satisfies $E(\varepsilon_i | V_i, X_i) = 0$; define also $\sigma^2(v, x) = \mathrm{var}(\varepsilon_i | V_i = v, X_i = x) = G(v|x)[1 - G(v|x)]$. Define $\mu_2(k) = \int t^2 k(t) dt$. Let ∇, ∇^2 denote the first and second derivative operators.

THEOREM 2. *Suppose that assumptions B1 and B2 in the appendix hold and that the bandwidth sequence $b = b(n)$ satisfies $b \to 0$ and $nb^{d+2}/\log n \to \infty$. Then, for $j = 1, 2$,*

$$\frac{\widehat{\mu}_{jr}(x) - \mu_r(x) - b^2 \beta_j(x)}{\sqrt{\omega_j(x)/nb^d}} \xrightarrow{d} N(0, 1),$$

where

$$\beta_1(x) = \frac{\mu_2(k)}{2}\mathrm{tr}(\nabla^2\mu_r(x)) \text{ and } \beta_2(x) = \frac{\mu_2(k)}{2}\int \mathrm{tr}(\nabla^2 m(v,x))d\lambda(v),$$

$$\begin{aligned}\omega_1(x) &= \|K\|^2\frac{\mathrm{var}[s_r(Z)|X=x]}{f_X(x)} \\ \omega_2(x) &= \|K\|^2\int_{\alpha_0(x)}^{\alpha_1(x)}\sigma^2(v,x)\left(\frac{r'(v,x)}{f_{V,X}(v,x)}\right)^2 f_{V,X}(v,x)dv.\end{aligned}$$

The two estimators are not in general rankable in terms of mean squared error, but can be compared in some special cases, which we do next.

Suppose that $r(w,x)=w$. In this case

$$\omega_1(x) \propto \mathrm{var}\left(\frac{[Y-1(V<\kappa)]}{h(V\mid X)}|X=x\right) \text{ and } \omega_2(x)\propto\int_{\alpha_0(x)}^{\alpha_1(x)}\left(\frac{G(v|x)[1-G(v|x)]}{f_{V|X}(v|x)}\right)dv.$$

If furthermore, $V|X$ is uniform on $[0,1]$,

$$\omega_1(x)\propto\int_0^1 G(v|x)dv[1-\int_0^1 G(v|x)dv]+\kappa(1-\kappa)+2\left[\int_0^\kappa G(v|x)dv-\kappa\int_0^1 G(v|x)dv\right]$$

$$\omega_2(x)\propto\int_0^1 G(v|x)[1-G(v|x)]dv.$$

Generally, $\omega_1(x)$ depends on κ except in the special case that W is uniform on $[0,1]$. If $W|X$ is uniform on $[0,1]$, the asymptotic variance of $\widehat{\mu}_{2r}(x)$ is proportional to $1/6$, while the asymptotic variance of $\widehat{\mu}_{1r}(x)$ is proportional to $1/4$ [taking out the factor to do with the bandwidth, kernel, and covariate density]. In this case, $\widehat{\mu}_{2r}(x)$ is more efficient in variance terms.[8]

[8] The estimators for different κ are correlated, but not perfectly so, so that there is scope for improving efficiency of $\widehat{\mu}_{1r}(x)$. The covariance function of $\widehat{\mu}_{1r}(x;\kappa_1),\widehat{\mu}_{1r}(x;\kappa_2)$ is proportional to

$$C(\kappa_1,\kappa_2)=\frac{1}{4}+\kappa_1\wedge\kappa_2-\kappa_1\kappa_2-\left[\frac{\kappa_1}{2}-\int_0^{\kappa_1}G(v)dv\right]-\left[\frac{\kappa_2}{2}-\int_0^{\kappa_2}G(v)dv\right],$$

which is equal to

$$\frac{1}{4}+\kappa_1\wedge\kappa_2-\kappa_1\kappa_2-\frac{\kappa_1}{2}+\frac{\kappa_1^2}{2}-\frac{\kappa_2}{2}+\frac{\kappa_2^2}{2}$$

in the uniform case. The correlation function is maximized at $\kappa_1=\kappa_2$ whereupon it is one, but is minimized at the point where $\kappa_1=\kappa_2\pm 1/2$ and the minimizing value is 0.5. Furthermore, we can establish a functional central limit theorem in κ. Now consider the class of estimators $\int\widehat{\mu}_{1r}(x;\kappa)\omega(\kappa)d\kappa$ for some weighting function $\omega(.)$. One can show that with the optimal combination we achieve the same variance factor, $1/6$ in the uniform case, as the estimator $\widehat{\mu}_{2r}(x)$.

Regarding the bias of the two estimators in the case that $r(w, x) = w$:

$$\begin{aligned} \operatorname{tr}(\nabla_x^2 \mu_r(x)) &= \sum_{j=1}^{d} \int v \frac{\partial^2 g(v|x)}{\partial x_j^2} dv = \sum_{j=1}^{d} \int \operatorname{tr}(\nabla_x^2 m(v,x)) dv \\ \int \operatorname{tr}(\nabla^2 m(v,x)) d\lambda(v) &= -\int \left[\sum_{j=1}^{d} \frac{\partial^2 G(v|x)}{\partial x_j^2} + \frac{\partial^2 G(v|x)}{\partial v^2} \right] dv, \end{aligned}$$

where $g(v|x)$ is the conditional density of $W|X$. Under certain conditions these two biases are the same applying integration by parts. In order for $\int [\partial^2 G(v|x)/\partial v^2] dv = \int [\partial g(v|x)/\partial v] dv = 0$ it should be the case that the conditional density and its derivative is zero on the boundary.

In conclusion, we have shown than it in a fairly important special case $\widehat{\mu}_{2r}(x)$ has smaller mean squared error than $\widehat{\mu}_{1r}(x)$, although this result does not hold more generally. There are some other comparisons between the estimators that are relevant. For example, estimator $\widehat{\mu}_{1r}(x)$ requires more smoothness than $\widehat{\mu}_{2r}(x)$, as can be seen from the bias expressions given above. On the other hand $\widehat{\mu}_{1r}(x)$ requires a lower dimensional smoothing operation than $\widehat{\mu}_{2r}(x)$, which may be important in small samples. An advantage of the estimator $\widehat{\mu}_{1r}(x)$ is that since it is a standard nonparametric regression estimator, all of the sophisticated regression bandwidth selection methods can be automatically applied, whereas the comparable theory relevant for $\widehat{\mu}_{2r}(x)$ is not so well developed. Similar comments apply to standard error construction based on asymptotic or bootstrap principles.

5.2 Semiparametric Estimators

Here we state the asymptotic properties of the conditional moment estimators based on Corollary 1. Let $\widehat{\theta}$ be some consistent estimator of θ_0. Define:

$$\widehat{\mu}_{3r}(x) = r[g(x,\widehat{\theta}), x] + \frac{1}{n} \sum_{i=1}^{n} \frac{r'[g(x,\widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\widehat{\psi}(\widehat{U}_i)}$$

$$\widehat{\mu}_{4r}(x) = r[g(x,\widehat{\theta}), x] + \frac{1}{n} \sum_{i=1}^{n} \frac{r'[g(x,\widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\widetilde{\psi}(\widehat{U}_i)},$$

where $\widehat{U}_i = g(X_i, \widehat{\theta}) - V_i$ and

$$\widehat{\psi}(\widehat{U}_i) = \frac{1}{n} \sum_{j=1}^{n} h[g(X_j,\widehat{\theta}) - \widehat{U}_i] \quad ; \quad \widetilde{\psi}(\widehat{U}_i) = \frac{1}{nb} \sum_{j=1}^{n} k\left(\frac{\widehat{U}_i - \widehat{U}_j}{b} \right).$$

We shall suppose that

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(Z_i, \theta_0) + o_p(1) \tag{6}$$

for some function m that is mean zero and has finite variance. The estimators $\widehat{\mu}_{3r}^*(x)$ and $\widehat{\mu}_{4r}^*(x)$ are the special cases of $\widehat{\mu}_{3r}(x)$ and $\widehat{\mu}_{4r}(x)$ in which θ is known, and so correspond to the case of m being identically zero.

For each $\theta \in \Theta$ and $x \in \mathcal{X}$, define:

$$\begin{aligned}
f_0(Z_i,\theta) &= \frac{r'[g(x,\theta)-U_i(\theta),x][Y_i - 1(U_i(\theta)>0)]}{\psi(U_i)} \\
f_1(Z_i,\theta) &= r[g(x,\theta),x] + \frac{r'[g(x,\theta)-U_i(\theta),x][Y_i - 1(U_i(\theta)>0)]}{\psi(U_i)} \\
\Gamma_F &= \left(\frac{\partial}{\partial\theta} E\left[f_1(Z_i,\theta)\right]\right)_{\theta=\theta_0},
\end{aligned}$$

where $U_i(\theta) = g(X_i,\theta) - V_i$. The quantities f_0, f_1 and Γ_F depend on x but we have suppressed this notationally. Note also that $Ef_1(Z_i,\theta_0) = \mu_r(x)$. Finally, let

$$\gamma_i = \frac{\partial g}{\partial \theta}(X_i,\theta_0) - E\left[\frac{\partial g}{\partial \theta}(X_i,\theta_0)\right].$$

THEOREM 3. *Suppose that Assumptions C1-C3 in the Appendix hold. Then, as $n \to \infty$,*

$$\frac{\sqrt{n}[\widehat{\mu}_{3r}(x) - \mu_r(x)]}{\sigma_\eta(x)} \xrightarrow{d} N(0,1), \tag{7}$$

where $0 < \sigma_\eta^2(x) = \operatorname{var}(\eta_j) < \infty$ with $\eta_j = \eta_{1j} + \eta_{2j} + \eta_{3j}$, where:

$$\begin{aligned}
\eta_{1j} &= f_1(Z_j,\theta_0) - Ef_1(Z_j,\theta_0) \\
\eta_{2j} &= \left(\Gamma_F - E\left[f_0(Z_i,\theta_0)\frac{\psi'(U_i)}{\psi(U_i)}\gamma_i\right]\right) m(Z_j;\theta_0) \\
\eta_{3j} &= -E\left[f_0(Z_i,\theta_0)\frac{h[g(X_j,\theta_0)-U_i] - \psi(U_i)}{\psi(U_i)}|X_j\right].
\end{aligned}$$

The three terms η_{1j}, η_{2j}, and η_{3j} are all mean zero and have finite variance. They are generally mutually correlated. When θ_0 is known, the term $\eta_{2j} = 0$ and this term is missing from the asymptotic expansion.

We next give the distribution theory for the semiparametric estimator $\widehat{\mu}_{4r}(x)$. Let

$$\gamma_i^* = \frac{\partial g}{\partial \theta'}(X_i,\theta_0) - E\left[\frac{\partial g}{\partial \theta'}(X_i,\theta_0)|U_i\right],$$

and $\phi_u(U_i) = E[\partial g(X_i,\theta_0)/\partial\theta | U_i]$.

THEOREM 4. *Suppose that assumptions B1, B2 and C1-C4 in the Appendix hold. Then*

$$\frac{\sqrt{n}[\widehat{\mu}_{4r}(x) - \mu_r(x)]}{\sigma_\eta^*(x)} \xrightarrow{d} N(0,1),$$

where $0 < \sigma_\eta^{*2}(x) = \mathrm{var}(\eta_j^*) < \infty$, *with:* $\eta_j^* = \eta_{1j}^* + \eta_{2j}^* + \eta_{3j}^*$, *where* $\eta_{1j}^* = \eta_{1j}$, *while*

$$\eta_{2j}^* = \left(\Gamma_F - E\left[f_0(Z_i,\theta_0)\left(\frac{\psi'(U_i)}{\psi(U_i)}\gamma_i^* - \phi_u'(U_i)\right)\right]\right) m(Z_j;\theta_0)$$

$$\eta_{3j}^* = -\frac{r'[g(x,\theta_0) - U_j, x]}{\psi(U_j)}\left(Y_j - E[Y_j|U_j]\right).$$

The three terms η_{1j}^*, η_{2j}^*, and η_{3j}^* are all mean zero and have finite variance. They are generally correlated. When θ_0 is known, the term $\eta_{2j}^* = 0$ and this term is missing from the asymptotic expansion.

Standard errors can be constructed by substituting population quantities by estimated ones along the lines discussed in Newey and McFadden (1994). Alternatively, and perhaps preferably, one can use the bootstrap as we do in our application below.

Regarding efficiency, it is not possible to provide a ranking of the two estimators $\widehat{\mu}_{3r}(x)$ and $\widehat{\mu}_{4r}(x)$ uniformly throughout the 'parameter space'. This partly depends on the choice of $\widehat{\theta}$. It may be possible to develop an efficiency bound for estimation of the function $\mu_r(.)$ by following the calculations of Bickel, Klaassen, Ritov and Wellner (1993, Chapter 5). Since there are no additional restrictions on μ_r, the plug-in estimator with efficient $\widehat{\theta}$ should be efficient.

5.2.1 Quantile estimators

The distribution theory is trivial. The estimator $\widehat{w}_q(x) = \widehat{G}^{-1}(q \mid x)$ has the distribution theory for standard conditional quantile estimators. The distribution theory for $\widetilde{w}_q(x) = g(x,\widehat{\theta}) - \widehat{G}^{*-1}(1-q)$ is the same as the distribution theory for $\widetilde{w}_q(x) = g(x,\theta_0) - \widetilde{G}^{*-1}(1-q)$, where

$$\widetilde{G}^*(u) = \widehat{E}(Y \mid U = u),$$

which is again basically a standard one-dimensional conditional quantile estimator. This is because $\widehat{\theta}$ converges at rate root-n, so the estimation error in $\widehat{\theta}$ is asymptotically irrelevant given the slower convergence rate of quantiles.

6 Numerical Results

6.1 Monte Carlo

We report the results of a small simulation experiment based on a design of Crooker and Herriges (2000). Let

$$W_i = \beta_1 + \beta_2 X_i + \sigma\varepsilon_i,$$

where X_i is uniformly distributed on $[-30, 30]$ and ε_i is standard normal. We take $\beta_1 = 100$ and $\beta_2 = 2$, which guarantees that the mean WTP is equal to 100. We vary the value of $\sigma \in \{5, 10, 50\}$ and sample size $n \in \{100, 300, 500\}$. The bid values are chosen equally randomly from $\{25, 50, 75, 125, 175\}$, and $\kappa = 100$. This design violates our assumptions on two grounds: first, the bid distribution is continuous (though later we also consider a continuous bid design); and second, the range of the bid values, at least in the large σ case, does not include the full support of W. This design was chosen because it permits direct comparison with the parametric and SNP estimators of WTP considered by Crooker and Herriges (2000), and to check robustness of the estimators against violations of assumptions that may occur in applications.

The moments we estimate are $E[W|X = x]$ and $\text{std}(W|X = x) = \sqrt{E[W^2|X = x] - E^2[W|X = x]}$. We compute estimators $\widehat{\mu}_\lambda(.)$ for $\lambda = 1, 3, 4, 5$. We do not compute $\widehat{\mu}_2(.)$ here, because it is very time consuming, and the small sample performance of this class of estimators (integrals of nonparametric conditional expectations) has been extensively documented in Sperlich, Linton, and Härdle (1999) and elsewhere. In the computation of $\widehat{\mu}_1(.)$ and $\widehat{\mu}_4(.)$ we used a Gaussian kernel and Silverman's rule of thumb bandwidth. Although this is in no way likely to be optimal for this problem, it is a convenient and hence fairly widely used method in practice.

In Table 1 and 2 we report four different performance measures: pointwise root mean squared error (PRMSE), pointwise mean absolute error (PMAE), integrated root mean squared error (IRMSE), and integrated mean absolute error (IMAE). Crooker and Herriges (2000) just report the pointwise results, at the central point $x = 0$, so we chose this point for our corresponding calculations. Thus, their Table 2a ($n = 100$) and Appendix Table 1a ($n = 300$) are directly comparable with a subset of our results. Our conclusions are:

(A1) The performance of our estimators generally improves with sample size according to all measures: the pointwise measures at approximately the right rate, while the integrated measures improve much more slowly.

(A2) Estimators $\widehat{\mu}_1(.)$ and $\widehat{\mu}_4(.)$ despite using nonparametric methods, which $\widehat{\mu}_3(.)$ and $\widehat{\mu}_5(.)$ do not use, perform very well. From Table 1, the ranking of the estimators seems to be[9]:

By pointwise performance criteria: In small samples $\widehat{\mu}_1 \prec \widehat{\mu}_5 \prec \widehat{\mu}_4 \prec \widehat{\mu}_3$ and in large samples $\widehat{\mu}_1 \prec \widehat{\mu}_4 \prec \widehat{\mu}_5 \prec \widehat{\mu}_3$

By integrated performance criteria: In small samples $\widehat{\mu}_5 \prec \widehat{\mu}_4 \prec \widehat{\mu}_1 \prec \widehat{\mu}_3$ and in large samples $\widehat{\mu}_4 \prec \widehat{\mu}_5 \prec \widehat{\mu}_1 \prec \widehat{\mu}_3$

(A3) The only consistent ranking across designs is that $\widehat{\mu}_3$ always performs the worst.

(A4) Estimators $\widehat{\mu}_1(.)$, $\widehat{\mu}_4(.)$, and $\widehat{\mu}_5(.)$ seem to perform better than Crooker and Herriges SNP estimator, especially in the large σ case.

[9] Here, $a \prec b$ means a better than b.

(A5) The estimates of $\mathrm{std}(W|X=x)$ are subject to much more variability than the estimates of $E[W|X=x]$, and there does seem to be some suggestion of inconsistency in the large σ case.

Although the method seems to work reasonably well in this discrete bid case, based on asymptotic theory we should obtain better results when the bid distribution is actually continuous. We repeated the above experiments with bid distribution uniform on $[25, 175]$ and report the results in Tables 3 and 4. Our conclusions are:

(B1) The performance in the continuous design is somewhat better than in the discrete design, e.g., 70% of the numbers are larger in Table 1 than in Table 3. Surprisingly, for some designs the pointwise results in Table 1 are better, but the integrated results are always better in Table 3.

(B2) The ranking of the estimators is the same in Table 3 as Table 1. Once again $\widehat{\mu}_3$ always performs the worst, but the rankings of the other estimators vary depending on the criterion and sample size.

(B3) The results for standard deviation are much better in most case in Table 4.

Perhaps the reason for the good pointwise performance in Table 3 is that the chosen point of evaluation $x=0$ corresponds to $E[W|X=0]=100$ and there is a point mass in the distribution of the bids at this point.

6.2 Application

We examine a dataset used in An (2000), which is from a contingent valuation study conducted by Hanemann et al. (1991) to elicit the WTP for protecting wetland habitats and wildlife in California's San Joaquin Valley. Each respondent was assigned a bid value. They were then also given a second bid that was either higher or lower than the first, depending on their acceptance or rejection of the first bid. The total number of bid values is 14. The dataset consists of bid responses and some personal characteristics of the respondents. The covariates are age and number of years resident in California, education and income bracket, and binary indicators of sex, race, and membership in an environmental organization. The sample size, after excluding nonrespondents, incomplete responses, etc., is $n=530$.

Because there are seven covariates and only a limited number of bid values, we primarily consider semiparametric specifications for W, in particular:

$$W = X_i'\theta - \varepsilon \text{ and } \log(W) = X_i'\theta - \varepsilon.$$

With these linear and log linear specifications we estimate the quantity

$$\mu_w(x) = E(W|X=x)$$

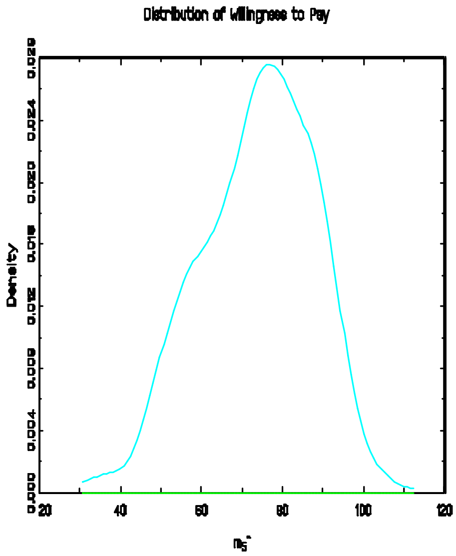

Figure 1:

using our semiparametric estimators $\widehat{\mu}_j(x)$, $j = 3, 4, 5$.[10] To check for possible framing effects, we estimate this conditional mean WTP separately using first bid data and second bid data.

Figure 1 shows a kernel based estimate of the density of the data points $\widehat{\mu}_5(X_i)$, $i = 1, ..., n$ using the linear specification with first bid data. This may be interpreted as the distribution of estimated WTP across the sample. Similar results are obtained using $\widehat{\mu}_3$ and $\widehat{\mu}_4$ and (in shape but not location) using second bid data. There is obviously quite a bit of dispersion around the mean, but the distribution looks quite symmetric.

In Table 5 we report the sample average of the estimates of $E(W \mid X = X_i)$, denoted $\overline{\widehat{\mu}_j}$, as well as $E(W \mid X = \overline{X})$, denoted $\widehat{\mu}_j(\overline{X})$, along with bootstrap standard errors. The bootstrap data were drawn with replacement from $\{Y_i, V_i, X_i\}_{i=1}^n$. The computation of $\widehat{\mu}_j$ is done exactly as described in the simulation section.

[10] For the linear model this means taking $r(x, w) = w$, and for the log linear model w, v, and r are replaced with $\log(w)$, $\log(v)$ and $r(x, w) = e^w$.

	Linear		Log Linear	
	bid1	bid2	bid1	bid2
$\overline{\widehat{\mu}_3}$/s.e.	68.3417/95.6285	274.0677/172.0449	62.0320/4.4683	306.0211/411.4603
$\widehat{\mu}_3(\overline{X})$/s.e.	68.3417/95.6285	274.0677/172.0449	61.5918/4.2751	302.4752/328.7766
$\overline{\widehat{\mu}_4}$/s.e.	61.3042/90.5139	213.0857/120.3624	64.6992/5.0823	369.2809/394.6291
$\widehat{\mu}_4(\overline{X})$/s.e.	61.3042/90.5139	213.0857/120.3624	63.7869/4.4995	472.5140/328.2098
$\overline{\widehat{\mu}_5}$/s.e.	73.7925/8.4186	143.4519/13.6322	99.1164/4.1348	141.5369/9.0742
$\widehat{\mu}_5(\overline{X})$/s.e.	73.7925/8.4186	143.4519/13.6322	98.7726/6.6526	134.0196/21.4996

Table 5: Estimates of WTP

Table 6 provides parameter estimates along with their bootstrap standard errors, and asterisks indicating significant departure from zero at the 5% level.

	Linear		Log Linear	
	bid1	bid2	bid1	bid2
YEARCA/s.e.	0.6869/0.4724	1.6964/0.8023	0.0021/0.0022	0.0131/0.0062*
SEX/s.e.	10.3763/12.3460	33.3371/22.4836	-0.0460/0.0632	0.2579/0.1740
ln(AGE)/s.e.	-42.0977/19.2617*	-62.0518/33.1249	-0.2040/0.1088	-0.4801/0.2563
EDUC/s.e.	-0.9696/2.9860	3.9614/5.2206	0.0119/0.0154	0.0307/0.0404
WHITE/s.e.	5.0222/14.3550	27.9634/28.0857	0.1338/0.0797	0.2164/0.2173
ENVORG/s.e.	3.4128/14.6966	12.2217/30.1263	-0.1085/0.0792	0.0946/0.2331
ln(INCOME)/s.e.	7.0079/9.3697	49.0606/19.0478*	0.0972/0.0500*	0.3796/0.1474*

Table 6

The most striking feature of Tables 5 and 6 is that the second bid data yield far larger coefficients, with corresponding larger estimates of WTP. This may be an indicator of framing, shadowing, or anchoring effects, in which hearing the first bid and replying to it affects responses to later bids. See, e.g., McFadden (1994), Green et al. (1998) and Hurd et al. (1998). These results may also be due to small sample problems associated with the survey design, in particular, the distribution of second bids differs markedly from the distribution of first bids, including some far larger bid values. In contrast, An (2000), using a very different methodology, accepts the hypothesis of no framing effects in these data, though he does report some large differences in coefficient estimates based on data using both bids versus just first bid data.

Looking across estimators and specifications, few of the regressors are statistically significant. Income is generally most significant, having a positive effect. Table 5 shows a moderate range of mean WTP estimates from the first bid, while the second bid WTP estimates are far more dispersed and have much larger standard errors. Using different estimators, An (2000) reports WTP at the mean ranging from 155 to 227 (plus one outlier estimate of 1341), which may be compared to our estimates of 62 to 99 for first bid data and 141 to 369 using only second bids.

Finally, we conducted a purely nonparametric analysis with each of the four continuous covariates, one at a time. In Table 7 we report the estimated value of $\mu_w(\overline{X}_j)$ along with bootstrap standard errors for each separate covariate. The implementation is as described in the simulation section. The estimated values of $\mu_w(\overline{X}_j)$ are quite precise and are in the ballpark reported earlier.

	$\widehat{\mu}_1(\overline{X}_j)$	
X_j	bid1	bid2
YEARCA/s.e.	75.5147/10.75004	167.9237/17.7617
AGE/s.e.	91.5519/9.6176	171.3500/22.1864
EDUC/s.e.	94.6993/11.4778	142.3537/34.2841
ln(INCOME)/s.e.	75.5148/10.7500	167.9237/17.7617

Table 7

In Figures 2 and 3 we provide the marginal smooths themselves along with a pointwise 95% confidence interval. In contrast to the semiparametric model which assumes a linear or loglinear relationship, these figures from the nonparametric estimator show some nonlinear effects.

7 Concluding Remarks

We have provided some semiparametric and nonparametric estimators of conditional moments and quantiles of the latent W. These results show the importance of experimental designs in which the distribution of bids V is ideally continuous, but at a minimum possesses a large number of mass points (and so is in some sense approximately continuous). The estimators appear to perform well with both simulated and actual data.

We have for convenience assumed throughout that the support of V (which must contain the support of W) is bounded. Most of the results in this paper should extend to the infinite support case, although some of the estimators may then require asymptotic trimming to deal with issues arising from division by a density estimate when the true density is not bounded away from zero.

The precision of these estimators depends in part on the density h. When designing experiments, one may wish to choose h to maximize efficiency based on the variance estimators.

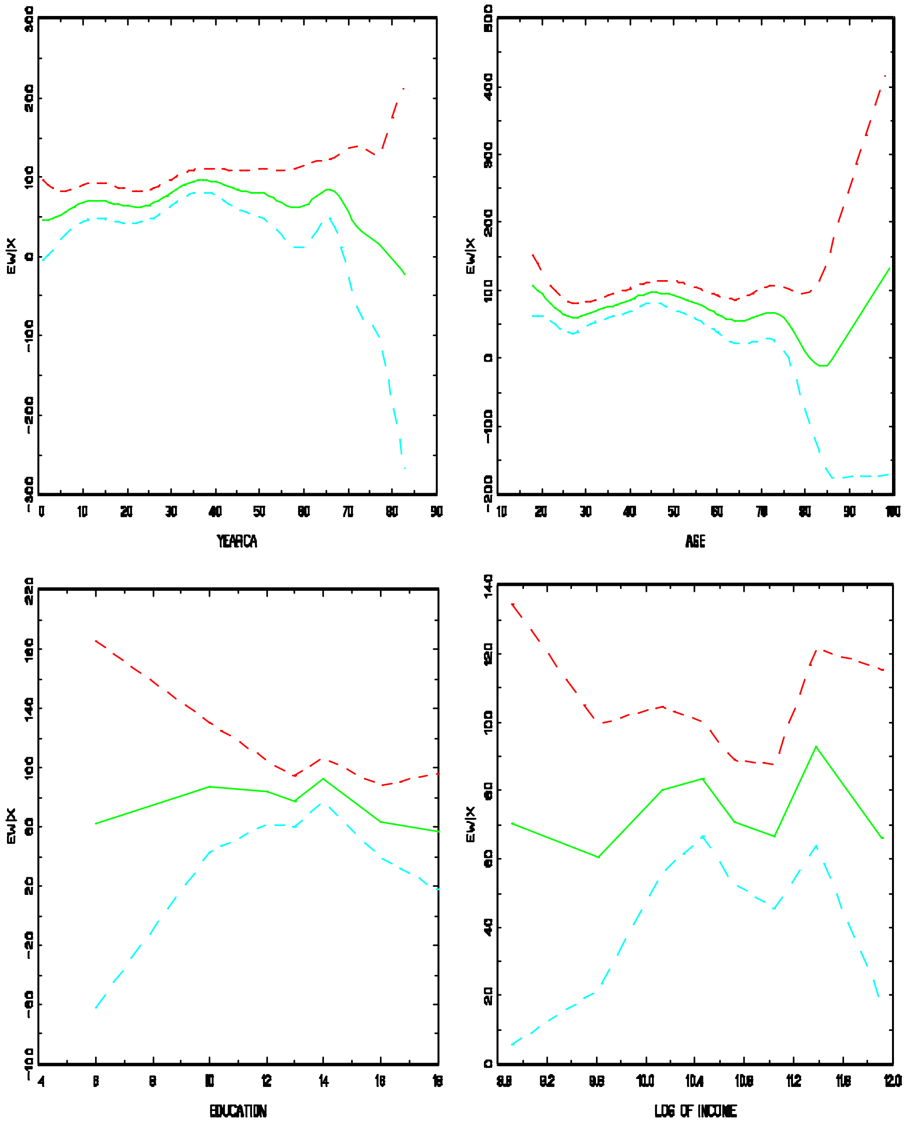

Figure 2: Marginal Local Linear regression Smooths using Bid1 data along with pointwise 95% confidence interval

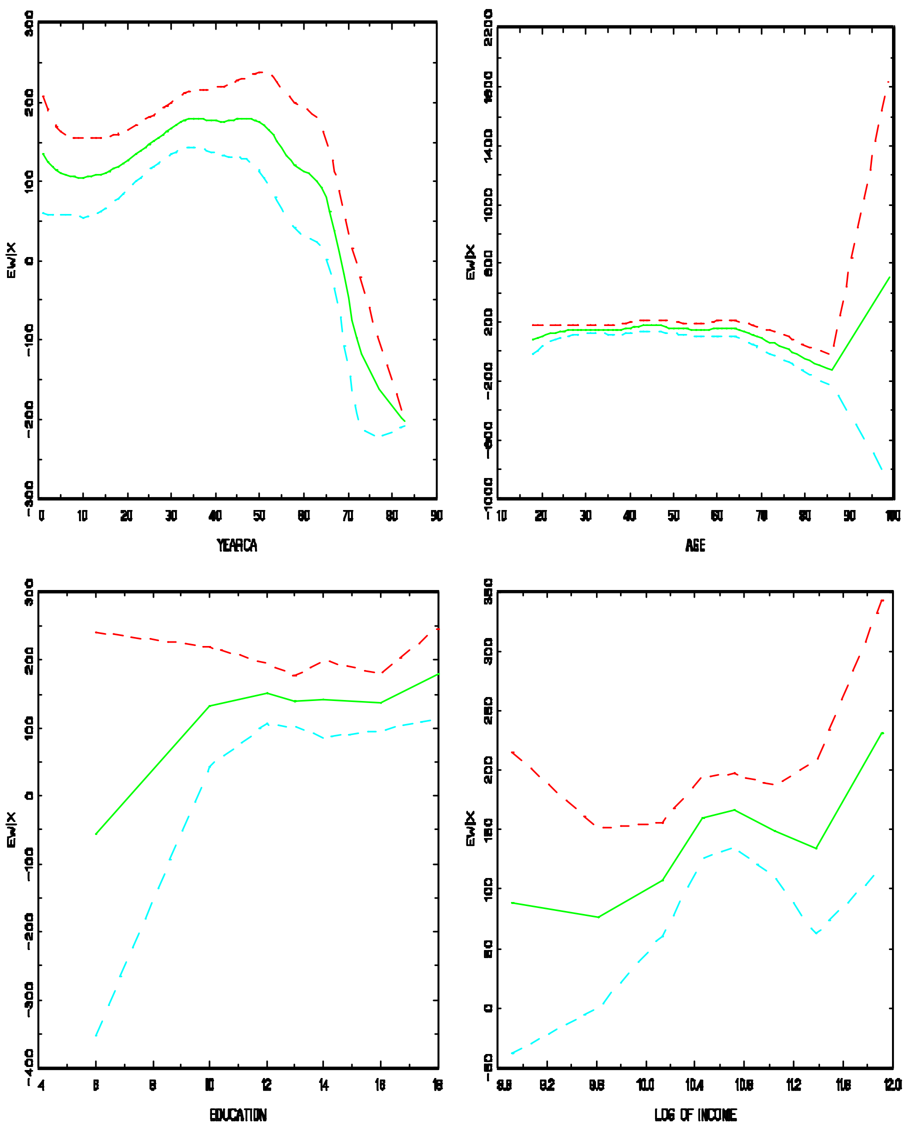

Figure 3: Marginal Local Linear regression Smooths using Bid2 data along with pointwise 95% confidence interval

8 Appendix

8.1 Nonidentification With Discrete Bids

The consistency of our estimators directly shows that moments $\mu_r(x) = E[r(W, X) \mid X = x]$ are nonparametrically identified, given our assumption that V is drawn from a continuous distribution. However, in most WTP applications the distribution of V is discrete. We show here that identification fails given discrete V in the special case of the nonparametric location model $W = g(X) - \varepsilon$ with $\varepsilon \perp X$ and g increasing in a scalar X. It follows a fortiori that identification will fail under our more general conditions. For simplicity in the proof V is assumed to only take on two values, but the basic logic can be extended to an arbitrary finite number of bids.

THEOREM 5. Assume $\operatorname{supp}(X)$ is some open or closed interval on the real line, $\operatorname{supp}(V) = \{-\delta, 0\}$ for some $\delta > 0$, and $W = g(X) - \varepsilon$ with ε having an unknown, strictly monotonic c.d.f. $F_\varepsilon(\varepsilon)$ and g strictly monotonically increasing in X. Assume V, X, ε are mutually independent. Let $Y = I(W > V)$. The functions $g(x)$ and $F_\varepsilon(\varepsilon)$ are not identified given the distribution of Y conditional on V, X.

PROOF OF THEOREM 5. Since Y is binary, the distribution of Y given X and V is $m(v, x) = E[Y \mid X = x, V = v] = F_\varepsilon[g(x) - v]$. Let $\zeta_0 = \inf[\operatorname{supp}(X)]$, $g_0 = g(\zeta_0)$, and $\zeta_j = g^{-1}(g_0 + j\delta)$ for integers j. Let $\widetilde{g}(x)$ be any strictly monotonic function on $x \in [\zeta_0, \zeta_1]$ such that $\widetilde{g}(\zeta_0) = g_0$ and $\widetilde{g}(\zeta_1) = g_0 + \delta$. Define $\widetilde{F}_\varepsilon(\varepsilon)$ on $\varepsilon \in [g_0, g_0 + \delta]$ by $\widetilde{F}_\varepsilon(\varepsilon) = m[0, \widetilde{g}^{-1}(\varepsilon)]$. Next, define $\widetilde{F}_\varepsilon(\varepsilon)$ on $\varepsilon \in (g_0 + \delta, g_0 + 2\delta]$ by $\widetilde{F}_\varepsilon(\varepsilon) = m[\delta, \widetilde{g}^{-1}(\varepsilon - \delta)]$, and define $\widetilde{g}(x)$ on $x \in (\zeta_1, \zeta_2]$ by $\widetilde{g}(x) = \widetilde{F}_\varepsilon^{-1}[m(0, x)]$. Now define $\widetilde{F}_\varepsilon(\varepsilon)$ on $\varepsilon \in (g_0 + 2\delta, g_0 + 3\delta]$ by $\widetilde{F}_\varepsilon(\varepsilon) = m[\delta, \widetilde{g}^{-1}(\varepsilon - \delta)]$, and define $\widetilde{g}(x)$ on $x \in (\zeta_2, \zeta_3]$ by $\widetilde{g}(x) = \widetilde{F}_\varepsilon^{-1}[m(0, x)]$. Continue on in this way until the support of x is exhausted By construction, the functions $\widetilde{g}$ and $\widetilde{F}_\varepsilon$ satisfy $m(v, x) = \widetilde{F}_\varepsilon[\widetilde{g}(x) - v]$ for all x and v on their support, and hence are observationally equivalent to $g(x)$ and $F_\varepsilon(\varepsilon)$. . ■

Notes.

This theorem shows that, with discrete V, nothing can be identified about the function $g(x)$ (except possibly its endpoints) over the interval $x \in [\zeta_0, \zeta_1]$, since the observable data are consistent with $g(x)$ equalling any regular function over that interval. Moreover, $g(x)$ is also not identified on any other interval, because the 'estimate' of its value on other intervals depends on its value in $[\zeta_0, \zeta_1]$.

The same proof could have been started by letting $\widetilde{F}_\varepsilon(\varepsilon)$ be any regular function with the correct endpoints on $\varepsilon \in [g_0, g_0 + \delta]$, then recovering the corresponding $\widetilde{g}$ on that interval, and then proceeding as before. Therefore, the function $\widetilde{F}_\varepsilon$ is also completely unknown (except possibly endpoints) over an initial interval, and unknown elsewhere

The nonidentification is not just an issue of location or scale. The proof assumes $g(x)$ may be known at two points, $g(\zeta_0)$ and $g(\zeta_1)$, which is equivalent to knowing (or choosing) a location and scale for $g(x)$. Similarly, the proof may be started by assuming $\widetilde{F}_\varepsilon(\varepsilon)$ is known at the two points and $\varepsilon = g_0$ and $\varepsilon = g_0 + \delta$, which is equivalent to knowing (or choosing) a location and scale for $\widetilde{F}$. These functions are therefore not identified up to location and scale.

This theorem's nonparametric location model implies $E[W \mid X = x] = g(x) - E(\varepsilon)$, so the nonidentification of $g(x)$ up to any location shows nonidentification of mean WTP. Other moments are likewise not identified.

This theorem can be applied to show nonidentification of other closely related models. In particular, it implies nonidentification of the nonparametric ordered choice model $Y = jI(\alpha_j < g(x) - \varepsilon \leq a_{j+1})$ for a set of integers j and threshold constants α_j (two of which can be normalized to zero and one to pin down the location of ε and the scaling of both ε and g) It also shows nonidentification of the model considered by Das (2002), in which $W = g(x) - \varepsilon$ and one only observes which of a few different fixed intervals each observation W lies in. With a partial parameterization, this model is what An (2000) and others call a double bounded dichotomous choice.

The implications of this theorem for bid design differ markedly from results on optimal bid design in parametric models. Summarizing Kanninen (1993), Crooker and Herriges (2000) say, ”estimates of the mean WTP are best with relatively few bid levels.”

Some existing estimators implicitly assume identification, such as the SNP estimator of Chen and Randall (1997), which they apply to data in which bids are drawn from a finite number of mass points. This theorem shows that such models are not identified with a finite discrete bid distribution.

Our empirical application also uses only available discrete bid data. Our Monte Carlo analysis shows that our estimators work reasonably well with a discrete bid distribution, though identification, consistency, and our limiting distribution theory assume bids drawn from a continuous distribution. A formal asymptotic theory to rationalize the use of available discrete bid data could be devised by assuming that the number of different bid values grows to infinity with the sample size, either by having the observed mass points be drawn from some underlying continuous distribution, or be deterministically chosen in a way that becomes dense in the range of possible WTP values. Crooker and Herriges' (2000) monte carlo design, which we also use, employs this feature of an increasingly fine grid of bid values.

With finite data, both our estimator and others like Chen and Randall (1997) and Das (2002) that are not parameterized essentially smooth between the different available bid values to yield estimates. An (2000) provides a semiparametric model that identifies and estimates the WTP distribution only at the available bid levels, and explicitly interpolates these estimates to obtain an estimate of WTP at the mean.

8.2 Regularity Conditions

We first state some regularity conditions that are needed for the nonparametric estimation of h:

Assumption B.1. k is a symmetric probability density with bounded support, and is Lipschitz continuous on its support, i.e.,

$$|k(t) - k(s)| \leq c|t - s|$$

for some constant c.

Assumption B.2. The variables (V, X) are continuously distributed with Lebesgue density $f_{V,X}(v, x)$ that satisfies $\inf_{\alpha_0(x) \leq v \leq \alpha_1(x)} f_{V,X}(v, x) > 0$. Furthermore, m and $f_{V,X}$ are twice continuously differentiable for all v with $\alpha_0(x) \leq v \leq \alpha_1(x)$. The set $[\alpha_0(x), \alpha_1(x)] \times \{x\}$ is strictly contained in the support of (V, X).

We also need some conditions on the estimator and on the regression functions and densities.

Assumption C.1. Suppose that

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(Z_i, \theta_0) + o_p(1)$$

for some function m such that $E[m(Z_i, \theta_0)] = 0$ and $\quad = E[m(Z_i, \theta_0) m(Z_i, \theta_0)'] < \infty$.

Assumption C.2. The function g is twice continuously differentiable in θ and

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \left\| \frac{\partial g}{\partial \theta}(x, \theta) \right\| \leq d_1(x) \quad ; \quad \sup_{\|\theta - \theta_0\| \leq \delta_n} \left\| \frac{\partial^2 g}{\partial \theta \partial \theta'}(x, \theta) \right\| \leq d_2(x)$$

with $Ed_1(X_i) < \infty$ and $Ed_2(X_i) < \infty$.

Assumption C.3. The density function h is continuous and is strictly positive on its support and is twice continuously differentiable.

Assumption C.4. The kernel k is twice continuously differentiable on its support, and therefore $\sup_t |k''(t)| < \infty$. The bandwidth b satisfies $b \to 0$ and $nb^6 \to \infty$.

8.3 Distribution Theory for Nonparametric Estimators

Proof of Theorem 2. The local linear estimator satisfies the following expansion

$$\widehat{m}(v, x) - m(v, x) = \sum_{i=1}^{n} \widetilde{w}_{ni}(v, x)\varepsilon_i + \delta_n \beta(v, x) + R_n(v, x), \tag{8}$$

where $\widetilde{w}_{ni}(v, x)$ are approximations to $w_{ni}(v, x)$ and $R_n(v, x)$ is a remainder term that contains the various approximation errors described below. We next state conditions under which an estimator with expansion (8) is asymptotically normal.

LEMMA 2.1. *Suppose that: (i)* $0 < \underline{\sigma}^2 \leq \sigma_i^2 \leq \overline{\sigma}^2 < \infty$; *(ii)* $R_n(v,x)/\min\{\alpha_n, \delta_n\} \to^p 0$, *where* $\alpha_n = 1/\sqrt{\sum_{i=1}^n \widetilde{w}_{ni}^2(v,x)} \to^p 0$; *and (iii)* $\max_{1\leq i\leq n} \widetilde{w}_{ni}^2(v,x)/\sum_{i=1}^n \widetilde{w}_{ni}^2(v,x) \to^p 0$. *Then*

$$\frac{\widehat{m}(v,x) - m(v,x) - \delta_n \beta(v,x)}{\sqrt{\sum_{i=1}^n \widetilde{w}_{ni}^2(v,x)\sigma_i^2}} \xrightarrow{d} N(0,1).$$

This result is a standard application of the Lindeberg-Feller central limit theorem. The magnitude of the bias term, δ_n, depends on the method used and on the smoothness of m (and perhaps also on the smoothness of the covariate density). The magnitude of α_n depends on the estimation method and on the covariate density in general. The optimal rate in the central limit theorem is achieved when α_n and δ_n are the same magnitude.

On substitution, it follows that

$$\begin{aligned}
\widehat{\mu}_{2r}(x) - \mu_r(x) &= \int_{\alpha_0(x)}^{\alpha_1(x)} r'(v,x)[\widehat{m}(v,x) - m(v,x)]dv \\
&= \sum_{i=1}^n \overline{w}_{ni}(x)\varepsilon_i + \delta_n \overline{\beta}(x) + \overline{R}_n(x),
\end{aligned}$$

where: $\overline{w}_{ni}(x) = \int \widetilde{w}_{ni}(v,x)d\lambda(v)$, $\overline{\beta}(x) = \int \beta(v,x)d\lambda(v)$, and $\overline{R}_n(x) = \int R_n(v,x)d\lambda(v)$. We next state conditions under which $\widehat{\mu}_{2r}(x)$ is asymptotically normal.

LEMMA 2.2. *Suppose that condition (i) from Lemma 2.1 is true, and that:* (i) $\overline{R}_n(x)/\min\{\overline{\alpha}_n, \delta_n\} \to^p 0$, *where* $\overline{\alpha}_n = 1/\sqrt{\sum_{i=1}^n \overline{w}_{ni}^2(x)} \to^p 0$; *and* (ii) $\max_{1\leq i\leq n} \overline{w}_{ni}^2(x)/\sum_{i=1}^n \overline{w}_{ni}^2(x) \to^p 0$. *Then,*

$$\frac{\widehat{\mu}_{2r}(x) - \mu_r(x) - \delta_n \overline{\beta}(x)}{\sqrt{\sum_{i=1}^n \overline{w}_{ni}^2(x)\sigma_i^2}} \xrightarrow{d} N(0,1). \tag{9}$$

Under Assumptions B1 and B2 the expansion (8) holds with $\delta_n = b^2$ and $\beta(v,x)$ is as stated above, the weights

$$\widetilde{w}_{ni}(v,x) = \frac{1}{f_{V,X}(v,x)} \frac{1}{nb^{d+1}} K\left(\frac{v - V_i}{b}\right) K\left(\frac{x - X_i}{b}\right),$$

while the remainder term satisfies

$$\sup_{\alpha_0(x)\leq v\leq \alpha_1(x)} |R_n(v,x)| = O_p\left(\frac{\sqrt{\log n}}{nb^{d+1}}\right) + o_p(b^2).$$

See for example Masry (1996a, 1996b).

Provided $nb^{d+2}/\log n \to \infty$, condition (i) of Lemma 2.2 is satisfied because $\overline{\alpha}_n = O_p(1/\sqrt{nb^d})$ as we now show. We have

$$
\begin{aligned}
& \left|\overline{w}_{ni}(x) - \frac{-1}{nb^d} K\left(\frac{x-X_i}{b}\right) \frac{r'(V_i,x)}{f_{V,X}(V_i,x)}\right| \\
= & \left|\frac{1}{nb^d} K\left(\frac{x-X_i}{b}\right)\left[\int_{\alpha_0(x)}^{\alpha_1(x)} \frac{r'(v,x)}{f_{V,X}(v,x)} \frac{1}{b} K\left(\frac{v-V_i}{b}\right) dv - \frac{r'(V_i,x)}{f_{V,X}(V_i,x)}\right]\right| \\
= & \left|\frac{1}{nb^d} K\left(\frac{x-X_i}{b}\right) \int_{\alpha_0(x)}^{\alpha_1(x)} \left[\frac{r'(v,x)}{f_{V,X}(v,x)} - \frac{r'(V_i,x)}{f_{V,X}(V_i,x)}\right] \frac{1}{b} K\left(\frac{v-V_i}{b}\right) dv\right| \text{ for large } n \\
= & O(b^2)
\end{aligned}
$$

by a change of variables and dominated convergence argument that is in wide use in nonparametrics; see, e.g., Newey and McFadden (1994, section 8). It works in this case because the set $[\alpha_0(x), \alpha_1(x)]$ is contained in the support of V and the conditions on K etc. Therefore, the asymptotic variance of $\widehat{\mu}_{2r}(x)$ is

$$
\sum_{i=1}^{n} \overline{w}_{ni}^2(x)\sigma_i^2 \simeq \frac{1}{n^2 b^{2d}} \sum_{i=1}^{n} K^2\left(\frac{x-X_i}{b}\right)\left(\frac{r'(V_i,x)}{f_{V,X}(V_i,x)}\right)^2 \sigma_i^2 = O_p(n^{-1}b^{-d}),
$$

as follows from Markov's inequality. Therefore, $\overline{\alpha}_n = O_p(1/\sqrt{nb^d})$ as required. In fact, $\sum_{i=1}^{n} \overline{w}_{ni}^2(x)\sigma_i^2$ satisfies a law of large numbers and is approximately

$$
\begin{aligned}
& \frac{1}{nb^d} E\left[\frac{1}{b^{d_X}} K^2\left(\frac{x-X_i}{b}\right)\left(\frac{r'(V_i,x)}{f_{V,X}(V_i,x)}\right)^2 \sigma^2(V_i,X_i)\right] \\
\simeq & \frac{1}{nb^d} \|K\|^2 \int \sigma^2(v,x)\left(\frac{r'(v,x)}{f_{V,X}(v,x)}\right)^2 f_{V,X}(v,x)dv,
\end{aligned}
$$

where $\sigma^2(V_i,X_i) = \sigma_i^2$, by a change of variables and dominated convergence. Furthermore, condition (ii) of Lemma 2.2 is satisfied by the arguments used in Gozalo and Linton (1999, Lemma CLT). ■

8.4 Distribution Theory for Semiparametric Quantities

Let E_i denote expectation conditional on Z_i.

Proof of Theorem 3. Recall that

$$
\widehat{\mu}_{3r}^*(x;\widehat{\theta}) = r[g(x,\widehat{\theta}),x] + \frac{1}{n}\sum_{i=1}^{n} \frac{r'[g(x,\widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\widehat{\psi}(\widehat{U}_i)},
$$

where $\widehat{U}_i = g(X_i,\widehat{\theta}) - V_i$ and

$$
\widehat{\psi}(\widehat{U}_i) = \frac{1}{n}\sum_{j=1}^{n} h[g(X_j,\widehat{\theta}) - \widehat{U}_i].
$$

By a geometric series expansion we can write

$$\begin{aligned}
\widehat{\mu}^*_{3r}(x;\widehat{\theta}) &= \frac{1}{n}\sum_{i=1}^n f_1(Z_i,\widehat{\theta}) - \frac{1}{n}\sum_{i=1}^n f_2(Z_i,\theta_0)[\widehat{\psi}(\widehat{U}_i) - \psi(U_i)] \\
&\quad -\frac{1}{n}\sum_{i=1}^n [f_2(Z_i,\widehat{\theta}) - f_2(Z_i,\theta_0)][\widehat{\psi}(\widehat{U}_i) - \psi(U_i)] \\
&\quad +\frac{1}{n}\sum_{i=1}^n \frac{r'[g(x,\widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\psi^2(U_i)\widehat{\psi}(\widehat{U}_i)}[\widehat{\psi}(\widehat{U}_i) - \psi(U_i)]^2,
\end{aligned}$$

where

$$f_2(Z_i,\theta) = \frac{r'[g(x,\theta) - U_i(\theta), x][Y_i - 1(U_i(\theta) > 0)]}{\psi^2(U_i)}.$$

Leading Terms. We make use of Lemmas 1 and 2 given below. Lemma 1 implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [f_1(Z_i,\widehat{\theta}) - Ef_1(Z_i,\theta_0)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\Gamma_F m(Z_i,\theta_0) + [f_1(Z_i,\theta_0) - Ef_1(Z_i,\theta_0)]\} + o_p(1). \quad (10)$$

Furthermore, by Lemma 2

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n f_2(Z_i,\theta_0)[\widehat{\psi}(\widehat{U}_i) - \psi(U_i) - \frac{1}{n}\sum_{j=1}^n L(Z_i,Z_j)]\right| \\
\le\ & \frac{1}{n}\sum_{i=1}^n |f_2(Z_i,\theta_0)| \times \max_{1\le i\le n}\left|[\widehat{\psi}(\widehat{U}_i) - \psi(U_i) - \frac{1}{n}\sum_{j=1}^n L(Z_i,Z_j)]\right| \\
=\ & o_p(n^{-1/2}),
\end{aligned}$$

where $L(Z_i,Z_j) = \xi_j(U_i) + \Gamma(Z_i)m(Z_j;\theta_0)$, and

$$\xi_j(u) = h[g(X_j,\theta_0) - u] - E(h[g(X_j,\theta_0) - u])$$

$$\Gamma(Z_i) = \psi'(U_i)\left(\frac{\partial g}{\partial\theta}(X_i,\theta_0) - E\left[\frac{\partial g}{\partial\theta}(X_i,\theta_0)\right]\right).$$

Then

$$\begin{aligned}
\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n f_2(Z_i,\theta_0)L(Z_i,Z_j) &= \sum_{i=1}^n\sum_{j=1}^n \varphi_n(Z_i,Z_j) \\
&= \frac{1}{n}\sum_{j=1}^n \omega(Z_j) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\omega(Z_j) = E\left[f_2(Z_i,\theta_0)\Gamma(Z_i)\right]m(Z_j;\theta_0) + E\left[f_2(Z_i,\theta_0)\xi_j(U_i)|Z_j\right]$$

by standard U-statistic theory. We have

$$
\begin{aligned}
E\left[f_2(Z_i,\theta_0)\xi_j(U_i)|Z_j\right] &= E\left[f_2(Z_i,\theta_0)h[g(X_j,\theta_0)-U_i]|Z_j\right] - E\left[f_2(Z_i,\theta_0)\psi(U_i)\right] \\
&= E\left[(f_1(Z_i,\theta_0)-r[g(x,\theta_0),x])\frac{h[g(X_j,\theta_0)-U_i]}{\psi(U_i)}|Z_j\right] \\
&\quad -E\left[(f_1(Z_i,\theta_0)-r[g(x,\theta_0),x])\right] \\
&= E\left[\frac{r'[g(x,\theta)-U_i,x][Y_i-1(U_i>0)]}{\psi(U_i)}\frac{h[g(X_j,\theta_0)-U_i]-\psi(U_i)}{\psi(U_i)}|Z_j\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[f_2(Z_i,\theta_0)\Gamma(Z_i)\right] &= E\left[f_2(Z_i,\theta_0)\psi(U_i)\frac{\psi'(U_i)}{\psi(U_i)}\left(\frac{\partial g}{\partial\theta}(X_i,\theta_0)-E\left[\frac{\partial g}{\partial\theta}(X_i,\theta_0)\right]\right)\right] \\
&= E\left[\frac{r'[g(x,\theta_0)-U_i,x][Y_i-1(U_i>0)]}{\psi(U_i)}\frac{\psi'(U_i)}{\psi(U_i)}\gamma_i\right],
\end{aligned}
$$

so that the leading terms are as stated.

Remainders. By the Cauchy-Schwarz inequality

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}[f_2(Z_i,\widehat{\theta})-f_2(Z_i,\theta_0)][\widehat{\psi}(\widehat{U}_i)-\psi(U_i)]\right| \\
\leq\ &\left(\frac{1}{n}\sum_{i=1}^{n}[f_2(Z_i,\widehat{\theta})-f_2(Z_i,\theta_0)]^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}[\widehat{\psi}(\widehat{U}_i)-\psi(U_i)]^2\right)^{1/2} \\
=\ &O_p(n^{-1})
\end{aligned}
$$

from another application of Lemmas 1 and 2.

We have assumed that $\inf_{u\in\mathcal{U}}\psi(u)>0$, which implies that

$$
\min_{1\leq i\leq n}\psi(\widehat{U}_i)=\inf_{u\in\mathcal{U}}\psi(u)+O_p(n^{-1/2})
$$

is bounded away from zero with probability tending to one. Therefore,

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\frac{r'[g(x,\widehat{\theta})-\widehat{U}_i,x][Y_i-1(\widehat{U}_i>0)]}{\psi^2(\widehat{U}_i)\widehat{\psi}(\widehat{U}_i)}[\widehat{\psi}(\widehat{U}_i)-\psi(\widehat{U}_i)]^2\right| \\
\leq\ &\frac{\sup_{u\in\mathcal{U}}[\widehat{\psi}(u)-\psi(u)]^2+O_p(n^{-1/2})}{\inf_{u\in\mathcal{U}}\psi^2(u)\widehat{\psi}(u)+O_p(n^{-1/2})}\frac{1}{n}\sum_{i=1}^{n}|r'[g(x,\widehat{\theta})-\widehat{U}_i,x]|\cdot(|Y_i|+1) \\
=\ &O_p(n^{-1}).
\end{aligned}
$$

In conclusion,

$$\sqrt{n}[\widehat{\mu}_{3r}^*(x;\widehat{\theta}) - \mu_r(x;\theta_0)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_i + o_p(1),$$

as required. The asymptotic distribution of $\sqrt{n}[\widehat{\mu}_{3r}^*(x;\widehat{\theta}) - \mu_r(x;\theta_0)]$ follows from the central limit theorem for independent random variables.

■

Proof of Theorem 4. By a geometric series expansion we can write

$$\begin{aligned}
\widehat{\mu}_{4r}^*(x;\widehat{\theta}) &= \frac{1}{n}\sum_{i=1}^n f_1(Z_i,\widehat{\theta}) - \frac{1}{n}\sum_{i=1}^n f_2(Z_i,\theta_0)[\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)] \\
&\quad -\frac{1}{n}\sum_{i=1}^n [f_2(Z_i,\widehat{\theta}) - f_2(Z_i,\theta_0)] \times [\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)] \\
&\quad +\frac{1}{n}\sum_{i=1}^n \frac{r'[g(x,\widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\psi^2(U_i)\widetilde{\psi}(\widehat{U}_i)}[\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)]^2.
\end{aligned}$$

Leading Terms. We make use of Lemma 3 given below. The term $n^{-1}\sum_{i=1}^n f_1(Z_i,\widehat{\theta})$ has already been analyzed above. By Lemma 3 we have with probability tending to one

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^n f_2(Z_i,\theta_0)\left[[\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)] - \frac{1}{n}\sum_{j=1}^n L^*(Z_i,Z_j)\right]\right| &\leq \frac{1}{nb^3}\left(\frac{1}{n}\sum_{i=1}^n |f_2(Z_i,\theta_0)|d(X_i)\right) \\
&= O_p(n^{-1}b^{-3}) \qquad (11)
\end{aligned}$$

for some function $d(.)$ and random variables $L^*(Z_i,Z_j) = b^{-1}k((U_i - U_j)/b) - \psi(U_i) + \Gamma^*(Z_i)\cdot m(Z_j,\theta_0)$, where

$$\Gamma^*(Z_i) = \psi'(U_i)\left[\frac{\partial g}{\partial\theta'}(X_i,\theta_0) - E\left[\frac{\partial g}{\partial\theta'}(X_i,\theta_0)|U_i\right]\right] - \psi(U_i)\phi_u'(U_i).$$

Under our bandwidth conditions, the right hand side of (11) is $o_p(n^{-1/2})$. Furthermore,

$$\frac{1}{n}\sum_{i=1}^n f_2(Z_i,\theta_0)\frac{1}{n}\sum_{j=1}^n L^*(Z_i,Z_j) = \sum_{i=1}^n\sum_{j=1}^n \varphi_n(Z_i,Z_j)$$

where

$$\varphi_n(Z_i,Z_j) = \frac{1}{n^2}f_2(Z_i,\theta_0)\left[\frac{1}{b}k\left(\frac{U_i - U_j}{b}\right) - \psi(U_i) + \Gamma^*(Z_i)\cdot m(Z_j,\theta_0)\right].$$

Note that $E_i\varphi_n(Z_i,Z_j) = 0$ but $E_j\varphi_n(Z_i,Z_j) \neq 0$. We write

$$T_{n4} = \frac{1}{n}\sum_{j=1}^n \psi(Z_j) + \sum_{\substack{i=1 \\ i\neq j}}^n\sum_{j=1}^n \omega_n(Z_i,Z_j),$$

where

$$\psi(Z_j) = n^2 E_j \varphi_n(Z_i, Z_j)$$

$$\omega_n(Z_i, Z_j) = \varphi_n(Z_i, Z_j) - E_j \varphi_n(Z_i, Z_j),$$

so that $\omega_n(Z_i, Z_j)$ is a degenerate kernel satisfying $E_i \omega_n(Z_i, Z_j) = E_j \omega_n(Z_i, Z_j) = 0$. We next compute $\psi_n(Z_j)$; using integration by parts, a change of variable, and dominated convergence we have

$$\psi(Z_j) = (f_2(Z_j, \theta_0) - E[f_2(Z_j, \theta_0)|U_j])\, \psi(U_j) + E\left[f_2(Z_i, \theta_0)\Gamma^*(Z_i)\right] \cdot m(Z_j, \theta_0) + O_p(b^2).$$

Finally,

$$(f_2(Z_i, \theta_0) - E[f_2(Z_i, \theta_0)|U_i])\, \psi(U_i) = \frac{r'[g(x, \theta_0) - U_i, x]}{\psi(U_i)} (Y_i - E[Y_i|U_i])$$

$$E[f_2(Z_i, \theta_0)\Gamma^*(Z_i)] = E\left[\frac{r'[g(x, \theta_0) - U_i, x]}{\psi(U_i)}[Y_i - 1(U_i > 0)] \times \left(\frac{\psi'(U_i)}{\psi(U_i)}\gamma_i^* - \phi_u'(U_i)\right)\right].$$

Remainder Terms. First,

$$\begin{aligned} & \left|\frac{1}{n}\sum_{i=1}^n [f_2(Z_i, \widehat{\theta}) - f_2(Z_i, \theta_0)][\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)]\right| \\ \leq & \left(\frac{1}{n}\sum_{i=1}^n [f_2(Z_i, \widehat{\theta}) - f_2(Z_i, \theta_0)]^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^n [\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)]^2\right)^{1/2} \\ = & \; o_p(n^{-1/2}). \end{aligned}$$

Second

$$\begin{aligned} & \left|\frac{1}{n}\sum_{i=1}^n \frac{r'[g(x, \widehat{\theta}) - \widehat{U}_i, x][Y_i - 1(\widehat{U}_i > 0)]}{\psi^2(U_i)\widetilde{\psi}(\widehat{U}_i)}[\widetilde{\psi}(\widehat{U}_i) - \psi(U_i)]^2\right| \\ \leq & \frac{\sup_{u\in\mathcal{U}}[\widetilde{\psi}(u) - \psi(u)]^2(1 + o_p(1))}{\inf_{u\in\mathcal{U}} \psi^2(u)\widetilde{\psi}(u) + o_p(1)} \frac{1}{n}\sum_{i=1}^n |r'[g(x, \widehat{\theta}) - \widehat{U}_i, x]| \cdot (|Y_i| + 1) \\ = & \; o_p(n^{-1/2}). \end{aligned}$$

■

8.5 Subsidiary Results

Define

$$F_n(\theta) = \frac{1}{n}\sum_{i=1}^n f(Z_i, \theta)$$

for some function f, and let $F(\theta) = EF_n(\theta)$ and $\Gamma_F = \partial F(\theta_0)/\partial\theta$.

LEMMA 1. *Assume:*

(i) For some vector m

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(Z_i, \theta_0) + o_p(1)$$

where $E[m(Z_i, \theta_0)] = 0$ *and* $\quad = E[m(Z_i, \theta_0)m(Z_i, \theta_0)'] < \infty$.

(ii) There exists a finite matrix Γ_F *of full (column) rank such that*

$$\lim_{\|\theta-\theta_0\|\to 0} \frac{\|F(\theta) - \Gamma_F(\theta - \theta_0)\|}{\|\theta - \theta_0\|} = 0.$$

(iii) For every sequence of positive numbers $\{\delta_n\}$ *such that* $\delta_n \to 0$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \left\|\sqrt{n}[F_n(\theta) - F(\theta)] - \sqrt{n}[F_n(\theta_0) - F(\theta_0)]\right\| = o_p(1).$$

Then

$$\sqrt{n}[F_n(\widehat{\theta}) - F(\theta_0)] \Longrightarrow N(0, V),$$

where

$$\begin{aligned} V &= \mathrm{var}[\Gamma_F m(Z_i, \theta_0) + f(Z_i, \theta_0)] \\ &= \Gamma_F \quad \Gamma_F{}' + \mathrm{var}[f(Z_i, \theta_0)] + 2\Gamma_F E m(Z_i, \theta_0) f(Z_i, \theta_0). \end{aligned}$$

See below for a discussion on the verification of (iii).

PROOF. Since $\widehat{\theta}$ is root-n consistent, there exists a sequence $\delta_n \to 0$ such that

$$\Pr[||\sqrt{n}(\widehat{\theta} - \theta_0)|| > \delta_n] \to 0$$

as $n \to \infty$. We can therefore suppose that $||\sqrt{n}(\widehat{\theta} - \theta_0)|| \le \delta_n$ with probability tending to one. We have

$$\begin{aligned} \sqrt{n}[F_n(\widehat{\theta}) - F(\theta_0)] &= \sqrt{n}[F(\widehat{\theta}) - F(\theta_0)] + \sqrt{n}[F_n(\widehat{\theta}) - F(\widehat{\theta})] \\ &= \Gamma_F\sqrt{n}(\widehat{\theta} - \theta_0) + \sqrt{n}[F_n(\theta_0) - F(\theta_0)] + o(\|[\sqrt{n}(\widehat{\theta} - \theta_0)\|) \\ &\quad + \sqrt{n}\{[F_n(\widehat{\theta}) - F(\widehat{\theta})] - [F_n(\theta_o) - F(\theta_0)]\} \\ &= \Gamma_F\sqrt{n}(\widehat{\theta} - \theta_0) + \sqrt{n}[F_n(\theta_0) - F(\theta_0)] + o_p(1)[\text{by (ii) and (iii)}] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Gamma_F m(Z_i, \theta_0) + [f(Z_i, \theta_0) - Ef(Z_i, \theta_0)]\} + o_p(1), \end{aligned}$$

and the result now follows from standard CLT arguments. ■

LEMMA 2. *As* $n \to \infty$

$$\max_{1\le i\le n}\left|\widehat{\psi}(\widehat{U}_i)-\psi(U_i)-\frac{1}{n}\sum_{j=1}^{n}L(Z_i,Z_j)\right|=o_p(n^{-1/2}), \tag{12}$$

where $L(Z_i,Z_j)=\xi_j(U_i)+\Gamma(Z_i)m(Z_j;\theta_0)$ *and*

$$\xi_j(u)=h[g(X_j,\theta_0)-u]-E(h[g(X_j,\theta_0)-u])$$

$$\Gamma(Z_i)=\psi'(U_i)\left(\frac{\partial g}{\partial\theta}(X_i,\theta_0)-E\left[\frac{\partial g}{\partial\theta}(X_i,\theta_0)\right]\right).$$

PROOF. We have for any u,

$$\begin{aligned}
\widehat{\psi}(u)-\psi(u) &= \frac{1}{n}\sum_{j=1}^{n}h[g(X_j,\widehat{\theta})-u]-E(h[g(X_j,\theta_0)-u]) \\
&= \frac{1}{n}\sum_{j=1}^{n}h[g(X_j,\theta_0)-u]-E(h[g(X_j,\theta_0)-u]) \\
&\quad +\frac{1}{n}\sum_{j=1}^{n}h'[g(X_j,\theta_0)-u]\frac{\partial g}{\partial\theta}(X_j,\theta_0)(\widehat{\theta}-\theta_0)+R_n(u),
\end{aligned}$$

where

$$\begin{aligned}
R_n(u) &= \frac{1}{2n}\sum_{j=1}^{n}h''[g(X_j,\overline{\theta})-u](\widehat{\theta}-\theta_0)'\frac{\partial g}{\partial\theta}(X_j,\overline{\theta})\frac{\partial g}{\partial\theta'}(X_j,\overline{\theta})(\widehat{\theta}-\theta_0) \\
&\quad +\frac{1}{2n}\sum_{j=1}^{n}h'[g(X_j,\overline{\theta})-u](\widehat{\theta}-\theta_0)'\frac{\partial^2 g}{\partial\theta\partial\theta'}(X_j,\overline{\theta})(\widehat{\theta}-\theta_0),
\end{aligned}$$

where $\overline{\theta}$ are intermediate values between $\widehat{\theta}$ and θ_0. With probability tending to one for a sequence $\delta_n \to 0$ we have by the Cauchy Schwarz inequality

$$\begin{aligned}
|R_n(u)| &\le ||\widehat{\theta}-\theta_0||^2\frac{1}{2n}\sum_{j=1}^{n}\sup_{\|\theta-\theta_0\|\le\delta_n}|h''[g(X_j,\theta)-u]|\sup_{\|\theta-\theta_0\|\le\delta_n}||\frac{\partial g}{\partial\theta}(X_j,\theta)|| \\
&\quad +||\widehat{\theta}-\theta_0||^2\frac{1}{2n}\sum_{j=1}^{n}\sup_{\|\theta-\theta_0\|\le\delta_n}|h'[g(X_j,\theta)-u]|\sup_{\|\theta-\theta_0\|\le\delta_n}\left\|\frac{\partial^2 g}{\partial\theta\partial\theta'}(X_j,\theta)\right\| \\
&\le ||\widehat{\theta}-\theta_0||^2\cdot\left(\sup_t|h''(t)|\frac{1}{2n}\sum_{j=1}^{n}d_1(X_j)+\sup_t|h'(t)|\cdot\frac{1}{2n}\sum_{j=1}^{n}d_2(X_j)\right) \qquad (13) \\
&= O_p(n^{-1}).
\end{aligned}$$

Since the right hand side of (13) does not depend on u, this order is uniform in u. Furthermore because h' is bounded and continuous, by a standard uniform law of large numbers

$$\sup_{u\in\mathcal{U}}\left|\frac{1}{n}\sum_{j=1}^{n}h'[g(X_j,\theta_0)-u]\frac{\partial g}{\partial\theta}(X_j,\theta_0)-E\left[h'[g(X_j,\theta_0)-u]\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right]\right|=o_p(1),$$

where $\mathcal{U}$ is the support of $U_i=g(X_i,\theta_0)-V_i$. Therefore,

$$\sup_{u\in\mathcal{U}}\left|\widehat{\psi}(u)-\psi(u)-\frac{1}{n}\sum_{j=1}^{n}\xi_j(u)-J(u)(\widehat{\theta}-\theta_0)\right|=o_p(n^{-1/2}),\tag{14}$$

where $\xi_j(u)=h[g(X_j,\theta_0)-u]-E(h[g(X_j,\theta_0)-u])$ are i.i.d. with mean zero and finite variance, and

$$J(u)=E\left[h'[g(X_j,\theta_0)-u]\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right].$$

Because $\sqrt{n}(\widehat{\theta}-\theta_0)=O_p(1)$ the supremum over $u\in\mathcal{U}$ is the same as a maximum over $\widehat{U}_i$.

Furthermore, by a second order Taylor series expansion

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\xi_j(\widehat{U}_i)+J(\widehat{U}_i)(\widehat{\theta}-\theta_0) &= \frac{1}{n}\sum_{j=1}^{n}\xi_j(U_i)+\frac{1}{n}\sum_{j=1}^{n}\frac{\partial\xi_j}{\partial u}(U_i)\frac{\partial g}{\partial\theta}(X_j,\theta_0)(\widehat{\theta}-\theta_0)\\ &\quad +J(U_i)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})\\ &= \frac{1}{n}\sum_{j=1}^{n}\xi_j(U_i)+E_i\left[\frac{\partial\xi_j}{\partial u}(U_i)\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right](\widehat{\theta}-\theta_0)\\ &\quad +J(U_i)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2}),\end{aligned}$$

where

$$\frac{\partial\xi_j}{\partial u}(u)=-h'[g(X_j,\theta_0)-u]+E\left(h'[g(X_j,\theta_0)-u]\right),$$

and the error term is bounded in the same way as above using the continuous second derivatives of h,g. That is, $\max_{1\le i\le n}|J(\widehat{U}_i)-J(U_i)|=O_p(1)$ and

$$\sup_{||\theta-\theta_0||\le\delta_n}\max_{1\le i\le n}\left|\frac{\partial^2\xi_j}{\partial u^2}(U_i(\theta))\right|\le d(Z_i)$$

with $Ed(Z_i)<\infty$. Note that

$$\begin{aligned}J(U_i)+E_i\left[\frac{\partial\xi_j}{\partial u}(U_i)\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right] &= E_i\left[h'[g(X_j,\theta_0)-U_i]\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right]\\ &\quad +E_i\left[\left\{-h'[g(X_j,\theta_0)-U_i]+E_i\left(h'[g(X_j,\theta_0)-U_i]\right)\right\}\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right]\\ &= E_i\left(h'[g(X_j,\theta_0)-U_i]\right)E\left(\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right),\end{aligned}$$

so that

$$\widehat{\psi}(\widehat{U}_i)-\psi(\widehat{U}_i)=\frac{1}{n}\sum_{j=1}^{n}\xi_j(U_i)+\left(E_i\left[h'[g(X_j,\theta_0)-U_i]\right]\cdot E\left[\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right]\right)\cdot(\widehat{\theta}-\theta_0)+o_p(n^{-1/2}). \quad (15)$$

Finally,

$$\psi(\widehat{U}_i)-\psi(U_i)=-E_i\left(h'[g(X_j,\theta_0)-U_i]\right)\cdot\frac{\partial g}{\partial\theta}(X_i,\theta_0)\cdot(\widehat{\theta}-\theta_0)+o_p(n^{-1/2}). \quad (16)$$

These results are uniform under some additional conditions. Combining (15) and (16) we obtain the result (12).

■

LEMMA 3. *We have with probability tending to one*

$$\left|\widetilde{\psi}(\widehat{U}_i)-\psi(U_i)-\frac{1}{n}\sum_{j=1}^{n}L^*(Z_i,Z_j)\right|\leq\frac{k}{nb^3}d(X_i)$$

for some function d, *where*

$$L^*(Z_i,Z_j)=\frac{1}{b}k\left(\frac{U_i-U_j}{b}\right)-\psi(U_i)+\Gamma^*(Z_i)\cdot m(Z_j,\theta_0)$$

$$\Gamma^*(Z_i)=\psi'(U_i)\left[\frac{\partial g}{\partial\theta'}(X_i,\theta_0)-E\left[\frac{\partial g}{\partial\theta'}(X_i,\theta_0)|U_i\right]\right]-\psi(U_i)\phi_u'(U_i).$$

PROOF. Making a second order Taylor series expansion we have $\widetilde{\psi}(\widehat{U}_i)-\psi(U_i)=T_{ni}+R_{ni}$, where

$$T_{ni}=\overline{\psi}(U_i)-\psi(U_i)+\frac{1}{nb^2}\sum_{j=1}^{n}k'\left(\frac{U_i-U_j}{b}\right)\left[\frac{\partial g}{\partial\theta'}(X_i,\theta_0)-\frac{\partial g}{\partial\theta'}(X_j,\theta_0)\right](\widehat{\theta}-\theta_0)$$

$$\begin{aligned}R_{ni} &= \frac{1}{nb^3}\sum_{j=1}^{n}k''\left(\frac{U_i^*-U_j^*}{b}\right)\left[\frac{\partial g}{\partial\theta}(X_i,\theta_0)-\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right](\widehat{\theta}-\theta_0)(\widehat{\theta}-\theta_0)'\left[\frac{\partial g}{\partial\theta}(X_i,\theta_0)-\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right]'\\ &\quad+\frac{1}{nb^2}\sum_{j=1}^{n}k'\left(\frac{U_i-U_j}{b}\right)(\widehat{\theta}-\theta_0)'\left[\frac{\partial^2 g}{\partial\theta\partial\theta'}(X_i,\theta^*)-\frac{\partial^2 g}{\partial\theta\partial\theta'}(X_j,\theta^*)\right](\widehat{\theta}-\theta_0),\end{aligned}$$

where θ^* are intermediate values between $\widehat{\theta}$ and θ_0, and $U_i^*=U_i(\theta^*)$. We have with probability tending to one

$$\begin{aligned}|R_{ni}| &\leq b^{-3}\sup_u|k''(u)|\cdot||\widehat{\theta}-\theta_0||^2\cdot\left(\left\|\frac{\partial g}{\partial\theta}(X_i,\theta_0)\right\|^2+\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{\partial g}{\partial\theta}(X_j,\theta_0)\right\|^2\right)\\ &\quad+b^{-1}||\widehat{\theta}-\theta_0||^2\cdot\frac{1}{nb}\sum_{j=1}^{n}\left|k'\left(\frac{U_i-U_j}{b}\right)\right|(d_1(X_i)+d_2(X_j))\end{aligned}$$

by the Cauchy-Schwarz inequality. By the uniform convergence results of Masry (1996a,1996b):

$$\begin{aligned}
\max_{1\le i\le n} \frac{1}{nb}\sum_{j=1}^{n}\left|k'\left(\frac{U_i-U_j}{b}\right)\right| &= O_p(1)\\
\max_{1\le i\le n} \frac{1}{nb}\sum_{j=1}^{n}\left|k'\left(\frac{U_i-U_j}{b}\right)\right| d_2(X_j) &= O_p(1),
\end{aligned}$$

so that for suitable constants and dominating functions

$$|R_{ni}| \le \frac{k_1}{nb^3}(d_3(X_i)+k_2) + \frac{k_3}{nb}(d_1(X_i)+k_4)$$

with probability tending to one. This gives the result. Furthermore, if the functions d_j are bounded this translates into a uniform result

$$\max_{1\le i\le n} |R_{ni}| = O_p(n^{-1}b^{-3}).$$

Provided $nb^6 \to \infty$, this term is $o_p(n^{-1/2})$. With additional smoothness conditions on k this condition can be substantially weakened.

Furthermore, by Masry (1996a, 1996b)

$$\max_{1\le i\le n}\left|\frac{1}{nb^2}\sum_{j=1}^{n} k'\left(\frac{U_i-U_j}{b}\right) - E\left[\frac{1}{b^2}k'\left(\frac{U_i-U_j}{b}\right)|U_i\right]\right| = O_p(\sqrt{\frac{\log n}{nb^3}})$$

$$\max_{1\le i\le n}\left|\frac{1}{nb^2}\sum_{j=1}^{n} k'\left(\frac{U_i-U_j}{b}\right)\frac{\partial g}{\partial\theta'}(X_j,\theta_0) - E\left[\frac{1}{b^2}k'\left(\frac{U_i-U_j}{b}\right)\phi_u(U_j)\,|U_i\right]\right| = O_p(\sqrt{\frac{\log n}{nb^3}}).$$

Also,

$$\begin{aligned}
&\left|E\left[\frac{1}{b^2}k'\left(\frac{U_i-U_j}{b}\right)\phi_u(U_j)|U_i\right] - [\phi_u(U_i)\psi(U_i)]'\right|\\
=& \left|\int \frac{1}{b^2}k'\left(\frac{U_i-u}{b}\right)\phi_u(u)\psi(u)du - [\phi_u(U_i)\psi(U_i)]'\right|\\
=& \left|\int \frac{1}{b}k\left(\frac{U_i-u}{b}\right)[\phi_u(u)\psi(u)]'du - [\phi_u(U_i)\psi(U_i)]'\right|\\
=& \left|\int k(t)\left([\phi_u(U_i+tb)\psi(U_i+tb)]' - [\phi_u(U_i)\psi(U_i)]'\right)dt\right|\\
=& O_p(b^2)
\end{aligned}$$

by integration by parts, change of variables and dominated convergence using the symmetry of k. This order is uniform in i by virtue of the boundedness and continuity of the relevant functions. Therefore,

$$\max_{1\le i\le n} |T_{ni} - \frac{1}{n}\sum_{j=1}^{n} L^*(Z_i, Z_j)| = o_p(n^{-1/2}).$$

Finally, we have

$$\max_{1\le i\le n} |\frac{1}{n}\sum_{j=1}^{n} L^*(Z_i, Z_j)| = O_p(b^2 + \sqrt{\frac{\log n}{nb}})$$

by standard results for kernel estimates. ■

8.5.1 Stochastic Equicontinuity Results

We now show that condition (iii) of Lemma 1 is satisfied. Let $\Theta_n(c) = \{\theta\colon \sqrt{n}\left|\theta - \theta^0\right| \le c\}$. Since $\sqrt{n}(\widehat{\theta} - \theta^0) = O_p(1)$, for all $\epsilon > 0$ there exists a c_ϵ and an integer n_0 such that for all $n \ge n_0$, $\Pr[\widehat{\theta} \in \Theta_n(c_\epsilon)] \ge 1 - \epsilon$. Define the stochastic process

$$\nu_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(Z_i, \theta) - E[f(Z_i, \theta)], \quad \theta \in \Theta,$$

where

$$f(Z_i, \theta) = r[g(x, \theta), x] + \frac{r'[g(x, \theta) - U_i(\theta), x][Y_i - 1(U_i(\theta) > 0)]}{\psi(U_i)}$$

and define the pseudo-metric

$$\rho(\theta, \theta') = E\left([f(Z_i, \theta) - f(Z_i, \theta')]^2\right),$$

on Θ. Under this metric, the parameter space Θ is totally bounded. We are only interested in the behaviour of this process as θ varies in the small set Θ_n. By writing $\theta = \theta^0 + \gamma n^{-1/2}$, we shall make a reparameterization to $\nu_n(\gamma)$, where $\gamma \in \Gamma(c) \subset \mathbb{R}^p$. We establish the following result:

$$\sup_{\gamma\in\Gamma} |\nu_n(\gamma) - \nu_n(0)| = o_p(1) \tag{17}$$

To prove (17) it is sufficient to show a pointwise law of large numbers, e.g., $\nu_n(\gamma) - \nu_n(0) = o_p(1)$ for any $\gamma \in \Gamma$, and stochastic equicontinuity of the process ν_n at $\gamma = 0$. The pointwise result is immediate because the random variables are sums of i.i.d. random variables with finite absolute moment and zero mean; the probability limit of $\nu_n(\gamma)$ is the same for all $\gamma \in \Gamma$ by the smoothness of the expected value in γ. To complete the proof of (17) we shall use the following lemma, proved below, which states that ν_n is stochastically equicontinuous in θ. The difficulty in establishing the required equicontinuity arises solely because the function g inside U is nonlinear in θ.

LEMMA SE. *Under the above assumptions, the process* $\nu_n(\gamma)$ *is stochastically equicontinuous, i.e., for all* $\epsilon > 0$ *and* $\eta > 0$, *there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} \Pr\left[\sup_{\rho(t_1,t_2)<\delta} |\nu_n(t_1) - \nu_n(t_2)| > \eta\right] < \epsilon.$$

PROOF OF LEMMA SE. By a second order Taylor series expansion of $g(Z_i,\theta)$ around $g(Z_i,\theta^0)$:

$$g(Z_i,\theta^0+\gamma n^{-1/2}) = g(Z_i,\theta^0) + \frac{1}{\sqrt{n}}\sum_{k=1}^{p}\frac{\partial g}{\partial\theta_k}(Z_i,\theta^0)\gamma_k + \frac{1}{n}\sum_{k=1}^{p}\sum_{r=1}^{p}\frac{\partial^2 g}{\partial\theta_k\partial\theta_r}(Z_i;\overline{\theta})\gamma_k\gamma_r \tag{18}$$

for some intermediate points $\overline{\theta}$. Define the linear approximation to $g(Z_i,\theta^0+\gamma n^{-1/2})$,

$$T(Z_i,\gamma) = g(Z_i,\theta^0) + \sum_{k=1}^{p}\frac{\partial g}{\partial\theta_k}(Z_i,\theta^0)\gamma_k$$

for any γ. By assumption C2, for all k,r, $\sup_{\theta\in\Theta}|\partial^2 g(Z_i,\theta)/\partial\theta_k\partial\theta_r|^2 \le d(Z_i)$ with $Ed(Z_i) < \infty$. Therefore, for all $\delta > 0$ there exists an $\varepsilon > 0$ such that

$$\begin{aligned}
\Pr\left[\frac{1}{\sqrt{n}}\max_{i,k,r}\sup_{\theta\in\Theta_n}\left|\frac{\partial^2 g}{\partial\theta_k\partial\theta_r}(Z_i,\theta)\right| > \varepsilon\right] &\le n\sum_{k,r}\Pr\left[\frac{1}{\sqrt{n}}\sup_{\theta\in\Theta_n}\left|\frac{\partial^2 g}{\partial\theta_k\partial\theta_r}(Z_i,\theta)\right| > \varepsilon\right] \\
&\le \frac{\sum_{k,r}E[d(Z_i)]}{\varepsilon^2} \\
&\le \delta
\end{aligned}$$

by the Bonferroni and Chebychev inequalities. Therefore, with probability tending to one

$$\max_{1\le i\le n}\left|\frac{1}{n}\sum_{k=1}^{p}\sum_{r=1}^{p}\frac{\partial^2 g}{\partial\theta_k\partial\theta_r}(Z_i;\overline{\theta})\gamma_k\gamma_r\right| \le \frac{\overline{\pi}}{\sqrt{n}}$$

for some $\overline{\pi} < \infty$. Define the stochastic process

$$\nu_{n1}(\gamma,\pi) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\overline{f}(Z_i,\theta_0+\gamma n^{-1/2},\pi n^{-1/2}) - E\overline{f}(Z_i,\theta_0+\gamma n^{-1/2},\pi n^{-1/2})$$

on $\gamma \in \Gamma$ and $\pi \in \Pi = [0, \overline{\pi}]$, where

$$\begin{aligned} & \overline{f}(Z_i, \theta_0 + \gamma n^{-1/2}, \pi n^{-1/2}) \\ = \quad & r[g(x, \theta_0 + \gamma n^{-1/2}), x] + \frac{r'[g(x, \theta_0 + \gamma n^{-1/2}) - U_i(\theta_0 + \gamma n^{-1/2}), x]}{\psi(U_i)}[Y_i - 1(T(Z_i, \gamma n^{-1/2}) + \frac{\pi}{\sqrt{n}} > 0)] \end{aligned}$$

It suffices to show that $\nu_{n1}(\gamma, \pi)$ is stochastically equicontinuous in γ, π, and the deterministic centering term is of smaller order. The latter argument is a standard Taylor expansion. The argument for $\nu_{n1}(\gamma, \pi)$ is very similar to that contained in Sherman (1993) because we basically have a linear index structure in this part. One can apply Lemma 2.12 in Pakes and Pollard (1989). ■

References

[1] AN, M.Y. (2000). "A Semiparametric Distribution for Willingness to Pay and Statistical Inference with Dichotomous Choice CV Data,"

[2] ANDREWS, D.W.K. (1994): "Asymptotics for Semiparametric Econometric Models by Stochastic Equicontinuity." Econometrica 62, 43-72.

[3] BICKEL, P.J., C.A.J. KLAASSEN, J. RITOV, AND J. WELLNER (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. Springer: Berlin.

[4] CHEN, H. AND A. RANDALL (1997): "Semi-nonparametric Estimation of Binary Response Models With an Application to Natural Resource Valuation, *Journal of Econometrics*, 76, 323-340.

[5] CREEL, M., AND J. LOOMIS (1997): "Semi-nonparametric Distribution-free Dichotomous Choice Contingent Valuation," *Journal of Environmental Economics and Management*, 32, 341-358.

[6] CROOKER, J.R., AND J.A. HERRIGES (2000): "Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework," Manuscript, Texas Tech University.

[7] DAS, M. (2002), "Minimum Distance Estimators for Nonparametric Models With Grouped Dependent Variables," unpublished manuscript.

[8] Gozalo, P., and O.B. Linton (1999): "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression." *Journal of Econometrics,* 99, 63-106.

[9] Hardle, W., and O.B. Linton (1994), "Applied Nonparametric Methods," *The Handbook of Econometrics*, vol. IV*,* eds. D.F. McFadden and R.F. Engle III. North Holland.

[10] Kanninen, B. (1993), "Dichotomous Choice Contingent Valuation," *Land Economics*, 69, 138-146.

[11] Klein, R. and R. H. Spady (1993), "An efficient Semiparametric Estimator for Binary Response Models," *Econometrica* 61, 387-421.

[12] Lewbel, A. (1997), "Semiparametric Estimation of Location and Other Discrete Choice Moments," *Econometric Theory,* 13, 32-51.

[13] Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics* 97, 145-177.

[14] Linton, O. and J.P. Nielsen (1995): "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika*, 82, 93-100.

[15] Manski, C. and E. Tamer, (2000), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

[16] Matzkin, R., (1992), "Non-parametric and Distribution-free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica* 60, 239-270.

[17] Masry, E. (1996a), "Multivariate local polynomial regression for time series: Uniform strong consistency and rates," *J. Time Ser. Anal.* 17, 571-599.

[18] Masry, E., (1996b), "Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and their Applications* 65, 81-101.

[19] McFadden, D. (1994): " Contingent Valuation and Social Choice," *American Journal of Agricultural Economics,* 76, 4.

[20] Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

[21] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, vol. iv, ed. by R. F. Engle and D. L. McFadden, pp. 2111-2245, Amsterdam: Elsevier.

[22] Pakes, A. and D. Pollard. (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-57.

[23] Sherman, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123-37.

[24] Silverman, B. (1986): *Density estimation for statistics and data analysis*. London, Chapman and Hall.

[25] Sperlich, S, O.B. Linton, and W. Härdle (1999) A Simulation comparison between the Backfitting and Integration methods of estimating Separable Nonparametric Models *TEST* 8, 419-458.

[26] Stone, C.J. (1982): "Optimal global rates of convergence for nonparametric regression." *Annals of Statistics* 8, 1040-1053.

[27] Tjostheim, D. and B. H. Auestad (1994): "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of the American Statistical Association*, 89, 1398-1409.

		$\sigma = 5$			$\sigma = 10$			$\sigma = 50$		
		n=100	n=300	n=500	n=100	n=300	n=500	n=100	n=300	n=500
PRMSE	$\widehat{\mu}_1$	3.97	2.73	2.27	4.01	2.73	2.30	3.97	2.67	2.23
	$\widehat{\mu}_3$	14.10	12.79	12.56	14.04	12.85	12.60	13.92	12.79	12.52
	$\widehat{\mu}_4$	7.00	3.69	2.66	6.98	3.67	2.65	6.91	3.71	2.65
	$\widehat{\mu}_5$	4.90	3.48	3.14	4.98	3.48	3.13	4.97	3.50	3.16
PMAE	$\widehat{\mu}_1$	3.15	2.21	1.82	3.18	2.19	1.84	3.18	2.14	1.79
	$\widehat{\mu}_3$	12.49	12.13	12.15	12.37	12.17	12.20	12.27	12.12	12.12
	$\widehat{\mu}_4$	5.72	3.00	2.15	5.72	2.99	2.14	5.62	3.02	2.14
	$\widehat{\mu}_5$	3.94	2.88	2.66	4.01	2.87	2.65	3.99	2.89	2.68
IRMSE	$\widehat{\mu}_1$	14.56	12.63	12.21	14.55	12.64	12.20	14.59	12.59	12.17
	$\widehat{\mu}_3$	17.59	16.04	15.74	17.55	16.07	15.79	17.57	16.10	15.79
	$\widehat{\mu}_4$	12.74	10.41	9.89	12.73	10.38	9.90	12.85	10.51	10.00
	$\widehat{\mu}_5$	11.74	10.33	10.02	11.76	10.31	10.03	11.92	10.43	10.14
IMAE	$\widehat{\mu}_1$	10.96	10.16	10.05	10.95	10.15	10.04	10.98	10.08	9.97
	$\widehat{\mu}_3$	14.42	13.36	13.13	14.35	13.38	13.18	14.34	13.40	13.16
	$\widehat{\mu}_4$	10.08	8.50	8.22	10.06	8.48	8.23	10.17	8.59	8.32
	$\widehat{\mu}_5$	9.23	8.44	8.31	9.25	8.42	8.32	9.38	8.53	8.42

Table 1: Conditional Mean in 5-bid design; 10,000 replications; Bandwidth by Silverman's Thumb; Pointwise Root Mean Squared and Absolute Errors (PRMSE and PMAE) calculated at $x = 0$.

		$\sigma = 5$			$\sigma = 10$			$\sigma = 50$		
		n=100	n=300	n=500	n=100	n=300	n=500	n=100	n=300	n=500
PRMSE	$\widehat{\mu}_1$	7.90	3.07	1.24	4.12	2.98	4.50	37.49	41.47	43.19
	$\widehat{\mu}_3$	11.18	6.43	5.40	12.09	9.99	9.86	45.97	48.60	49.37
	$\widehat{\mu}_4$	9.91	7.47	6.85	11.10	9.70	9.30	45.90	46.78	46.75
	$\widehat{\mu}_5$	28.14	26.41	26.03	24.18	22.07	21.33	28.74	24.32	22.50
PMAE	$\widehat{\mu}_1$	7.10	2.61	0.94	3.26	2.61	4.36	37.29	41.42	43.16
	$\widehat{\mu}_3$	8.12	5.65	5.17	10.97	9.76	9.75	44.62	48.29	49.25
	$\widehat{\mu}_4$	7.59	6.34	5.98	10.33	9.26	8.84	44.78	46.19	46.24
	$\widehat{\mu}_5$	24.26	24.38	24.65	21.12	20.17	19.92	23.56	21.38	20.59
IRMSE	$\widehat{\mu}_1$	17.37	17.49	17.42	13.11	13.24	13.24	30.48	30.14	30.30
	$\widehat{\mu}_3$	11.18	6.43	5.40	12.09	9.99	9.86	45.97	48.60	49.37
	$\widehat{\mu}_4$	9.91	7.47	6.85	11.10	9.70	9.30	45.90	46.78	46.75
	$\widehat{\mu}_5$	21.93	20.17	19.75	18.66	16.69	16.08	33.37	32.01	31.66
IMAE	$\widehat{\mu}_1$	15.74	15.71	15.46	11.39	11.47	11.57	29.40	29.13	29.25
	$\widehat{\mu}_3$	8.12	5.65	5.17	10.97	9.76	9.75	44.62	48.29	49.25
	$\widehat{\mu}_4$	7.59	6.34	5.98	10.33	9.26	8.84	44.78	46.19	46.24
	$\widehat{\mu}_5$	18.05	17.50	17.39	15.83	14.71	14.38	29.72	29.33	29.25

Table 2: Conditional Standard Deviation in 5-bid design; 10,000 replications; Bandwidth by Silverman's Thumb

		$\sigma = 5$			$\sigma = 10$			$\sigma = 50$		
		n=100	n=300	n=500	n=100	n=300	n=500	n=100	n=300	n=500
PRMSE	$\widehat{\mu}_1$	5.59	3.32	2.60	6.19	3.80	3.05	11.30	7.34	5.92
	$\widehat{\mu}_3$	10.65	7.05	6.28	10.42	7.11	6.24	12.42	9.63	8.90
	$\widehat{\mu}_4$	6.27	3.33	2.51	6.42	3.45	2.67	9.12	5.11	3.96
	$\widehat{\mu}_5$	5.78	3.35	2.61	5.88	3.39	2.60	7.56	4.38	3.42
PMAE	$\widehat{\mu}_1$	4.40	2.63	2.06	4.91	3.01	2.43	9.00	5.83	4.72
	$\widehat{\mu}_3$	8.56	5.73	5.22	8.35	5.80	5.19	10.16	8.30	7.93
	$\widehat{\mu}_4$	5.02	2.68	2.01	5.13	2.75	2.14	7.26	4.09	3.16
	$\widehat{\mu}_5$	4.59	2.67	2.08	4.70	2.71	2.08	6.04	3.49	2.72
IRMSE	$\widehat{\mu}_1$	12.30	7.54	6.02	12.37	7.57	6.05	15.76	11.12	9.82
	$\widehat{\mu}_3$	12.65	8.05	6.98	12.46	8.10	6.94	15.90	12.38	11.55
	$\widehat{\mu}_4$	9.22	5.12	3.93	9.35	5.19	4.03	13.37	9.24	8.32
	$\widehat{\mu}_5$	8.91	5.13	4.00	9.00	5.15	3.99	12.31	8.85	8.07
IMAE	$\widehat{\mu}_1$	8.81	5.41	4.33	8.96	5.47	4.40	12.05	8.51	7.54
	$\widehat{\mu}_3$	9.99	6.46	5.70	9.80	6.52	5.66	12.35	9.80	9.24
	$\widehat{\mu}_4$	7.14	3.95	3.02	7.24	4.01	3.12	10.40	7.23	6.57
	$\widehat{\mu}_5$	6.87	3.97	3.08	6.95	3.98	3.09	9.53	6.90	6.36

Table 3: Conditional Mean in Continuous design; 10,000 replications; Bandwidth by Silverman's Thumb

		$\sigma = 5$			$\sigma = 10$			$\sigma = 50$		
		n=100	n=300	n=500	n=100	n=300	n=500	n=100	n=300	n=500
PRMSE	$\widehat{\mu}_1$	9.59	7.43	6.49	7.49	5.43	4.55	10.74	7.91	7.17
	$\widehat{\mu}_3$	40.77	41.70	41.96	37.21	37.90	38.04	18.75	18.18	18.08
	$\widehat{\mu}_4$	18.37	12.46	10.44	16.99	12.00	10.35	15.96	8.70	6.92
	$\widehat{\mu}_5$	20.38	15.14	13.32	19.07	14.52	12.72	24.20	15.84	12.30
PMAE	$\widehat{\mu}_1$	8.86	7.10	6.25	6.41	4.79	4.05	8.68	6.65	6.16
	$\widehat{\mu}_3$	38.72	41.16	41.66	35.33	37.34	37.73	17.15	17.58	17.73
	$\widehat{\mu}_4$	14.01	9.95	8.54	14.39	10.62	9.25	11.89	6.72	5.49
	$\widehat{\mu}_5$	15.15	11.68	10.55	15.88	12.62	11.21	18.37	11.54	9.12
IRMSE	$\widehat{\mu}_1$	12.13	10.00	9.08	10.98	8.98	8.08	18.80	14.35	12.92
	$\widehat{\mu}_3$	40.77	41.70	41.96	37.21	37.90	38.04	18.75	18.18	18.08
	$\widehat{\mu}_4$	18.37	12.46	10.44	16.99	12.00	10.35	15.96	8.70	6.92
	$\widehat{\mu}_5$	18.51	14.00	12.20	17.54	13.36	11.82	24.18	18.16	16.00
IMAE	$\widehat{\mu}_1$	10.05	8.43	7.69	9.35	7.59	6.77	14.06	11.05	10.33
	$\widehat{\mu}_3$	38.72	41.16	41.66	35.33	37.34	37.73	17.15	17.58	17.73
	$\widehat{\mu}_4$	14.01	9.95	8.54	14.39	10.62	9.25	11.89	6.72	5.49
	$\widehat{\mu}_5$	13.34	10.53	9.43	14.20	11.28	10.08	18.32	13.42	11.95

Table 4: Conditional Standard Deviation in Continuous design; 10,000 replications; Bandwidth by Silverman's Thumb