ESTIMATING FEATURES OF A DISTRIBUTION FROM BINOMIAL DATA

Arthur Lewbel and Daniel McFadden

May 17, 1997

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. One application is destructive duration analysis, where the process is censored at an observation test time. Another 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 an endangered species. Sampled consumers are asked whether they would vote for a referendum that would provide the good at a cost specified by experimental design. This paper provides practical estimators for moments and quantiles of the unknown distribution in this problem. Under mild regularity conditions and a randomized design for thresholds, we show that the moments estimators are root-N consistent and asymptotically normal, despite the limited information in binomial response, while quantile estimators converge at a lower rate equal to the optimal rate for nonparametric regression estimation of the distribution of responses.

ACKNOWLEDGEMENTS: The first author is from the Department of Economics, Brandeis University, and acknowledges research support from the National Science Foundation through grants SER-9514977 and SES-9210749. The second author is from the Department of Economics, University of California, Berkeley, and acknowledges research support from the E. Morris Cox Endowment. Correspondence should be directed to Daniel McFadden, Department of Economics, University of California, Berkeley CA 94720-3880, e-mail: mcfadden@econ.berkeley.edu, internet: http://elsa.berkeley.edu/~mcfadden. Arthur Lewbel has moved to Department of Economics, Boston College.

lewb@bc.edu http://www2.bc.edu/˜lewbel/

KEYWORDS: binomial response, duration analysis, contingent valuation, semiparametric regression.
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. In bioassay, one is interested in the distribution of survival times until onset of an abnormality in laboratory animals. The animals are sacrificed at times determined by experimental design, and tested for presence of the abnormality. An observation consists of a test time, an indicator for the event that the abnormality is absent at the test time, and a vector of covariates. This setup occurs in any non-retrospective destructive duration analysis where the observation process alters subsequent behavior and time of onset cannot be determined retrospectively, as in destructive testing of materials.

A similar problem arises 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. Sampled consumers are asked whether they would vote for a referendum that would provide the good at a cost specified by experimental design. This method is termed referendum contingent valuation. An observation consists of a cost threshold, or bid, a binomial indicator for whether willingness-to-pay (WTP) exceeds the bid, and a vector of covariates. There is evidence that presentation of a bid alters subsequent responses, so the problem resembles destructive duration analysis.¹

We assume that the unknown distribution in this problem depends on covariates through a single-index location shift. When the experimental design is randomized

¹ McFadden (1994) provides references and new evidence that responses to follow up bids are biased. There are additional issues of reliability of contingent valuation survey responses, particularly context effects, anchoring to bids, and consistency with the assumptions on consumer behavior needed to translate stated values into public policy; see Green et al (1997). Because of these effects, direct open-ended elicitations of willingness to pay, analyzed by robust methods that handle non-response and outliers, may be as reliable as referendum methods. This paper will ignore these issues, and provide statistical methods that are valid when binomial responses are reliable indicators for the underlying variable of interest.
with a strictly positive bid density, and some tail conditions are imposed, we find practical Root-N Consistent Asymptotically Normal (RCAN) estimators for the location shift parameters and conditional moments. Conditional quantiles can be estimated at an optimal rate, less than Root-N, that is associated with nonparametric regression estimation of the distribution of responses.

To set the problem formally, consider a linear model

\[ W = x\beta - \nu, \]

where \( x \) is a finite-dimensional vector of covariates, \( \nu \) is a mean-zero disturbance independent of \( x \) with an unknown CDF \( G(\nu) \). Consider a random variable \( B \) with a known conditional density \( h(b|\nu|x) \) given \( x \). Define an indicator for the event that \( W \) exceeds \( B \), \( Y = 1(W > B) \). Then, \( W \) is the random variable whose conditional distribution, given \( x \), is the target of the analysis, \( B \) is the bid drawn from the density \( h(b|\nu|x) \) specified by the experimental design, and \( Y \) is an indicator for the event that \( W \) exceeds the bid. For positive variables such as durations, interpret \( W \) and \( B \) as logs of the variables of interest.\(^2\) Throughout, we make the following assumptions:

(i) The covariates \( x \) have a compact support \( X \), and \( Ex'x \) is of full rank.
(ii) The parameter vector \( \beta \) lies in the interior of a compact set.
(iii) The CDF \( G(\nu) \) has mean zero and support \([-\alpha_0, \alpha_1]\) with \( \alpha_0, \alpha_1 \leq +\infty \), and for some \( \lambda_0 > 0 \) satisfies \( E_0 e^{\lambda_0|\nu|} < +\infty \).
(iv) The design bid \( B \) has a density \( h(b|x) \) that is bounded, positive and continuously differentiable on \( \{(b,x) \mid -\alpha_0 + x\beta < b < \alpha_1 + x\beta \ & \ x \in X\} \); and satisfies the tail conditions that for a small positive constant \( C_0 \),

\[ h(b|x) \geq C_0 e^{-|b|\lambda_0/3} \text{ for } x\beta - \alpha_0 < b < x\beta + \alpha_1. \]

The density \( h(b|x) \) is dominated by a square-integrable function \( \bar{h}(b) \).
(v) A random sample \( z_i = (x_i, b_i, y_i) \) for \( i = 1, \ldots, N \) is observed, where \( b_i \) is a realization of \( B \) and \( y_i \) is a realization of \( Y = 1(W > B) \).

---

\(^2\) Define \( F(s) = G(-\log(s)) \). Then \( F(te^{-x\beta}) \) is the complementary CDF of \( T = e^W \) given \( x \). In duration analysis, \( F(te^{-x\beta}) \) is the survival curve and \( e^{-x\beta} \) is a proportional hazards term.
Commenting on (III), the tail condition $E_u e^{\lambda_0 u} < +\infty$ implies that $u$ has a proper moment generating function; implies the exponential bounds

$$\lim_{u \to +\infty} e^{\lambda_0 u} G(u) = 0 \quad \text{and} \quad \lim_{u \to +\infty} e^{\lambda_0 u}[1 - G(u)] = 0 ;$$

and using integration by parts, implies the condition

$$\int_{-\infty}^{+\infty} e^{\lambda_0 u} \cdot |G'(u)| \cdot du < +\infty .$$

The existence of moments $E_u e^{ku}$ and $E_u e^{-ku}$ for an integer $k > 0$ implies this tail condition with $\lambda_0 \geq k$. The tail condition is met automatically for all $\lambda_0 > 0$ if the support of $u$ is bounded.

Commenting on (IV), the tail condition is automatically satisfied when $h$ has a compact support that contains the support of $W$. Note that (IV) is never vacuous when (III) holds, since the bilateral exponential density $h(b|x) = (1/\theta) \lambda_0 e^{-\lambda_0 |b|/\theta}$ satisfies the tail condition. The condition that $h(b|x)$ be a positive density requires a randomized experimental design. A design with a finite number of treatments that specify fixed bid levels will violate this condition. For the purpose of moment estimation, it may be possible to weaken this design requirement if there are continuous covariates; what is needed is that $xb - b$ have a positive density. When covariates are absent or discrete, a finite experimental design will not permit consistent estimation of moments of $W$ when $G$ is unknown.

A function $r(w,x)$ will be said to satisfy condition [R] if it is continuous in $(w,x)$ and continuously differentiable in $w$ for each $x$, and satisfies

$$\lim_{u \to +\infty} r(xb - u, x) e^{-\lambda_0 |u|} = 0$$

and

$$\sup_{-\alpha_0 < u < \alpha_1} \sup_{x} |r'(xb - u, x)| e^{-\lambda_0 |u|/3} < +\infty ,$$

where $\lambda_0$ is given by Assumption (III) and $r'$ denotes the derivative of $r$ with respect to its first argument.\(^3\) When condition [R] holds, the conditional expectation of

\(^3\) The first condition can be satisfied without loss of generality when $\alpha_0$ and/or $\alpha_1$
The conditional expectation \( E_{W|X} r(W,X) \) gives the expression

\[
E_{W|X} r(W,X) = r(x\beta, x) + \int_{-\infty}^{\infty} r(x\beta - u, x) [G(u) - 1(u > 0)] du .
\]

Let \( \mu = \mu(x, \beta, r) \) be a notational shorthand for \( E_{W|X} r(W,X) \). Leading cases are moments \( r(W,X) = W^k \) for \( k = 1, 2, ..., \) and when \( W \) is the log of the variable of interest, moments \( r(W,X) = e^{kW} \). In addition to moments, we are interested in conditional quantiles: the \( q \)-th quantile of \( W \) given \( X \) satisfies \( W_q = x\beta - G^{-1}(1-q) \). Section 2 discusses estimation of \( \beta \). Section 3 discusses conditional moment estimators for \( W \) when \( \beta \) is known. Section 4 modifies these estimators to use a "plug in" estimator of \( \beta \). Section 5 discusses conditional quantile estimators. Section 6 presents some Monte Carlo evidence. Section 7 discusses extensions.

2. ESTIMATION OF \( \beta \)

Given \( X \) and a realization \( b \) of \( B \), the indicator \( Y \) satisfies \( Y = 1(W > b) = 1(x\beta - b > 0) \), and its data generation process can be written

\[
P(Y=1 \mid X, b) = G(x\beta - b) .
\]

One can also write

\[
y = G(x\beta - b) + \varepsilon ,
\]

where \( G \) is a CDF and \( \varepsilon \) is mean-zero binomial disturbance. Then, (4) has the form of a standard semiparametric estimation problem; see Powell (1994). Lewbel (1997) has provided an easily computed estimator for \( \beta \) in model (4),

\[
\hat{\beta} = \left[ \sum_{i=1}^{N} x_i x_i \right]^{-1} \sum_{i=1}^{N} x_i [y_i - 1(b_i < 0)] / h(b_i \mid x_i)
\]

Since \( h \) is known, this is a simple regression that requires no iteration or

are finite by truncating \( r \) in tails that extend beyond the support of \( W \).
nonparametric elements such as choice of bandwidth. The statistical properties of
this estimator are given in the following result:

**Theorem 1.** Suppose assumptions (I)-(V) hold. Then \( \hat{\beta} \) is an unbiased RCAN estimator of \( \beta \), with \( \text{acov}(\sqrt{N}(\hat{\beta} - \beta)) = H^{-1} \cdot J \cdot H^{-1} \), where

\[
H = \text{plim} \frac{1}{N} \sum_{i=1}^{N} x_i x_i' \quad \text{and} \quad J = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{x_i' [y_i - 1(b_i < 0)]}{h(b_i \mid x_i)} - x_i' \hat{\beta} \right] .
\]

Proof: Assumption (I) implies that \( \frac{1}{N} \sum_{i=1}^{N} x_i x_i' \xrightarrow{\text{as}} \mathbb{E} x'x = H \) positive definite.

Consider the random variable \( \frac{Y - 1(b < 0)}{h(b \mid x)} = \frac{1(x \beta - b > u)}{h(b \mid x)} - \frac{1(b < 0)}{h(b \mid x)} \equiv z(x, b, u) \). Assumptions (I)-(IV) guarantee that the conditional expectation of \( z \) exists, with

\[
E_{b, u} z(x, b, u) = E_{v} \int_{-\infty}^{\infty} \left[ 1(x \beta - b > u) - 1(b < 0) \right] db = E_{v} \left\{ \begin{array}{ll}
\int_{0}^{\infty} 1(x \beta - b > u) db & \text{if } \int_{-\infty}^{0} 1(x \beta - b < u) db = 0 \\
\int_{-\infty}^{0} 1(x \beta - b < u) db & \text{otherwise} \end{array} \right.
\]

\[
= E_{v} \{ \max(x \beta - u, 0) - \min(x \beta - u, 0) \} = E_{v} \{ x \beta - u \} = x \beta .
\]

Then \( \hat{\beta} \) is an unbiased estimator of \( \beta \). The second moment of \( z(x, b, u) \) is

\[
E_{b, u} z(x, b, u)^2 = E_{v} \left\{ \begin{array}{ll}
\int_{0}^{\infty} 1(x \beta - b > u) h(b \mid x) db & \text{if } \int_{-\infty}^{0} 1(x \beta - b < u) h(b \mid x) db = 0 \\
\int_{-\infty}^{0} 1(x \beta - b < u) h(b \mid x) db & \text{otherwise} \end{array} \right.
\]

\[
= \int_{0}^{\infty} \frac{G(x \beta - b) db}{h(b \mid x)} + \int_{-\infty}^{0} \frac{1 - G(x \beta - b) db}{h(b \mid x)} .
\]

But (IV) implies for \( b > 0 \) that
\[
G(x^{-b}|h(b|x)) \leq \left\{ \sup_{b>0} e^{\lambda_0(b-x^{-b})} G(x^{-b}) \right\} e^{\lambda_0 x^{-b}} e^{-b / 2\lambda_0 / 3} / C_0 .
\]

The term in brackets is bounded, by (III), and hence \(G(x^{-b}|h(b|x))\) is integrable over \(0 < b < +\infty\). Similarly, (IV) implies \([1 - G(x^{-b})] / h(b|x)\) is integrable over \(-\infty < b < 0\). Therefore, \(E_{x,b,v} x(x,b,v)^2 < +\infty\). These arguments establish that the expression

\[
\frac{1}{N} \sum_{i=1}^N \frac{[x_i(y_i - 1(b_i<0))]}{h(b_i|x_i)} - x_i^{x_i\beta} .
\]

is a normalized sum of i.i.d. random variables that have mean zero and finite variance. The Lindeberg-Levy CLT then implies the expression is asymptotically normal, with covariance matrix \(J\). Then, the linear transformation (5) is asymptotically normal with \(acov(\sqrt{N}(\hat{\beta} - \beta)) = H^{-1}JH^{-1}\).

Lewbel (1997, Theorems 1,2) obtains a more general result that with the addition of trimming factors provides a RCAN estimator of \(\beta\) when \(h(b|x)\) is unknown and is replaced by a kernel estimator, there is measurement error or random coefficients, and the tail conditions are weakened. Other semiparametric estimators for \(\beta\) are available under Assumptions I-V, which ensure that \(y = 1(x^{-b}>v)\), with \(b\) having a continuous density and \(v\) being independent of \(x\) with zero mean; see Cossette (1983), Ichimura (1992), and Klein and Spady (1992). In particular, the Klein-Spady estimator, which maximizes a quasi-likelihood function in \(\beta\), with a kernel estimator used for the unknown CDF \(G(\cdot)\), and appropriate trimming, satisfies the semiparametric efficiency bound for this problem derived by Chamberlain (1986) and Cossette (1987).

The efficiency of the Lewbel estimator will depend on the design density \(h\). For a distribution \(G\) whose tails behave like \(e^{-\lambda |v|}\), a design density with tails that behave like \(e^{-\lambda |\beta| / 2}\) will be relatively efficient. In many applications, the computational advantages of the Lewbel estimator will outweigh its efficiency loss relative to the semiparametric bound. Alternately, one can start from the Lewbel estimator and obtain a fully efficient estimator in one Newton-Raphson iteration of the Klein-Spady quasi-likelihood function.
3. MOMENT ESTIMATION WITH KNOWN $\beta$

Consider the problem of estimating $E_W|X_r(W,x)$ when $\beta$ is known and $G$ is unknown. This case obviously applies when there are no covariates. More generally, we show in Section 4 that the $\beta$ vector appearing in the estimators in this section can be replaced by $\hat{\beta}$ from Theorem 1, or any other RCAN estimator.

Let $v_i = x_i^p b_i$. Let $\psi(v) = E_{x_i} h(x_i^p - v|x_i)$ denote the unconditional density of $V = x_i^p - V$. Note that $\psi(v)$ inherits the properties of $h(b|x)$ from Assumption (IV). In particular, (I) and (IV) imply $\psi(v) \geq C_1 e^{-|v|/\lambda_0/3}$ for some small positive constant $C_1$. An estimator of $\psi$ is

$$
\psi_N(v) = \frac{1}{N} \sum_{i=1}^{N} h(x_i^p - v|x_i)
$$

(7)

A preliminary result gives properties of $\psi_N(v)$ that will be used later:

**Lemma 1.** If Assumptions I and IV, then $E \psi_N(v) = \psi(v)$; for any sequence $v_1, v_2, \ldots$, one has $\sup_{i \leq N} N^{1/4} |\psi_N(v_i) - \psi(v_i)| \xrightarrow{a.s.} 0$; and $\psi_N(v) \geq C_2 e^{-|v|/\lambda_0/3}$ for some small positive constant $C_2$.

Proof: Hoeffding's inequality implies $P(N^{1/4} |\psi_N(v_i) - \psi(v_i)| > \epsilon) < 2 \exp(-2N\epsilon^2/M^2)$, where $M$ is a bound on $h$ from Assumption (IV). Then,

$$
\sum_{N=N_0}^{\infty} P(\sup_{i \leq N} N^{1/4} |\psi_N(v_i) - \psi(v_i)| > \epsilon) < \sum_{N=N_0}^{\infty} 2N \exp(-2N\epsilon^2/M^2).
$$

The right-hand-side of this inequality converges to zero as $N_0$ increases, establishing almost sure convergence. The tail condition (IV) applied to each term in the sum defining $\psi_N$ establishes that $\psi_N$ satisfies the stated tail condition. ■

Consider a function $r(W|x)$ that satisfies condition $[R]$, so that the conditional expectation $\mu(x_0, p, r) = E_W|x_0 r(W,x_0)$ exists. The following estimator for this moment is motivated by (2), and extends an estimator of the mean of $e^W$ from discrete response data proposed independently by McFadden (1994) for the case without
covariates and by Lewbel (1997) for the case with covariates:

\[ \mu_N(x_0, \beta, \eta) = r(x_0 \beta, x_0) + \frac{1}{N} \sum_{i=1}^{N} r'(x_0 \beta - \nu_i x_0) \frac{y_i - 1(\nu_i > 0)}{\psi_N(\nu_i)} \]

This estimator requires no iteration, is simple to compute, and converges to \( \mu(x_0, \beta, \eta) \) at a \( \sqrt{N} \) rate. The precision of the estimator depends on the experimental design. In the absence of covariates, the asymptotically efficient design has

\[ \psi(v) = C_3 \left| r(-v) \right| \left| G(v) - 1(v > 0) \right| \]

where \( C_3 \) is a normalizing constant. The statistical properties of this estimator are given by the following result:

**Theorem 2.** Suppose Assumptions (I)-(V) and Condition [R]. Then \( \mu_N(x_0, \beta, \eta) \) converges almost surely to \( \mu(x_0, \beta, \eta) \), and

\[ \sqrt{N}(\mu_N(x_0, \beta, \eta) - \mu(x_0, \beta, \eta)) \rightarrow N(0, \sigma^2) \]

with \( \sigma^2 \) the probability limit of the estimator (9).

Proof: We write the statistic as a U-statistic plus asymptotically negligible terms, and then appeal to the large sample theory of U-statistics. We start with a several bounds. For brevity, write \( a_i = r'(x_0 \beta - \nu_i x_0) \frac{y_i - 1(\nu_i > 0)}{\psi(\nu_i)} \).

By condition [R],

\[ E|a_i| \leq \int_{-\infty}^{\infty} |e^{-\lambda_0 |v|} r(x_0 \beta - v x_0) e^{\lambda_0 |v|} |G(v) - 1(v > 0)| dv \]

\[ \leq C_4 \int_{-\infty}^{\infty} e^{\lambda_0 |v|} |G(v) - 1(v > 0)| dv \]

where \( C_4 \) is a large constant. The last integral is finite by (III). Hence,

\[ E_a = \int r'(x_0 \beta - v x_0) [G(v) - 1(v > 0)] dv \] exists. Write \( z_i = (x_i \nu_i y_i) \) and define
Commenting on (III), the tail condition $E_v e^{-\lambda_0 |v|} < +\infty$ implies that $v$ has a proper moment generating function; implies the exponential bounds

$$\lim_{v \to \infty} e^{-\lambda_0 v} G(v) = 0 \quad \text{and} \quad \lim_{v \to \infty} e^{-\lambda_0 v} [1 - G(v)] = 0 ;$$

and using integration by parts, implies the condition

$$\int_{-\infty}^{+\infty} e^{-\lambda_0 |v|} G(v) - 1_{(v>0)} |v| \, dv < +\infty .$$

The existence of moments $E_v e^{kv}$ and $E_v e^{-kv}$ for an integer $k > 0$ implies this tail condition with $\lambda_0 \geq k$. The tail condition is met automatically for all $\lambda_0 > 0$ if the support of $v$ is bounded.

Commenting on (IV), the tail condition is automatically satisfied when $h$ has a compact support that contains the support of $W$. Note that (IV) is never vacuous when (III) holds, since the bilateral exponential density $h(b|x) = (1/\theta) e^{-\lambda_0 |b|/\theta}$ satisfies the tail condition. The condition that $h(b|x)$ be a positive density requires a randomized experimental design. A design with a finite number of treatments that specify fixed bid levels will violate this condition. For the purpose of moment estimation, it may be possible to weaken this design requirement if there are continuous covariates; what is needed is that $x\beta - b$ have a positive density. When covariates are absent or discrete, a finite experimental design will not permit consistent estimation of moments of $W$ when $G$ is unknown.

A function $r(w,x)$ will be said to satisfy condition $[R]$ if it is continuous in $(w,x)$ and continuously differentiable in $w$ for each $x$, and satisfies

$$\lim_{v \to \infty} r(x \beta - v, x) e^{-\lambda_0 |v|} = 0$$

and

$$\sup_{-\alpha_0 < v < \alpha_1} \sup_x \left| r(x \beta - v, x) e^{-\lambda_0 |v|/3} \right| < +\infty ,$$

where $\lambda_0$ is given by Assumption (III) and $r'$ denotes the derivative of $r$ with respect to its first argument.\(^3\) When condition $[R]$ holds, the conditional expectation of $W$...\(^3\)

\(^3\) The first condition can be satisfied without loss of generality when $\alpha_0$ and/or $\alpha_1$...
\[ \mu_N^r(x_0, \beta, r) - \mu(x_0, \beta, r) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j \neq i} u(z_i, z_j) + \frac{1}{N} \sum_{i=1}^{N} u(z_i, z_i) \]

\[ + \frac{1}{N} \sum_{i=1}^{N} \frac{(\psi_N(\nu_i) - \psi(\nu_i))^2}{\psi(\nu_i) \cdot \psi_N(\nu_i)} . \]

The term A is a U-statistic, with a symmetric kernel \( u(z_i, z_j) \) having mean zero and a positive finite conditional variance \( \sigma_i^2 \). Then, \( A \overset{as}{\to} 0 \) and \( \sqrt{N} A \overset{d}{\to} N(0, \sigma_i^2) \); see Serfling (1980, p. 190, Theorem A and p. 192, Theorem A).

Consider the remaining terms. Since \( \mathbf{E} u(z_i, z_i) \) exists, \( \frac{1}{N} \sum_{i=1}^{N} u(z_i, z_i) \) converges almost surely to a constant, implying that \( \sqrt{N} \mathbf{B} \overset{as}{\to} 0 \). Finally,

\[ \sqrt{N} |C| \leq \left( \frac{1}{N} \sum_{i=1}^{N} \frac{|a_i|}{|\psi(\nu_i) \cdot \psi_N(\nu_i)|} \sup_{i \leq N} N^{1/4} |\psi_N(\nu_i) - \psi(\nu_i)| \right)^2 \]

\[ \leq \left( \frac{1}{N} \sum_{i=1}^{N} \frac{|a_i| \cdot |\nu_i| \cdot 2\lambda \cdot \sqrt{3}}{C_0 \cdot C_2} \sup_{i \leq N} N^{1/4} |\psi_N(\nu_i) - \psi(\nu_i)| \right)^2 . \]

The first term in the last line of this expression is an average of i.i.d. random variables whose expectation we have proved to be finite. Then, a SLLN implies this term converges almost surely to a constant. Lemma 1 establishes that the squared term in the last line converges almost surely to zero. Therefore, the product converges to zero; that is, \( \sqrt{N} C \overset{as}{\to} 0 \).

An estimate of \( \sigma_i^2 \) is obtained by plugging empirical analogs into its formula; define
(9) \[ \hat{\sigma}_1^2 = \frac{1}{N} \sum_{j=1}^{N} \hat{\xi}_j^2 , \]

with \( \hat{a}_i = r(x_{0\beta} - y_i, x_o) \frac{y_i \cdot 1(y_i > 0)}{\psi_N(y_i)} \) and

\[ \hat{\xi}_i = \frac{\hat{a}_i \cdot E_N \hat{\sigma}_1}{2} \cdot \sum_{j \neq i} \frac{r(x_{0\beta} - y_j, x_o) \cdot [y_j \cdot 1(y_j > 0)] \cdot (h(x_{0\beta} - y_j, x_o) - \psi_N(y_j))}{2(N-1) \cdot \psi_N(y_j)} . \]

An alternative approach to estimating \( \mu(x, \beta, r) \) uses the ordered observations and avoids the need to obtain a preliminary estimate of \( \psi \). Assume now that the support of \( v \) is bounded, so that \( \alpha_0 \) and \( \alpha_1 \) are finite.\(^4\) Augment the observed \((v, y)\) pairs with the artificial observations \((-\alpha_0, 0)\) and \((\alpha_1, 1)\). Recode each observation \((v, y)\) with \( v < -\alpha_0 \) as \((-\alpha_0, 0)\), and each observation with \( v > \alpha_1 \) as \((\alpha_1, 1)\). Then, index the observations so that the \( v \)'s, including the artificial ones, are in non-decreasing order, and denote them by \( v_{N0} \leq ... \leq v_{Ni, Ni+1} \). The probability of ties in the interior of the support is zero. Let \( y_{Ni} \) denote the observed \( y \) associated with \( v_{Ni} \). Then

(10) \[ \mu(x_{0\beta}, x_o) = r(x_{0\beta}, x_o) + \int_{-\infty}^{v_{Ni, Ni+1}} r(x_{0\beta} - v, x_o) \cdot [G(v) - 1(v > 0)] dv \]

\[ = r(x_{0\beta}, x_o) + \frac{1}{2} \sum_{i=1}^{N} \int_{v_{Ni, Ni+1}}^{v_{Ni, i-1}} r(x_{0\beta} - v, x_o) \cdot [G(v) - 1(v > 0)] dv \]

\[ = r(x_{0\beta}, x_o) + \sum_{i=1}^{N} \int_{v_{Ni, i-1}}^{v_{Ni, i+1}} r(x_{0\beta} - v, x_o) \cdot [G(v_{Ni}) - 1(v_{Ni} > 0)] \cdot \frac{v_{Ni, i+1} - v_{Ni, i-1}}{2} , \]

\(^4\) The analysis below can be extended to the case of unbounded support for \( v \) by trimming the observations to a finite interval that expands with sample size.
where the \( \tilde{v}_{Ni} \) are intermediate between \( v_{N,i-1} \) and \( v_{N,i} \). Replacing the \( \tilde{v}_{Ni} \) by \( v_{Ni} \) corresponds to the extended trapezoid rule for numerical integration, with its attendant approximation properties. This suggests the following estimator, obtained by replacing the unknown \( G(v_{Ni}) \) by \( y_{Ni} \) in the last formula:

\[
 \mu_N(x_o, \beta, r) = r(x_o \beta, x_o) + \sum_{i=1}^{N} r(x_0 \beta - v_{Ni} x_o)[y_{Ni} - 1(v_{Ni} > 0)] \frac{v_{N,i+1} - v_{N,i}}{2}
\]

Before stating the statistical properties of this estimator, we give a preliminary lemma that establishes some properties of differences of order statistics:

**Lemma 2.** Suppose the support of a random variable \( V \) is a compact interval, and that on the interior of this interval, \( V \) has a continuously differentiable density \( \psi \) that is bounded below by \( \gamma > 0 \). Given \( v \) in the interior of the support, and a random sample of size \( N-1 \), let \( L \) be the length of the longest interval containing \( v \) and no point from the random sample. Define \( Q_{jN}(v) = E((N-L)^j | v) \). Then, \( Q_{jN}(v) \leq (j+1)! \gamma^j / \gamma \) and \( Q_{jN}(v) = (j+1)! / \gamma(v)^j + O((\log(N))^2 / N) \). Also, for a universal constant \( C_\gamma \),

\[
 \text{Prob} ( \max_{1 \leq i \leq N} (v_{N,i+1} - v_{N,i-1}) > c ) \leq C_\gamma \exp(-N \gamma^2 c^2 / 8).
\]

Proof: Let \([v_o, v_i]\) denote the support of \( V \), and let \( A = v - v_o \) and \( B = v_i - v \). The probability that an interval \((v-a, v+b)\) contains no element of the random sample is \( G(a,b) = (1 - \Psi(v+b) + \Psi(v-a))^N \). Integration by parts applied to the expectation of a twice continuously differentiable non-negative \( f(a,b) \) yields
\[
\int_0^A \int_0^B f(a,b) G_{ab}(a,b) \, da \, db = f(A,B) G(A,B) - f(0,B) G(0,B) - f(A,0) G(A,0) \\
+ f(0,0) G(0,0) - \int_0^A \left[ f_a(a,B) G(a,B) - f_a(a,0) G(a,0) \right] da \\
- \int_0^B \left[ f_b(A,b) G(A,b) - f_b(0,b) G(0,b) \right] db + \int_0^A \int_0^B f_{ab}(a,b) G(a,b) \, da \, db.
\]

Applying this formula to the function \( f(a,b) = N_j^i (a+b)^j \), dropping negative terms, and using the bound \( G(a,b) \leq (1 - \gamma (a+b))^{-1} \) yields the inequality \( Q_{Nj}(v) \leq (j+1) \gamma^j \).

Next consider large \( N \). Define \( c = (j+1) \log(N) / \gamma (N-1) \). If \( a > c \) or \( b > c \), then \( \log G(a,b) \leq -(N-1) \gamma (a+b) \leq -(j+1) \log(N) \), implying \( G(a,b) \leq N^{-1} \). Substituting this bound in the expression for the expectation then gives

\[
\int_0^A \int_0^B N_j^i (a+b)^j G_{ab}(a,b) \, da \, db = \int_0^c N_j^i a^{-1} G(a,0) \, da + \int_0^c N_j^i b^{-1} G(0,b) \, db \\
+ \int_0^c \int_0^c N_j^i(j-1)(a+b)^j 2 G(a,b) \, da \, db + O(N^{-1})
\]

For \( 0 \leq a, b \leq c \), a Taylor's expansion of \( \log G(a,b) \) yields a \textit{uniform} approximation \( G(a,b) = e^{-\Psi(v)(a+b)(N-1)} [1 + O((\log(N))^2/N)] \). Substituting this approximation in the expression for the expectation yields the result \( Q_{Nj}(v) = (j+1) \gamma^j / \Psi(v) + O((\log(N))^2/N) \).

Let \( \Psi_N \) denote the empirical distribution of \( \Psi \). If \( v_{Ni} - v_{N,i-1} > a \), then \( \Psi_N(v_{N,i-1}+a) - \Psi_N(v_{N,i-1}) = 0 \) and \( \Psi(v_{N,i-1}+a) - \Psi(v_{N,i-1}) \geq \gamma a \). Therefore, if \( \Psi_N(v_{N,i-1}) - \Psi(v_{N,i-1}) \leq \gamma a/2 \), then \( \Psi_N(v_{N,i-1}+a) - \Psi(v_{N,i-1}+a) \leq -\gamma a/2 \), implying \( D_N \leq \sup_{v} |\Psi_N(v) - \Psi(v)| \leq \gamma a/2 \). Similarly, \( v_{N,i+1} - v_{Ni} > b \) implies \( D_N \geq \gamma b/2 \).

Since \( a + b = c \), \( \max{a,b} \geq \gamma c/4 \). But \( D_N \) is the Kolmogorov-Smirnov statistic, and the Dvoretzky-Kiefer-Wolfowitz theorem (see Serfling, 2.1.3,
Theorem A) states that \( \text{Prob}(D_N > d) \leq C e^{-2Nd^2} \), where \( C \) is a universal constant. This completes the proof of Lemma 2. 

The following result establishes the statistical properties of \( \mu^2_{N}(x_0, \beta, r) \):

**Theorem 3.** Suppose Assumptions (I)-(V) and condition [R]. Assume that \( v \) has bounded support, so that \( \alpha_0, \alpha_1 < +\infty \); and that \( r'(w, x_0) \) and \( G(v) \) satisfy Lipschitz conditions. Then \( \mu^2_{N}(x_0, \beta, r) \to \mu(x_0, \beta, r) \), and \( \sqrt{N}[\mu^2_{N}(x_0, \beta, r) - \mu(x_0, \beta, r)] \) converges in distribution to \( N(0, \sigma^2_2) \), with \( \sigma^2_2 \) given by (13) and consistently estimated by (14).

**Proof:** Rewrite the estimator (11) as \( \mu^2_{N}(x_0, \beta, r) - \mu(x_0, \beta, r) = A + B \), with

\[
A = \sum_{i=1}^{N} r'(x_0 \beta - v_{N_i} x_0) [y_{N_i} - G(v_{N_i})] \frac{v_{N_i,i+1} - v_{N_i,i-1}}{2}
\]

\[
B = \sum_{i=1}^{N} r'(x_0 \beta - v_{N_i} x_0) [G(v_{N_i}) - 1(v_{N_i} > 0)] \frac{v_{N_i,i+1} - v_{N_i,i-1}}{2}
\]

\[
- \int_{-\infty}^{\infty} r'(x_0 \beta - v, x_0) [G(v) - 1(v > 0)] dv.
\]

Define \( \xi_{N_i} = \sqrt{N} r'(x_0 \beta - v_{N_i} x_0) [y_{N_i} - G(v_{N_i})] \frac{v_{N_i,i+1} - v_{N_i,i-1}}{2} \) and let \( \mathcal{E}_{N_i} \) denote the \( \sigma \)-field of events generated by \( Z_{N_i} = (X_{N_i}^i, Y_{N_i}^i, V_{N_i}^i) \) for \( j \neq i \). Then, one has the conditional expectations

\[
E(\xi_{N_i} | \mathcal{E}_{N_i}) = 0,
\]

and
(13) \[
\sigma_{Ni}^2 = E(\zeta_{Ni}^2 | \tau_{Ni}) = N \cdot \frac{v_{N,i+1} \int r'(x_0 \beta - v, x_0)^2 \cdot G(v) \cdot [1 - G(v)] \psi(v) dv}{\left[ \frac{v_{N,i+1} - v_{N,i-1}}{2} \right]^2}.
\]

Also, letting \( \kappa(v) = G(v) \cdot [1 - G(v)] \cdot \alpha(v) \cdot \psi(v) \),

\[
E(\zeta_{Ni}^4 | \tau_{Ni}) = N^2 \cdot \frac{v_{N,i+1} \int r'(x_0 \beta - v, x_0)^4 \cdot \kappa(v) \cdot \psi(v) dv}{\left[ \frac{v_{N,i+1} - v_{N,i-1}}{2} \right]^4}.
\]

Lemma 2 implies \( E(\zeta_{Ni}^4 | \tau_{Ni}) = O(N^{-2}) \) and

\[
E \sum_{i=1}^{N} \sigma_{Ni}^2 = \sigma_2^2 + O(N^{-1}),
\]

with

\[
\sigma_2^2 = \frac{\alpha_1}{\alpha_0} \int \frac{r' (x_0 \beta - v, x_0)^2 \cdot G(v) \cdot [1 - G(v)] \psi(v)}{\psi(v)} dv.
\]

A consistent estimator of this variance is obtained from (13) by approximating the conditional expectation by an intermediate value and approximating \( G(v) \cdot [1 - G(v)] \) by an estimator that has an asymptotically negligible bias,

(14) \[
\hat{\sigma}_2^2 = N \cdot \sum_{i=1}^{N} r'(x_0 \beta - v_{Ni}, x_0)^2 \cdot y_{Ni} \cdot \left[ 1 - \frac{y_{Ni,i+1} + y_{Ni,i-1}}{2} \right] \cdot \left[ \frac{v_{N,i+1} - v_{N,i-1}}{2} \right]^2
\]

15
Since \( \text{Var}(\sigma_{Ni}^2) \leq \mathbb{E} \mathbb{E}(\xi_{Ni}^4 | \mathcal{F}_{Ni}) = O(N^2) \), Chebyshev's inequality implies \( \sum_{i=1}^{N} \sigma_{Ni}^2 \) converges in probability to a positive finite limit. Then, \( \{\xi_{Ni}\} \) forms a Martingale difference array, with the expectation \( \mathbb{E}(\xi_{Ni}^4 | \mathcal{F}_{Ni}) = O(N^2) \) guaranteeing that a Lindeberg condition holds. Then the central limit theorem for Martingale difference arrays guarantees that \( \sqrt{N} \mathbf{A} = \sum_{i=1}^{N} \xi_{Ni} \xrightarrow{d} N(0, \sigma_2^2) \); see Pollard (1984, Theorem VIII.1.1).

Next, consider \( \mathbf{B} \). Apply the Lipschitz property of \( r' \) and \( G \) to obtain the bound

\[
|\mathbf{B}| \leq C_8 \sum_{i=1}^{N} (\nu_{N,i+1} - \nu_{N,i-1})^2 + C_9 (\nu_{N,i_0+1} - \nu_{N,i_0-1})^2,
\]

where \( C_8 \) is determined by the Lipschitz constant, \( C_9 \) is a bound on \( r' \), and \( i_0 \) indexes one end of the interval that brackets zero. Then, Lemma 2 implies

\[
\mathbb{E} |\mathbf{B}| \leq C_8 6/\gamma^2 N + C_9 2/\gamma N.
\]

Markov's inequality then gives the result that \( \sqrt{N} \mathbf{B} \xrightarrow{p} 0 \). □

When \( G \) is known to be smooth, a variant of the estimator (8) that exploits this property replaces the binomial observation \( y_i \) with a kernel estimate of \( G(y_i) \). This smoothing requires additional computation, but may increase precision. Let \( \hat{G} \) denote a kernel estimator of \( G \); it is the Nadaraya-Watson estimator of the nonparametric regression \( \mathbb{E} Y | Y = G(Y) \) in (4):

\[
\hat{G}(y) = \left[ \frac{1}{\lambda_N N} \sum_{j=1}^{N} y_j k \left( \frac{y - y_j}{\lambda_N} \right) \right] + \left[ \frac{1}{\lambda_N N} \sum_{j=1}^{N} k \left( \frac{y - y_j}{\lambda_N} \right) \right],
\]

(15)

where \( \lambda_N \to 0 \) is a bandwidth and \( k(\cdot) \) is a kernel satisfying the assumption:

16
(VI) $k(u)$ is symmetric, non-negative, and twice continuously differentiable, and $k$ and its derivatives are bounded and integrable, and $N^{-4} \lambda_N \to 0$.

The rate requirement on $\lambda_N$ guarantees that bias in the estimator $\hat{G}(v_i)$ is asymptotically negligible. There is no upper limit on the rate at which $\lambda_N \to 0$, and "undersmoothing" causes no problem in estimating $\mu(x_0, \beta, r)$ at a $\sqrt{N}$ rate, as Theorem 2 shows. Let $\mu_N^3(x, \beta, r)$ denote the variant of $\mu_N^4(x, \beta, r)$ where $\hat{G}(v_i)$ replaces $y_i$ in (8).

A smoothed variant of the estimator (11) replaces $y_{Ni}$ by a nearest neighbor estimator

\begin{equation}
\tilde{G}(v_{Ni}) = \sum_{j=0}^{N+1} y_{Nj} \kappa_N(i,j),
\end{equation}

where $\kappa_N(i,j)$ is a weight function satisfying

(VI') $\kappa_N(i,j)$ is non-negative, $\sum_{j=0}^{N+1} \kappa_N(i,j) = 1$, and $\sum_{j=0}^{N+1} |N^{1/2} j - i| \kappa_N(i,j) \to 0$.

The last condition in (VI') guarantees that the bias in the nearest neighbor estimator is asymptotically negligible. As in the case of the kernel smoother, there is no upper limit on the rate at which the nearest neighbor weights shrink, as undersmoothing causes no problem for the moment estimator. Let $\mu_N^3(x, \beta, r)$ denote the variant of estimator $\mu_N^3(x, \beta, r)$ obtained by replacing $y_{Ni}$ by $\tilde{G}(v_{Ni})$.

The following result gives the statistical properties of the moment estimators $\mu_N^3(x, \beta, r)$ and $\mu_N^4(x, \beta, r)$.

**Theorem 4.** Suppose Assumptions (I)-(V). Assume $v$ has bounded support, so that $\alpha_0, \alpha_1 < +\infty$. Assume that $G$ is continuously differentiable, and that $r(w, x)$ satisfies condition [R] and is twice continuously differentiable. For the estimator $\mu_N^3(x, \beta, r)$, assume (VI); and for the estimator $\mu_N^4(x, \beta, r)$, assume (VI'). Then, these modified estimators are RCAN.
Proof: The arguments for these results parallel the proofs of Theorems 2 and 3, and we give only an outline of the essential elements. Essentially, the projection methods used in those proofs are applied again, with the projections now extended into one more dimension to handle the averaging over observations that is contained in the kernel or nearest neighbor smoothers.\(^5\)

Consider first the estimator \( \mu^3_N(x, \beta, \tau) \). Define

\[
a_{in} = \frac{r(x_0 \beta - \nu_i x_0)}{\psi(\nu_i)} - 1(\nu_i > 0)
\]

and consider a kernel \( u(z_{i,j}, z_n) \) that is the average, over permutations of the indices \( i,j,n \), of the expression

\[
a_{in} - E_{a_{in}} = \left[ a_{in} - r(x_0 \beta - \nu_i x_0) \frac{y_n}{\lambda_N \psi(\nu_i)} k \left( \frac{v_n - \nu_i}{\lambda_N} \right) \right] \frac{h(x_0 \beta - \nu_i x_0)}{\psi(\nu_i)} .
\]

Then \( \mu^3_N(x, \beta, \tau) - \mu(x, \beta, \tau) \) can be written as the sum of a U-statistic with this kernel that is RCAN, plus higher order terms that can be shown by the same arguments as in Theorem 2 to be asymptotically negligible. Because no kernel estimators appear in the denominator of the statistic, there is no issue of trimming, and additional moment conditions are not needed, even through they are available from the assumption

\(^5\) When the window width \( \lambda_N \) for the estimator \( \hat{G} \) satisfies \( N \lambda_N^2 \to \infty \), sufficient conditions are met to apply McFadden and Newey (1994, Theorem 8.1), which provides general conditions under which an estimator containing an embedded semiparametric estimator is RCAN. This case includes, for example, the window width \( \lambda_N \sim N^{-1/3} \).

The theorem includes undersmoothed cases, with \( N \lambda_N^2 \to 0 \), where the estimator \( \mu^3_N \) remains well-behaved even though the general theory for embedded semiparametric estimators does not apply. Note that the window width \( \lambda_N \sim N^{-1/5} \) that minimizes the integrated mean square error in \( \hat{G} \) is not rapid enough to eliminate asymptotic bias in the moment estimator. A similar situation holds for the weights in the nearest neighbor estimator \( \hat{G} \).
in this theorem that \( \nu \) has bounded support.

Next consider the estimator \( \hat{\mu}_N(x, \beta, \nu) \). Define

\[
\zeta_{Nij} = \sqrt{N} r(x_{0i} \beta - \nu_{Ni} x_0) \{ y_{Nj} - G(\nu_{Nj}) - \kappa_N(i, j) \} \frac{\nu_{Ni,i+1} - \nu_{Ni,i-1}}{2}
\]

and let \( \xi_{Nij} \) denote the \( \sigma \)-field of events generated by \( z_{Nn} = (x_{Nn}, \nu_{Nn}, y_{Nn}) \) for \( n \neq j \). Then \( E(\xi_{Nij} | \xi_{Nj}) = 0 \), and the \( \{ \zeta_{Nij} \} \) form a Martingale difference array. Write the estimator as

\[
\mu_N^*(x, \beta, \nu) - \mu(x, \beta, \nu) = A + B + C,
\]

where

\[
A = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \zeta_{Nij},
\]

\[
B = \sum_{i=1}^{N} r'(x_{0i} \beta - \nu_{Ni} x_0) \{ G(\nu_{Ni}) - 1(\nu_{Ni} > 0) \} \frac{\nu_{Ni,i+1} - \nu_{Ni,i-1}}{2} - \int r'(x_{0i} \beta - \nu, x_0) \{ G(\nu) - 1(\nu > 0) \} d\nu,
\]

\[
C = \sum_{i=1}^{N} \sum_{j=1}^{N} r'(x_{0i} \beta - \nu_{Ni} x_0) \{ G(\nu_{Nj}) - G(\nu_{Ni}) \} \kappa_N(i, j) \frac{\nu_{Ni,i+1} - \nu_{Ni,i-1}}{2}.
\]

As in the proof of Theorem 3, the CLT for Martingale difference arrays establishes that \( \sqrt{N} A \) is RCAN. That proof established that \( \sqrt{N} B \to 0 \). Assumption (VI') controls the bias in \( \tilde{G} \), and establishes that \( \sqrt{N} C \to 0 \).


The unsmoothed \( y \) observations are of course not in general monotone in \( \nu \). Similarly, the smoothers \( \hat{G} \) in (15) and \( \tilde{G} \) in (16) need not be monotone. It is possible that some improvement in estimator precision can be achieved by constraining \( \hat{G} \) and \( \tilde{G} \) to be monotone, although the analogy to nonparametric regression where
monotonicity restrictions fail to yield improvements in convergence rates suggests that the performance gain may be slight. Enforcement of monotonicity is most easily done using a pool adjacent violators algorithm (PAVA); see Cosslett (1983). We have not investigated the large sample properties of estimators in which PAVA is applied.

4. MOMENT ESTIMATES WITH UNKNOWN $\beta$

Consider the usual case in applications where one is interested in moments of $W$ conditioned on the index $x_0\beta$, with $\beta$ unknown. Suppose one substitutes the estimator $\hat{\beta}$ from (5), or any other RCAN estimator, into one of the estimators $\mu_N^i(x_0,\beta,\eta)$ for $i = 1,\ldots,4$ from Theorems 2, 3, or 4. In each case, the estimator is continuously differentiable in each $v_i = b_i - x_i\beta$. (For the estimators employing order statistics, the probability of ties is zero, and hence with probability one the order will be unchanged by an infinitesimal change in $\beta$. Therefore, these estimators will be continuously differentiable with probability one.) These conditions, plus compactness, are sufficient to give the following result:

**Theorem 5.** Suppose the assumptions of Theorems 2-4 hold. Suppose $\hat{\beta}$ is the RCAN estimator of $\beta$ given by Theorem 1, or another RCAN estimator for this problem, and let $\Omega$ denote the asymptotic covariance matrix of $\sqrt{N}(\hat{\beta} - \beta))$. Then

$$\Gamma_i = \lim_{N} V\beta [\mu_N^i(x_0,\beta,\eta) - \mu(x_0,\beta,\eta)]$$

is a finite vector, and

$$\sqrt{N}[\mu_N^i(x_0,\hat{\beta},\eta) - \mu(x_0,\beta,\eta)] \overset{d}{\to} N(0,\sigma_i^2 + \Gamma_i'\Omega\Gamma_i).$$


5. QUANTILE ESTIMATES

In contrast to the results for estimation of unconditional or conditional moments, we show that the unconditional median of $G$ cannot be estimated at a root-$N$ rate, and give an estimator that attains the best rate. The reason for this difference in moment and quantile estimates is that the latter depend critically on local properties of $G$ that cannot be estimated nonparametrically with the same
precision as smooth functionals of $G$ such as moments that depend only on its global properties.

The $q$-th quantile satisfies $w_q = x^\beta - G^{-1}(1-q)$, and a quantile estimator satisfies $\hat{w}_q = x^\beta - \hat{G}^{-1}(1-q)$, where $\hat{G}$ is a nonparametric estimator of $G$. We assume $\hat{G}$ is the kernel estimator (15), and note that this is one of a number of nonparametric estimators that can achieve the best rate for mean square error. Write the kernel estimator as $\hat{G}(v) = G(v) + A(v) + B(v)$, where

$$A(v) = \frac{1}{N\lambda N^{-1}\sum_{i=1}^N} \sum_{i=1}^N \left( \frac{v_i - \lambda N}{\lambda N} \right) y_i - G(v_i)$$

$$B(v) = \frac{1}{N\lambda N^{-1}\sum_{i=1}^N} \sum_{i=1}^N \left( \frac{v_i - \lambda N}{\lambda N} \right) (G(v_i) - G(v))$$

The standard theory of nonparametric kernel regression implies that under the assumptions (I)-(VI), plus assumptions that $G$ is twice continuously differentiable and $\psi$ is continuously differentiable, the bandwidth that minimizes integrated mean square error in $\hat{G}(v)$ is proportional to $N^{-1/5}$, and that with this bandwidth, $N^{2/5}(A(v) + B(v))$ is asymptotically normal (with a nonzero mean that can be removed, if desired, by use of a higher-order kernel). Further, this random function is stochastically equicontinuous in $v$.

The theorem of the mean applied to

$$1-q = G(x_0\beta - w_{q'}) = G(x_0\beta - \hat{w}_q') + A(x_0\beta - \hat{w}_q') + B(x_0\beta - \hat{w}_q')$$

gives

$$\hat{w}_q' - w_q = \frac{A(x_0\beta - \hat{w}_q') + B(x_0\beta - \hat{w}_q')}{G(x_0\beta - w_{q'})}$$

where $w_q'$ is intermediate between $\hat{w}_q$ and $w_q$. Then, $N^{2/5}(\hat{w}_q - w_q')$ is asymptotically normal at any quantile that has $G(x_0\beta - w_{q'})$ positive. Better rates can be attained only if $G$ and $\psi$ satisfy stronger smoothness assumptions, and even then only if a
higher-order kernel that exploits this smoothness is used. The root-N rate achieved for moments of continuously differentiable functions \( r(w,x) \) cannot be achieved by quantile estimates.

6. A MONTE CARLO STUDY OF ESTIMATOR CHARACTERISTICS

Monte Carlo analysis gives an idea of the finite-sample properties of the estimators proposed in this paper. For illustration, we consider a problem with a single covariate, uniformly distributed on \([0,4] \), \( \beta = 1 \), and \( G \) normal with variance 4, truncated to \( |u| < 6 \). We use a truncated bilateral exponential design density

\[
h(b) = e^{-\frac{|b-2|}{\sqrt{8}/32}(1 - e^{-2})}, \quad |b - 2| \leq 16.\]

For the kernel-smoothed estimator \( \hat{\mu}^3 \), we use an Epanechnikov kernel and a bandwidth of 0.05, and for the nearest-neighbor-smoothed estimator \( \hat{\mu}^4 \) we use triangular weights with 10 neighbors. Table 1 gives the mean and standard error of the estimator \( \hat{\beta} \) from 1000 Monte Carlo trials, as well as the mean of the asymptotic estimated standard error. The results demonstrate that \( \hat{\beta} \) is unbiased, and that the asymptotic standard error estimate is accurate in samples of this size:

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \beta )</td>
<td>1.000</td>
</tr>
<tr>
<td>Mean Estimated ( \hat{\beta} )</td>
<td>0.993</td>
</tr>
<tr>
<td>Monte Carlo Estimated SE of ( \hat{\beta} )</td>
<td>0.163</td>
</tr>
<tr>
<td>Mean Asymptotic SE of ( \hat{\beta} )</td>
<td>0.162</td>
</tr>
</tbody>
</table>

We use these trials and their \( \hat{\beta} \)'s to form estimates of the conditional mean of \( e^W \), given \( x_0 = 2 \). Table 2 gives the mean and standard error of each of the estimators \( \hat{\mu}^i \) for \( i = 1, \ldots, 4 \), and the asymptotic standard errors for each estimator. Due to the strong dependence of \( \mathbb{E}_W|X e^W \) on the upper tail of the distribution and the relatively limited information in binomial data, estimates of this quantity will necessarily be fairly noisy. All the estimators show a small finite sample upward bias, which in none of the cases is statistically significant. There appears to be some advantage to smoothing \( \mu^2 \), but not to smoothing \( \mu^1 \).
Table 2

<table>
<thead>
<tr>
<th></th>
<th>$\mu^1$</th>
<th>$\mu^2$</th>
<th>$\mu^3$</th>
<th>$\mu^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True $\mu = E_W</td>
<td>x^W$</td>
<td>46.060</td>
<td>46.060</td>
<td>46.060</td>
</tr>
<tr>
<td>Mean Estimate $\mu'</td>
<td>$</td>
<td>46.924</td>
<td>47.047</td>
<td>47.059</td>
</tr>
<tr>
<td>Monte Carlo Estimated SE of $\mu'</td>
<td>$</td>
<td>33.650</td>
<td>42.097</td>
<td>33.979</td>
</tr>
</tbody>
</table>

In this experiment, the estimators based on importance sampling are clearly superior to the ones based on order statistics. However, the importance sampling estimator is likely to be somewhat less robust, and exhibit large deviations when a poorly chosen design density is small in regions where $G$ is not near zero or one. Further analysis is needed before we can reach clear conclusions on the circumstances under which the importance sampling estimators dominate the order statistic estimators.

7. EXTENSIONS

(i) In non-destructive duration analysis, a subject is tested sequentially for the presence of an abnormality, at observation times set by experimental design. The analysis is retrospective if time of onset can be determined when the abnormality is present, and non-retrospective otherwise. Traditional duration analysis models handle the case of continuous monitoring or retrospective measurement in which the only problem is that of censoring of incomplete spells. There is an additional family of problems where testing is done sequentially and retrospective information is not available. These resemble the problem of non-retrospective destructive duration analysis. Similarly, economic surveys and contingent valuation experiments may employ a series of gates, with responses trapping $W$ in an interval. An unfolding bracket method presents the gates sequentially, but there are also designs where the subject chooses among multiple brackets. A feature of all these methods is that they determine a largest bid with a $y = 1$ response and a smallest bid with a $y = 0$ response. The estimators in this paper can be extended to handle such problems. When perturbation of response behavior by the observation process is not an issue, one can treat the smallest unsuccessful bid $(b', 0)$ and largest successful bid $(b', 1)$ as if they were independent observations, and then estimate moments or quantiles of the unknown distribution as before, with the expanded data.

(ii) Non-random designs. This paper obtains RCAN estimators for index parameters and moments when randomized designs are used that give a positive density
for bid levels over the support of \( W \), but notes that when this condition is not met, consistent estimators are not available. This is an issue for applications, as many studies use fixed designs with a finite number of bid treatments. If some component of the covariates is continuous, it may be possible to estimate the index parameters consistently using the variant of (5) given by Lewbel (1997) in which a kernel estimator of the density of this covariate is used. A continuous covariate may be sufficient to yield \( \psi(v) \) positive, so that the estimators \( \mu^i \) are RCAN for moments of \( W \), even for fixed designs. More generally, estimator \( \mu^4 \) will in many cases provide a reasonable approximation of \( \mu \), even when consistency is technically unattainable. From (10) and the approximation properties of the extended trapezoid rule for numerical integration, when \( r' \) and \( G \) are continuously differentiable, \( \mu(x_0, \beta, r) \) and the expression

\[
(17) \quad r(x_0, \beta, x_0) + \sum_{i=1}^{N} r'(x_0, -v_{Ni} x_0) [G(v_{Ni}) - 1(v_{Ni} > 0)] \frac{v_{Ni,i+1} - v_{Ni,i-1}}{2},
\]

differ by a factor that is of the order of

\[
\sum_{i=1}^{N} \left( \frac{v_{Ni,i+1} - v_{Ni,i-1}}{2} \right) \times \text{bound on the slope of } r'G.
\]

This quantity will be small for a well-specified fixed design. The estimator \( \mu^4 \), which replaces \( G \) by a nearest neighbor estimator \( \hat{G} \), will be a RCAN estimator of (17), and hence will approximate \( \mu \).

8. CONCLUSIONS

This paper has considered the problem of estimating single index parameters and features of an unknown distribution from binomial observations on whether draws from the distribution exceed thresholds set by experimental design. This situation arises in destructive duration analysis and in economic survey research and contingent valuation experiments. We have proposed practical estimators for these problems and have shown that under suitable design conditions they are root-N consistent and asymptotically normal.
REFERENCES


