Estimating Conditional Moments of a Survival Curve from Interval-Censored Data

Arthur Lewbel, Boston College
Oliver Linton, LSE
Daniel McFadden, UC Berkeley

Revised 2003

Outline

1. **A willingness-to-pay (WTP) experiment**
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Willingness to Pay for Seabirds

Green, Jakowitz, Kahneman, McFadden (1997)

What value would your household place on saving about 50,000 seabirds each year from offshore oil spills?

- Several million seabirds live out of sight off Pacific coast
- Small oil spills kill estimated 50,000+ seabirds per year
- Usually not possible to force tanker companies to pay
- Public money would have to be spent yearly to save the birds, extra funds required

Contingent Valuation Experimental Design

- Control subjects asked open-ended WTP
- In what is termed referendum format, treatment subjects were asked if WTP ≥ v, where bid v was set by experimental design
- In the GJKM study, bids were set at quantiles of the controls' WTP distribution
- Questions:
  - What are median and mean WTP?
  - Does the format (open-ended vs. referendum) matter?

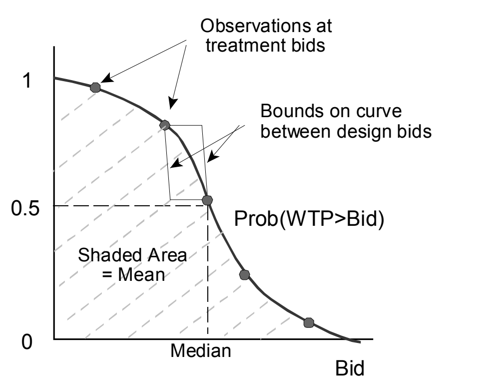
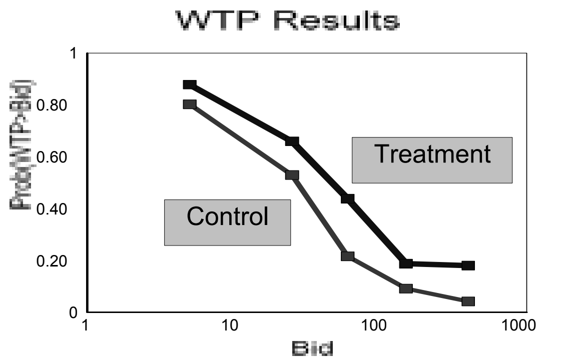


Table 1. Willingness-to-Pay to Save 50,000 Off-Shore Seabirds per Year

Distribution	Open-Ended	Starting Point Bid				
		$5	$25	$60	$150	$400
$0-4.99	19.8%	12.2%	8.5%	0.0%	8.3%	12.0%
$5-24.99	27.3%	67.4%	25.5%	41.7%	29.2%	22.0%
$25-59.99	31.4%	12.2%	53.2%	14.6%	27.1%	20.0%
$60-149.99	12.4%	8.2%	8.5%	41.7%	16.7%	18.0%
$150-399.99	5.0%	0.0%	2.1%	2.1%	18.8%	10.0%
$400+	4.1%	0.0%	2.1%	0.0%	0.0%	18.1%
Sample size	121	49	47	48	48	50
P(Open-Ended Response>Bid)		80.2%	52.9%	21.5%	9.1%	4.1%
(Std. Error)		5.7%	7.1%	5.9%	4.1%	2.8%
P(Anchored Response>Bid)		87.8%	66.0%	43.8%	18.8%	18.0%
(Std. Error)		4.7%	6.9%	7.2%	5.6%	5.4%
Median Response	$25.00	$10.00	$25.00	$25.00	$43.00	$50.00
(Std. Error)	$6.03	$2.33	$1.16	$14.04	$10.87	$23.41
Mean Response (a)	$64.25	$20.30	$45.43	$49.42	$60.23	$143.12
(Std. Error)	$13.22	$3.64	$12.61	$6.51	$8.59	$28.28
		Coefficient	Std. Error			
Marginal effect of starting point bid		0.284	0.32			
K-J Interquartile Anchoring Index		0.273	0.136			
Nonparametric referendum mean (b)		$167.33	$76.90			
Referendum multiplier		2.60	1.31			
Parametric referendum mean		$265.59	$138.96			
Referendum multiplier		4.13	2.32			



Outline

1. A willingness-to-pay (WTP) experiment
2. **Conditional survival curves**
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Survival Curves

- Let T denote a random failure time, and $G(t|x) = \text{Prob}(T \geq t|x)$, $t \geq 0$, denote the survival curve conditioned on a d-vector of (time-invariant) covariates x.
- In most applications, t is time. Alternately:
  - t is the administered dose of a toxin, T is lethal dose, and G is the dose-response curve
  - t is a bid in referendum Contingent Valuation and T is the subject's Willingness-to-Pay (WTP).

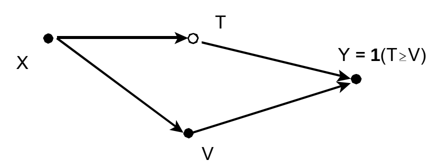
Survival Data

- Survival analysis often assumes a size N sample of i.i.d. observations (x_n,t_n), where x_n is a d-vector of covariates and t_n is completed (or censored) duration.
- Some applications provide interval-censored data: T_n is latent and one observes (x_n,v_n,y_n), where v_n is a test level in the t dimension set by experimental design, and independent of T_n given x_n. $y_n = \mathbf{1}(T_n \geq v_n)$ is a binary indicator for the event $T_n \geq v_n$. The conditional mean of y_n given x_n) is $G(v_n|x_n)$.

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. **Interval-censored data and examples**
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Interval-Censored Data



- Critical feature: T and V conditionally independent, given X
- Failure times not observed retrospectively.
- Analysis of a single test level v and a binary status indicator y can be extended to multiple (adaptive) test levels and multinomial status.

Interval-Censored Data Examples

- Animal experiments: At time v, the animal is sacrificed, y is one iff abnormality is present
- Materials Testing: At treatment level v, y is one iff material meets requirement; e.g., crash test at speed v.
- Dose-Response: At treatment level/dose v, y is one iff lethal dose exceeds v.
  - Referendum contingent valuation has the dose-response form, testing if willingness-to-pay (WTP) exceeds a bid v.

Longitudinal Interval-Censored Data

- Panels with periodic waves yield interval-censored data if retrospective data on T is unavailable or unreliable when failure occurs between waves.
- Statistical issue: If x is not time-invariant, then even if x(s), $0 \le s \le t$, is predetermined for T given $T \ge t$, intra-wave feedbacks may nevertheless make x(v) endogenous.
- Statistical issue: If the failure time T interacts with interview scheduling, then the inter-wave duration (v) becomes endogenous, biasing conditional hazard rate estimates.
  - Measured hazard rates in the Health and Retirement Study depend on interview timing within a wave

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. **Survival models and specializations**
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Alternative Formulation of Survival Model

- If $T = \exp(m^*(x,\eta))$, with m* decreasing in a disturbance η that has a continuous CDF $F(\cdot|x)$, then $G(t|x) = F(M^*(x,\log t)|x)$, where M* is the inverse of m* in its 2nd argument.
- Normalization: Define $\xi = F(\eta|x)$ and $T = \exp(m(x,\xi)) \equiv \exp(m^*(x,F^{-1}(\xi|x)))$. Then ξ is uniform [0,1] and $G(t|x) = M(x,\log t)$, where M is the inverse of m in its 2nd argument.

Specializations

- $T = \Lambda(m(x,\theta_0) - \eta)$
  $\rightarrow G(t|x) = F(m(x,\theta_0) - \Lambda^{-1}(t)|x)$
  a <u>semiparametric model</u> when F, θ_0 unknown, m, Λ known
  - $T = \exp(m(x,\theta_0) - \eta)$ $\rightarrow G(t|x) = F(m(x,\theta_0) - \log t|x)$
- $T = \exp(x\cdot\theta_0 - \eta) \rightarrow G(t|x) = F(x\cdot\theta_0 - \log t|x)$
  - Type 1 extreme value $F(\eta) = \exp(-\exp(-\alpha\eta))$ gives the <u>parametric</u> Weibull proportional hazards model,
    $G(t|x) = \exp(-t^{\alpha}\cdot\exp(-x\cdot\theta_0\alpha))$

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. **Regularity and design assumptions**
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Assumptions to set the problem

- **A.1.** Covariate vectors x are distributed in the population with a CDF H_x that has a compact support in a d-dimensional space. The survival curve G has a continuously differentiable positive density g(t|x) with a compact support.
- **A.2.** The generalized moment function r(t,x) is continuous in (t,x), and for each x is twice continuously differentiable in t.

Experimental Design Assumption

- **A.3.** There is an asymptotic distribution $H(v,x) = \int_{w\le v}\int_{z\le x} h(w|z)dwH_x(dz)$ of the treatments and covariates, where h(t|x) is a continuous density that for each x is strictly positive on a compact interval containing the support of G(t|x). The experimental design is described by an empirical CDF $H_N(v,x)$ such that $N^{1/2}[H_N(v,x) - H(v,x)]$ converges weakly to a Gaussian process.
- For some nonparametric estimators, a rate less than $N^{1/2}$ suffices and Gaussianity is not required
- A.3 implies $\sup_{v,x} |H_N(v,x) - H(v,x)| \to 0$ a.s.

Experimental Design Examples

- $H_N(v,x)$ is a random sample from H(v,x).
  - A.3 holds by Shorack-Wellner on convergence of triangular arrays of empirical processes, and a.s. convergence holds by Glivenko-Cantelli.
- At N, x_n is sampled randomly from H_x. A fixed design for v with J_N possible values of v is selected. v_n is drawn randomly from a density $h_N(v|x)$ on this finite support that for each x converges weakly to a positive continuous density h(v|x).
  - Sufficient: $J_N/N^{1/2} \to \infty$, max gap of order $1/J_N$, and CDF's of h_N and h coinciding at design points

Inference Problems

- Survival curve features of interest are moments, quantiles, and percentiles (unconditional, or conditional on x).
- The generalized moment problem to estimate $\mu(x) = \mathbf{E}_{T|x}r(T,x)$, for a C^2 function r(t,x) approximates many cases of interest.
- The estimation problem is semiparametric when unknown G depends on x through a known function $m(x,\theta_0)$ of an unknown parameter vector θ_0; e.g., the index $x\cdot \theta_0$.

Mathematical elements

- $\int r(v)G(dv) = r(0) + \int r'(v)G(v)dv$
- If Z is an importance random variable with a positive density q on the support of G, then
  $$\int r'(v)G(v)dv = \int[r'(z)G(z)/q(z)]q(z)dz = \mathbf{E}_Z\, r'(Z)G(Z)/q(Z)$$
- If V is a design random variable with a positive density h on the support of G, and Y is a status indicator with $\mathbf{E}_{Y|V}\, Y = G(V)$, then
  $$\int r'(v)G(v)dv = \mathbf{E}_V\, r'(V)G(V)/h(V) = \mathbf{E}_{V,Y}\, r'(V)Y/h(V)$$

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. **Nonparametric estimators**
7. Semiparametric estimators
8. Monte Carlo evidence
9. Application

Estimating $\mu(x) = \mathbf{E}_{T|x} r(T,x)$,

- $Y = \mathbf{1}(T \geq v)$ satisfies $G(v|x) = \mathbf{E}_{Y|v,x} Y$.
- Plug estimate $G^\wedge(t|x)$ into $\mu(x) = \int_{t\geq 0} r(t,x)G(dt|x)$ to get an estimator $\mu^\wedge_0(x)$
- Parametric problem is standard.
- In semiparametric or nonparametric problem
  - The curse of dimensionality applies -- undersmooth $G^\wedge$ to get a best rate for $\mu(x)$.
  - For practical estimation of $\mu(x)$, avoid explicit computation of $G^\wedge$ if possible

Estimators

Uncon-ditional	Con-ditional	Known	Unknown
$\mu^\wedge_0$	$\mu^\wedge_0(x)$	$H(v|x)$	$G(t|x)$
$\mu^\wedge_1$	$\mu^\wedge_1(x)$	$H(v|x)$	$G(t|x)$
$\mu^\wedge_2$	$\mu^\wedge_2(x)$	- - -	$H(v|x)$, $G(t|x)$
$\theta^\wedge$	- - -	m, Λ	$F(\eta)$, θ_0
$\mu^\wedge_3$	$\mu^\wedge_3(x)$	$H(v|x)$, m, Λ	$F(\eta)$
$\mu^\wedge_4$	$\mu^\wedge_4(x)$	m, Λ	$H(v|x)$, $F(\eta)$
$\mu^\wedge_5$	$\mu^\wedge_5(x)$	$H(v|x)$, m, Λ	$F(\eta)$

Estimators $\mu^\wedge_j$ are nonparametric for $j \leq 2$, semiparametric for $j > 2$.
Estimator $\mu^\wedge_5$ requires moment $r(t,x) = [\Lambda(t)^{-1}]^k$, k a positive integer.

Estimator descriptions

- Nonparametric
  - $\mu^\wedge_0(x)$ plug-in estimator
  - $\mu^\wedge_1(x)$ ragged integrand, design density
  - $\mu^\wedge_2(x)$ smooth integrand, any importance density
- Semiparametric
  - $\theta^\wedge$ nonlinear least squares
  - $\mu^\wedge_3(x)$ ragged integrand, design density
  - $\mu^\wedge_4(x)$ smooth integrand, uniform importance
  - $\mu^\wedge_5(x)$ special ragged integrand, design density

Integration-by-parts formulation of $\mu(x)$

- Notation: $r'(t,x) = \partial r(t,x)/\partial t$
- Define $s(x,v,y) = yr'(v,x)/h(v|x)$

  $\tau(x,v) = G(v|x)r'(v,x)/h(v|x) = \mathbf{E}_{Y|x}\, s(x,v,Y)$
- Integrating by parts,

$$\begin{aligned}\mu(x) &= \textstyle\int_{t\geq 0} r(t,x)g(t|x)dt = r(0,x) + \int_{t\geq 0} G(t|x)r'(t,x)dt \\ &= r(0,x) + \textstyle\int_{v\geq 0} \tau(x,v)H(dv|x) \\ &= r(0,x) + \mathbf{E}_{V|x}\, \tau(x,V) \qquad \text{smooth} \\ &= r(0,x) + \mathbf{E}_{Y,V|x}\, s(x,V,Y) \qquad \text{ragged}\end{aligned}$$

Example: Unconditional Moment Estimator
McFadden (1994), Lewbel (1997)

- Target: $\mu = \mathbf{E}_X\, \mu(X)$

  $= \mathbf{E}_X\, r(0,X) + \mathbf{E}_{Y,V,X}\, s(X,V,Y)$
- Estimate μ by a sample average

$$\mu^\wedge_1 = N^{-1}\textstyle\sum_{n\leq N} \{r(0,x_n) + s(x_n,v_n,y_n)\}$$

  - Assumptions 1-3 imply $\mu^\wedge_0$ is root-N CAN (elementary)
  - If x_n, v_n are sampled from $H(v,x)$, $\mu^\wedge_0$ is unbiased

Estimators $\mu^\wedge_0(x)$ and $\mu^\wedge_1(x)$

- Nonparametric estimator $G^\wedge(v|x)$ from, say, nearest neighbor regression of Y on v,x, is plugged into formula for $\mu(x)$ to get $\mu^\wedge_0(x)$. Estimator will have an IRMSE determined by $G^\wedge(v|x)$.
- Let $K_b(\cdot)$ denote a kernel of dimension d with bandwidth b. For each x, regress

  $s(x_n,v_n,y_n) = \alpha_0(x) + (x_n-x)\alpha(x)$,

  weighting the observations by $K_b(x-x_n)^{1/2}$.

  $\mu^\wedge_1(x) = r(0,x) + \alpha_0(x)$

  The large sample properties of $\mu^\wedge_1(x)$ are those of the local regression estimator of $\alpha_0(x)$

Estimator $\hat{\mu}_2(x)$

- Base estimator on smooth integrand

$$\mu(x) = r(0,x) + \mathbf{E}_{V|x}\, \tau(x,V)$$

- Replace G(v|x) by a local linear smooth of Y
- Replace h(v|x) by a uniform importance density

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. **Semiparametric estimators**
8. Monte Carlo evidence
9. Application

Semiparametric problem

- **A.4.** Assume
  $T = \Lambda(m(x,\theta_0) - \eta)$
  $\rightarrow G(t|x) = F(m(x,\theta_0) - \Lambda^{-1}(t)|x)$
  with F and θ_0 unknown, Λ and m known, Λ invertible and continuously differentiable, F a C^2 CDF independent of x with compact support containing 0
- $\mathbf{E}_{T|x}\, \Lambda^{-1}(T) = \alpha_0 + m(x,\theta_0)$, with $\alpha_0 = -\,\mathbf{E}\,\eta$
- Targets: The finite parameter vector θ_0 and the generalized conditional moment μ(x)

Definitions

(location-adjusted design)

- $U = m(x,\theta_0) - \Lambda^{-1}(V)$ and $u_n = m(x_n,\theta_0) - \Lambda^{-1}(v_n)$
- $\Psi_N(u)$ empirical CDF of U, with weak limit Ψ(u) that has a positive continuous density ψ on a support that contains the support of G
  - This property of $\Psi_N(u)$ follows from A.3., but may hold without A.3. if some components of x are continuously distributed
- $s^*(x,u,y) = r'(\Lambda(m(x,\theta_0) - u),x)\Lambda'(m(x,\theta_0) - u) \cdot(y-\mathbf{1}(u>0))/\psi(u)$

  $\tau^*(x,u) = \mathbf{E}_{Y|x}\, s^*(x,u,Y)$

Corollary 1

If A1-A4, then

$$E_{Y|u}\, Y = F(u)$$

$$\mu(x) = r(\Lambda(m(x,\theta_0)),x) + \int s^*(x,u)\psi(du)$$

$$\psi_N(u) = N^{-1}\sum_{n\le N} h(\Lambda(m(x_n,\theta_0)-u)|x_n)-u|x_n) \cdot\Lambda'(m(x_n,\theta_0)-u) \rightarrow \psi(u)$$

Estimator $\hat{\theta}$

- Define $s^\#(x,v,y) = y(d\Lambda^{-1}(v)/dv)/h(v|x)$
- $\Lambda^{-1}(0) + \mathbf{E}_{Y,V|x}\, s^\#(x,v,y) = \alpha_0 + m(x,\theta_0)$
- A nonlinear regression of $s^\#(x_n,v_n,y_n)$ on $\alpha_0 + m(x_n,\theta)$ provides a root-N CAN estimator of θ_0 if identification conditions are met

Estimator $\mu^\wedge_3(x)$

- When θ_0 is known:
  Replace $\psi(u)$ by $\psi_N(u)$ in $s^*(x,u,y)$
  $\mu^\wedge_3(x) = r(\Lambda(m(x,\theta_0)),x) + N^{-1}\sum_{n\le N} s^*(x,u_n,y_n)$
- When θ_0 is unknown, plug the estimator $\theta^\wedge$ into the definition of U and the formula $\mu^\wedge_3(x)$
- Theorem 3. The estimator $\mu^\wedge_3(x)$, with θ_0 either known or replaced by the plug in estimator $\theta^\wedge$, is root-N CAN

Estimator $\mu^\wedge_4(x)$

- When θ_0 is known:
  Form a kernel estimator $\psi^\sim_N(u)$ from the empirical density at the points $u_n = m(x_n,\theta_0) - \Lambda^{-1}(v_n)$.
  Replace $\psi(u)$ by $\psi^\sim_N(u)$ in $s^*(x,u,y)$,

  $$\mu^\wedge_4(x) = r(\Lambda(m(x,\theta_0)),x) + N^{-1}\sum_{n\le N} s^*(x,u_n,y_n)$$
- When θ_0 is unknown, plug the estimator $\theta^\wedge$ into the definition of U and the formula $\mu^\wedge_4(x)$
- Theorem 4. The estimator $\mu^\wedge_4(x)$, with θ_0 either known or replaced by the plug in estimator $\theta^\wedge$, is root-N CAN

Estimator $\mu^\wedge_5(x)$

- $r(v,x) = [\Lambda^{-1}(v)]^k$, k a positive integer
- $\mathbf{E}_{T|x} [\Lambda^{-1}(T)]^k = \mathbf{E}_{T|x} [m(x,\theta_0) - \eta]^k$
  $= \sum_{j\le k} (-1)^j {}_nC_j\, m(x,\theta_0)^{k-j}\, \mathbf{E}_\eta\, \eta^j$
  $= \sum_{j\le k} m(x,\theta_0)^j\, \alpha_j$
- Regress $s(x,v,y)$, defined for this special r, on $m(x,\theta_0)^j$ for $j \le k$ to estimate the α_j, and plug these into the formula above.

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. **Monte Carlo evidence**
9. Application

Monte Carlo Study

- $\log T = \beta_1 + \beta_2 X - \sigma\eta$
- X uniformly distributed on [-30,30]
- η standard normal
- $\beta_1 = 100$, $\beta_2 = 2$
- Treatments:
  - 5-bid design at {25, 50, 75, 125, 175}
  - Continuous design uniform on [25,175]
- Bandwidths chosen using Silverman's thumb
- 10,000 repetitions

Table 1. Conditional Mean in 5-bid design, 10,000 repetitions

		$\sigma = 5$		
		n=100	n=300	n=500
IRMSE	$\widehat{\mu}_1$	14.56	12.63	12.21
	$\widehat{\mu}_3$	17.59	16.04	15.74
	$\widehat{\mu}_4$	12.74	10.41	9.89
	$\widehat{\mu}_5$	11.74	10.33	10.02
IMAE	$\widehat{\mu}_1$	10.96	10.16	10.05
	$\widehat{\mu}_3$	14.42	13.36	13.13
	$\widehat{\mu}_4$	10.08	8.50	8.22
	$\widehat{\mu}_5$	9.23	8.44	8.31

Table 3. Conditional Mean in continuous design, 10,000 repetitions

		$\sigma = 5$ n=100	n=300	n=500
IRMSE	$\widehat{\mu}_1$	12.30	7.54	6.02
	$\widehat{\mu}_3$	12.65	8.05	6.98
	$\widehat{\mu}_4$	9.22	5.12	3.93
	$\widehat{\mu}_5$	8.91	5.13	4.00
IMAE	$\widehat{\mu}_1$	8.81	5.41	4.33
	$\widehat{\mu}_3$	9.99	6.46	5.70
	$\widehat{\mu}_4$	7.14	3.95	3.02
	$\widehat{\mu}_5$	6.87	3.97	3.08

Outline

1. A willingness-to-pay (WTP) experiment
2. Conditional survival curves
3. Interval-censored data and examples
4. Survival models and specializations
5. Regularity and design assumptions
6. Nonparametric estimators
7. Semiparametric estimators
8. Monte Carlo evidence
9. **Application**

Application

- WTP to protect California wetlands
- "Double Referendum" contingent valuation format: first bid drawn from design, second bid half if "No", double if "Yes"
- Covariates: Age, years in California, education, income bracket, sex, race, membership in environmental organization
- N = 530
- 14 bid levels total (number of first bid levels = ?)
- Data collected by Hanemann *et al*
- Model: log T = x·θ - η

Density of WTP

First bid data

(kernel-smoothed)

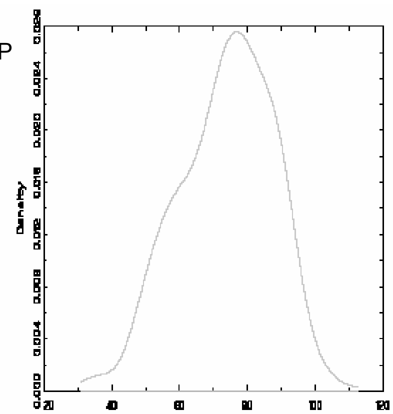


Table 5. Estimates Of Mean WTP

	Log Linear bid1	bid2
$\overline{\widehat{\mu}_3}$	62.0320 (4.4683)	306.0211 (411.4603)
$\widehat{\mu}_3(\overline{X})$	61.5918 (4.2751)	302.4752 (328.7766)
$\overline{\widehat{\mu}_4}$	64.6992 (5.0823)	369.2809 (394.6291)
$\widehat{\mu}_4(\overline{X})$	63.7869 (4.4995)	472.5140 (328.2098)
$\overline{\widehat{\mu}_5}$	99.1164 (4.1348)	141.5369 (9.0742)
$\widehat{\mu}_5(\overline{X})$	98.7726 (6.6526)	134.0196 (21.4996)

Table 6

	Log Linear bid1	bid2
YEARCA	0.0021 (0.0022)	0.0131 (0.0062*)
SEX	−0.0460 (0.0632)	0.2579 (0.1740)
ln(AGE)	−0.2040 (0.1088)	−0.4801 (0.2563)
EDUC	0.0119 (0.0154)	0.0307 (0.0404)
WHITE	0.1338 (0.0797)	0.2164 (0.2173)
ENVORG	−0.1085 (0.0792)	0.0946 (0.2331)
ln(INCOME)	0.0972 (0.0500*)	0.3796 (0.1474*)