Nonparametric Estimation of a Survivor Function with Across-Interval-Censored Data [1]

Mark Yuying An and Roberto A. Ayala [2]
Duke University and *Central Bank, Ecuador*

April 1996

---

[1] We thank Professors Ron Cummings, Timothy Haab, John Loomis and Bruce Turnbull for writen comments and Barbara Harrow for editing assitance. We thank Duke University and the Foundation Francisco Marroquín for financial support. An electronic version of the paper can be obtained in postscript format as "pub/man/papers/npecross.ps" on lewis.econ.duke.edu.

[2] Direct correspondence to Mark An, Department of Economics, Duke University, Durham, NC 27708-0097. Telephone: (919) 660-1809. Fax: (919) 684-8974. Email: man@econ.duke.edu.

Abstract

In this paper we present a *generalized self-consistent algorithm* that estimates a survivor function with *across-interval-censored* data. This algorithm is an iterative procedure based on Turnbull's (1974) *reallocation* idea. At each step of the iteration, the procedure first reduces the *across-interval-censored* problem to a *singly-censored* one, and then it applies the Kaplan-Meier estimation method. The main result of this paper is that our algorithm produces the maximum likelihood estimate. Unlike Turnbull (1974, 1976), we explicitly discuss situations in which corner solutions are encountered. The investigation is motivated from environmental economics where data from contingent valuation surveys are often used to nonparametrically estimate the willingness to pay distribution. In this estimation, the algorithm of Turnbull (1974, 1976) plays an instrumental role. However, there is a data grouping mechanism found in some contingent valuation surveys to which Turnbull's method does not apply. We refer to these cases as distinct bids and mixed bids, where *across-interval-censored* observations are common.

Key Words: Across-Interval Censoring, Contingent Valuation, Nonparametric Estimation, Self-Consistent Algorithm.

1 Introduction

This paper considers the estimation of a survivor function $S(x)$ from sample observations subject to a special case of *grouping* mechanism which we call *across-interval-censoring. Across-interval-censored* data are common in contingent valuation studies. Contingent valuation [3] is a technique that uses survey methods to elicit consumers' preferences for a change in the level of provision of a public good.

The study of the estimation of a survivor function from grouped data goes back at least to Ayer *et al* (1955) for binary threshold data, and Kaplan and Meier (1958) for possibly right-censored data. Turnbull's (1974) *self-consistent* algorithm, especially his own generalization –Turnbull (1976)– put the literature to rest with a high note. In this paper we revisit the problem for two reasons. First, *across-interval-censored* data, such as found in some contingent valuation surveys, are not covered by Turnbull (1976). Second, to prove the validity of his *self-consistent* algorithm, Turnbull imposes a sufficient condition to the data. This condition essentially guarantees an interior solution to the constrained maximum likelihood problem. As we will see later, this condition is not always satisfied in typical contingent valuation data. In this paper, we seek to generalize Turnbull's (1974) algorithm to cover the new data setting. The generalization takes a different direction as compared with Turnbull (1976). We also discuss the validity of our algorithm when the corner solution is encountered.

The rest of the paper is organized as follows. In section 2 we begin with some definitions of different grouping mechanisms. We also review various estimation methods proposed for different data settings. In section 3 we describe why the available methods cannot be used for some data sets with *across-interval-censored* observations and present a *generalized self-consistent* algorithm that applies to these cases. Section 4 discusses the validity of the algorithm under possible corner solutions. Section 5 presents an application to the data setting motivated by contingent valuation studies. Field data results are reported in section 6. The last section concludes and discusses

---

[3] For a comprehensive review of the contingent valuation method see Cummings *et al* (1986) and Mitchell and Carson (1989). For a critical view of the method see Hausman (1993) and McFadden (1994).

some directions for further research.

2 Previous Methods

In this section we briefly introduce previous methods used in the estimation of a distribution function with *grouped* data. We begin with some definitions of data grouping.

Let $\{X_1, X_2, ..., X_n\}$ be a random sample of size n from an absolutely continuous distribution function $F(x)$ whose support is an interval $\Omega = (a, b) \subset \mathbf{R}$. The researcher does not see the exact $\{X_i\}$ but some nonlinear transformation Y_i.

Definition 1 *An observation Y_i is* grouped *if for some set $A_i \subset \mathbf{\Omega}$,*

$$Y_i = 1_{X_i \in A_i}. \tag{1}$$

In this paper we only consider the cases where A_i is some interval. Specifically, suppose that there exists a set of W real numbers $t_1, t_2, ..., t_W$ such that $a = t_0 < t_1 < ... < t_W < t_{W+1} = b$. Then,

Definition 2 *If $A_i = (a, t_j]$, then Y_i is a "left-censored" observation; if $A_i = (t_j, b)$, then Y_i is a "right-censored" observation; if $A_i = (t_{j-1}, t_j]$, then Y_i is an "interval-censored" observation, and if $A_i = (t_j, t_k]$ for $j < k$, then Y_i is an "across-interval-censored" observation.*

In addition we use the following notation. For $j, k = 1, 2, ..., W$ and $j < k$, let

- $\lambda_j = \sum_{i=1}^n 1_{X_i > t_j}$ be the number of observations whose value is known to fall above t_j;
- $\mu_j = \sum_{i=1}^n 1_{X_i \leq t_j}$ be the number of observations whose value is known to fall below t_j;
- $\delta_j = \sum_{i=1}^n 1_{t_{j-1} < X_i \leq t_j}$ be the number of observations whose value is known to fall in the interval $(t_{j-1}, t_j]$; and in general,

- $\gamma_{jk} = \sum_{i=1}^{n} 1_{t_j < X_i \leq t_k}$ be the number of observations whose value is known to fall in the interval $(t_j, t_k]$. It is clear that $\gamma_{j,W+1} = \lambda_j$, $\gamma_{0j} = \mu_j$ and $\gamma_{j-1,j} = \delta_j$.

In general, we have the following matrix of sufficient statistics:

$$\begin{bmatrix} t_1 & \lambda_1 & \gamma_{01} & - & - & \cdots & - & - \\ t_2 & \lambda_2 & \gamma_{02} & \gamma_{12} & - & \cdots & - & - \\ t_3 & \lambda_3 & \gamma_{03} & \gamma_{13} & \gamma_{23} & \cdots & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ t_{W-1} & \lambda_{W-1} & \gamma_{0,W-1} & \gamma_{1,W-1} & \gamma_{2,W-1} & \cdots & \gamma_{W-2,W-1} & - \\ t_W & \lambda_W & \gamma_{0W} & \gamma_{1W} & \gamma_{2W} & \cdots & \gamma_{W-2,W} & \gamma_{W-1,W} \end{bmatrix} \tag{2}$$

It is easy to see that with the data information (2), we can at most identify and estimate W points of the cumulative distribution function $F_j = F(t_j)$ or in terms of the survivor function $S_j \equiv 1 - F(t_j)$, for $j = 1, 2, ..., W$. In practice, W is fixed as the sample size n increases. It is in this sense that the *grouped* data restrict the model to be fundamentally parametric with fixed dimension. [4]

When the data contain only *interval-* and *right-censored* observations, i.e. $\gamma_{0j} = \mu_j = 0$ and $\gamma_{jk} = 0$ for $k > j + 1$, Kaplan and Meier (1958) propose the following product limit estimate for the survivor function:

$$\begin{cases} \hat{S}_1 = 1 - \frac{\delta_1}{n_1} \\ \\ \hat{S}_j = \left[1 - \frac{\delta_j}{n_j}\right] \hat{S}_{j-1}, \; j = 2, 3, ..., W, \end{cases} \tag{3}$$

where $n_j = \sum_{k=j}^{W} (\lambda_k + \delta_k)$ is the number of observations at risk at t_j.

When the data contain only *right-* and *left-censored* observations, i.e. $\gamma_{j-1,j} = \delta_j = 0$ and $\gamma_{jk} = 0$ for $j \neq 0$, Ayer *et al* (1955) propose a "pooling adjacent violators" algorithm. Their method suggests the use of the unconstrained maximum likelihood estimate $\tilde{S}_j$ as long as it satisfies the constraint $1 \geq \tilde{S}_1 \geq ... \geq \tilde{S}_W \geq 0$, where

$$\tilde{S}_j = \frac{\lambda_j}{\mu_j + \lambda_j}, \; j = 1, 2, ..., W. \tag{4}$$

---

[4] For a detailed discussion of identification issues with *grouped* data in semi-parametric duration analysis see Kiefer (1988), Ryu (1994) and An (1996a).

If the sequence $\tilde{S}_1, ..., \tilde{S}_W$ is not monotonically decreasing, the algorithm pools the adjacent violators. In other words, if $\tilde{S}_j < \tilde{S}_{j+1}$, then let $\bar{S}_j = \bar{S}_{j+1}$, where

$$\bar{S}_j = \frac{\lambda_j + \lambda_{j+1}}{\lambda_j + \lambda_{j+1} + \mu_j + \mu_{j+1}}. \tag{5}$$

This procedure is repeated until a monotonic sequence emerges. The surprising result is that this "pooling adjacent violators" algorithm produces the maximum likelihood estimates.

When the data contains *right-censored*, *left-censored* and *interval-censored* observations, i.e. $\gamma_{jk} = 0$ for $k > j+1$, Turnbull (1974) proposes an iterative *self-consistent* algorithm.[5] The main idea is to *reallocate* the μ_j's into an adjusted number of *interval-censored* observations defined by

$$\delta'_j = \delta_j + \sum_{k=j}^{W} \mu_k \frac{S^0_{j-1} - S^0_j}{1 - S^0_k}, \; 1 \leq j \leq W, \tag{6}$$

where the S^0_j's are the trial values of S_j. Notice that the adjustment factor $(S^0_{j-1} - S^0_j)/(1 - S^0_k)$ is the estimate of $P[t_{j-1} < X \leq t_j | X \leq t_k]$. The next step is to apply the Kaplan-Meier estimation to this "adjusted" data set where δ_j is replaced by δ'_j and μ_j by 0. With the new trial values for the S_j's, the process is repeated until it converges. Therefore, Turnbull's *self-consistent* algorithm is the solution to the following system of equations:

$$\begin{cases} \hat{S}_1 = 1 - \frac{\delta'_1}{n'_1} \\ \\ \hat{S}_j = \left[1 - \frac{\delta'_j}{n'_j}\right] \hat{S}_{j-1}, \; j = 2, 3, ..., W, \end{cases} \tag{7}$$

where

$$n'_j = \sum_{k=j}^{W} (\lambda_k + \delta'_k),$$

and δ'_j is given by expression (6) with $\hat{S}_l$ replacing $\hat{S}^0_l$.

---

[5] The *self-consistent* algorithm is a variant of the EM algorithm proposed by Dempster *et al* (1977).

Turnbull shows that if

$$\lambda_W > 0,\ \delta_j > 0 \qquad \forall\ j, \tag{8}$$

any solution $\{\hat{S}_j, 1 \le j \le W\}$ to the system of equations given by expression (7) also satisfies the first order conditions from the unconstrained maximization of the log-likelihood function; therefore, is the maximum likelihood estimate. [6]

As a generalization of his *self-consistent* algorithm, Turnbull (1976) considers the estimation of the distribution function $F(x)$ when the data are *arbitrarily grouped* into a set of multiple disjoint closed intervals. [7] In particular, he examines the case in which X_i is censored into A_i, where

$$A_i = \bigcup_{j=1}^{k_i} [L_{ij}, R_{ij}],\ i = 1, 2, ..., N.$$

The data must be grouped in such a way that it is possible to construct a set of disjoint intervals $[l_1, r_1], [l_2, r_2], ..., [l_W, r_W]$ for $W << n$ where the left and right end points are respectively

$$l_j \in \{L_{ij}; 1 \le j \le k_i, 1 \le i \le n\},$$

and

$$r_j \in \{R_{ij}; 1 \le j \le k_i, 1 \le i \le n\},$$

such that

$$l_1 \le r_1 < l_2 \le r_2 < ... < l_W \le r_W. \tag{9}$$

For this type of *arbitrarily grouped* data, Turnbull describes an algorithm for obtaining the maximum likelihood estimates of the density function which is an extension of the idea he used in 1974 for *doubly-censored* data.

---

[6] For a comprehensive study of the nonparametric maximum likelihood estimation for *interval-censored* data, see Groeneboom and Wellner (1992).

[7] Turnbull (1976) also considers truncated observations.

3 Generalized Self-consistent Algorithm

In this section we propose an estimation method for a data grouping mechanism which has not been considered by any of the available methods presented before. In particular, we examine the case in which the researcher observes both *right-* and *across-interval-censored* observations. [8]

Notice that *across-interval-censored* observations do not necessarily fit into Turnbull's (1976) framework. In particular, if $\gamma_{ij} > 0$ and $\gamma_{kl} > 0$ for $i < k < j < l$, then a set of disjoint closed intervals as the one described by expression (9) cannot be obtained. Both Turnbull (1976) and the *generalized self-consistent* algorithm proposed here can be considered as different lines of generalization of the *doubly-censored* problem addressed in Turnbull (1974).

The algorithm maintains the same *reallocation* idea introduced by Turnbull (1974). In our case, however, the *across-interval-censored* data are reallocated into δ''_j, where:

$$\delta''_j = \sum_{i=0}^{j-1} \sum_{k=j}^{W} \gamma_{ik} \frac{S_{j-1} - S_j}{S_i - S_k}, \ 1 \leq j \leq W. \tag{10}$$

Notice that in this case the adjustment factor is given by the conditional probability $P(t_{j-1} < X \leq t_j | t_i < X \leq t_k)$. We can now apply the Kaplan-Meier estimation to these *singly-censored* data to derive a new set of values of $\{S_j\}$. At the convergence, $\{\hat{S}_j\}$ satisfies

$$\begin{cases} \hat{S}_1 = 1 - \frac{\delta''_1}{n''_1} \\ \\ \hat{S}_j = \left[1 - \frac{\delta''_j}{n''_j}\right] \hat{S}_{j-1}, \ j = 2, 3, ..., W, \end{cases} \tag{11}$$

where

$$n''_j = \sum_{k=j}^{W} (\lambda_k + \delta''_k)$$

and δ''_j is given by expression (10) with $\hat{S}_l$ replacing S_l.

---

[8] See section 5 for a further discussion of the cases where contingent valuation data generate *across-interval-censored* observations.

The algorithm provides an iterative procedure to solve this system of equations. Given a set of starting values $\{S^{\,0}_{j}, 1 \leq j \leq W\}$ that satisfies the restriction $1 > S^{\,0}_{1} > S^{0}_{2} > ... > S^{\,0}_{W} > 0$, the algorithm is the following: [9]

1) Calculate

$$\delta''_j = \sum_{i=0}^{j-1} \sum_{k=j}^{W} \gamma_{ik} \frac{S^0_{j-1} - S^0_j}{S^0_i - S^0_k}, \; 1 \leq j \leq W. \tag{12}$$

We can rewrite expression (12) in the following form:

$$\delta''_j = \gamma_{j-1,j} + \sum_{k=j}^{W} \gamma_{0k} \frac{S^0_{j-1} - S^0_j}{1 - S^0_k}$$

$$+ \sum_{k=j+1}^{W} \gamma_{j-1,k} \frac{S^0_{j-1} - S^0_j}{S^0_i - S^0_k} + \sum_{i=1}^{j-2} \sum_{k=j}^{W} \gamma_{ik} \frac{S^0_{j-1} - S^0_j}{S^0_i - S^0_k}, \; 3 \leq j \leq W, \tag{13}$$

or

$$\delta''_j = \delta_j + \sum_{k=j}^{W} \mu_k \frac{S^0_{j-1} - S^0_j}{1 - S^0_k}$$

$$+ \sum_{k=j+1}^{W} \gamma_{j-1,k} \frac{S^0_{j-1} - S^0_j}{S^0_i - S^0_k} + \sum_{i=1}^{j-2} \sum_{k=j}^{W} \gamma_{ik} \frac{S^0_{j-1} - S^0_j}{S^0_i - S^0_k}, \; 3 \leq j \leq W. \tag{14}$$

Comparing expressions (6) and (14) we can clearly see that they are equivalent except for the last two terms which are the weighted sum of the *across-interval-censored* observations.

2) Calculate

$$\begin{cases} S^1_1 = 1 - \frac{\delta''_1}{n''_1} \\ \\ S^1_j = \left[1 - \frac{\delta''_j}{n''_j}\right] S^1_{j-1}, \; j = 2, 3, ..., W, \end{cases} \tag{15}$$

where $n''_j = \sum_{k=j}^{W} (\lambda_k + \delta''_k)$.

3) Return to step 1 and replace $\{S^{\,0}_{j}, 1 \leq j \leq W\}$ with $\{S^{\,1}_{j}, 1 \leq j \leq W\}$, etc.

4) Stop when the required accuracy has been achieved.

---

[9] In what follows, we define $S^{\,0}_{0} = 1$.

An alternative to the algorithm is to construct a distribution-free log-likelihood function which is given by

$$l = \sum_{i=0}^{W-1} \sum_{k=i+1}^{W} \gamma_{ik} \log(S_i - S_k) + \sum_{i=1}^{W} \lambda_i \log S_i. \tag{16}$$

The maximum likelihood estimates are those values of S_j which maximize l subject to the constraint

$$1 \geq S_1 \geq ... \geq S_W \geq 0. \tag{17}$$

The unconstrained maximization of expression (16) gives the following first order conditions:

$$\frac{\partial l}{\partial S_W} = -\sum_{i=0}^{W-1} \frac{\gamma_{iW}}{S_i - S_W} + \frac{\lambda_W}{S_W} = 0 \tag{18}$$

$$\frac{\partial l}{\partial S_{W-1}} = -\sum_{i=0}^{W-2} \frac{\gamma_{i,W-1}}{S_i - S_{W-1}} + \frac{\gamma_{W-1,W}}{S_{W-1} - S_W} + \frac{\lambda_{W-1}}{S_{W-1}} = 0 \tag{19}$$

$$\vdots$$

$$\frac{\partial l}{\partial S_{j+1}} = -\sum_{i=0}^{j} \frac{\gamma_{i,j+1}}{S_i - S_{j+1}} + \sum_{k=j+2}^{W} \frac{\gamma_{j+1,k}}{S_{j+1} - S_k} + \frac{\lambda_{j+1}}{S_{j+1}} = 0 \tag{20}$$

$$\frac{\partial l}{\partial S_j} = -\sum_{i=0}^{j-1} \frac{\gamma_{ij}}{S_i - S_j} + \sum_{k=j+1}^{W} \frac{\gamma_{jk}}{S_j - S_k} + \frac{\lambda_j}{S_j} = 0 \tag{21}$$

$$\vdots$$

$$\frac{\partial l}{\partial S_2} = -\frac{\gamma_{02}}{1 - S_2} - \frac{\gamma_{12}}{S_1 - S_2} + \sum_{k=3}^{W} \frac{\gamma_{2k}}{S_2 - S_k} + \frac{\lambda_2}{S_2} = 0 \tag{22}$$

$$\frac{\partial l}{\partial S_1} = -\frac{\gamma_{01}}{1 - S_1} + \sum_{k=2}^{W} \frac{\gamma_{1k}}{S_1 - S_k} + \frac{\lambda_1}{S_1} = 0 \tag{23}$$

In order to demonstrate that the algorithm gives maximum likelihood estimates, we are going to prove that the solution to the system of equations given by (11) satisfies the first order conditions given by (18) through (23). This result is stated in the following theorem which is a direct generalization of the theorem proved in Turnbull (1974):

Theorem. *Let $\hat{S} = (\hat{S}_1, ..., \hat{S}_W)$ be a solution to the system of equations (11). Then,* *(i)* $1 \geq \hat{S}_1 \geq \hat{S}_2 \geq ... \geq \hat{S}_W \geq 0$*;*

(ii) If $1 > \hat{S}_1 > \hat{S}_2 > ... > \hat{S}_W > 0$, *then* $\hat{S}$ *also solves the system of equations (18) to (23). Hence, it is the maximum likelihood estimate.*

Proof. See the appendix.

4 The Validity of the algorithm under Corner Solutions

Theorem 1 asserts that for interior solutions, i.e. $1 > \hat{S}_1 > \hat{S}_2 > ... > \hat{S}_W > 0$, the algorithm produces maximum likelihood estimates. In this section we discuss the consequences of having corner solutions and the validity of the algorithm when the *strict inequality condition* is not satisfied.

A corner solution exists when the constraint (17) is satisfied with equality for some j, i.e. $\hat{S}_j = \hat{S}_{j-1}$. From expression (11), we can write:

$$\hat{S}_j = \frac{n''_{j+1} + \lambda_j}{n''_{j+1} + \lambda_j + \delta''_j} \hat{S}_{j-1}. \tag{24}$$

Therefore, $\hat{S}_j = \hat{S}_{j-1}$ if and only if $\delta''_j = 0$. Notice that $\gamma_{j-1,j} = \delta_j = 0$ is a necessary condition for $\delta''_j = 0$.[10] In fact, when proving the validity of his *self-consistent* algorithm, Turnbull (1974) rules out the possibility of corner solutions by imposing on the data the condition that $\delta_j > 0$ for all j.

The main result of this paper is that the algorithm always provides maximum likelihood estimates. The proof of theorem 1 shows that this is the case for interior solutions. For corner solutions, however, we provide a heuristic interpretation of why algorithm provides maximum likelihood estimates.

The fundamental logic that makes the algorithm work is its connection to the Kaplan-Meier estimation. As mentioned before, when the data consist of only *right-* and *interval-censored* observations, this method produces maximum likelihood estimates. Originally, it deals only with grouped data for which $\delta_j > 0$. However, if $\delta_j = 0$

---

[10] The adjustment factor for $\gamma_{j-1,j} = \delta_j$ is always equal to 1. See expressions (13) and (14).

for some j, it would act as if the interval (t_{j-1}, t_j) does not exist and take (t_{j-1}, t_{j+1}) as a new interval.[11] The algorithm essentially applies Kaplan-Meier estimation to the correctly adjusted data where the *left-* and *across-interval-censored* data have been reallocated into an adjusted number of *interval-censored* observations (δ_j''). If $\delta_j'' = 0$, Kaplan-Meier's method would suggest *pooling*. That produces $\hat{S}_j = \hat{S}_{j-1}$ as a corner solution. Therefore, the algorithm provides maximum likelihood estimates as well.

Additionally, the following lemma provides necessary and sufficient conditions for corner solutions.

Lemma. *If* $\hat{S}_j = \hat{S}_{j+1}$*, then* $\partial l(\hat{S})/\partial S_j = -\partial l(\hat{S})/\partial S_{j+1}$.

Proof. See the appendix.

Figures 1a and 1b give us more insight on this result. The axes represent S_j and S_{j+1}. The 45 degree line bisects the unit square into two areas: area 1 where $S_j > S_{j+1}$ and area 2 where $S_j < S_{j+1}$. If the likelihood function reaches its global maximum at point A in area 1 (figure 1a), the *strict inequality condition* is satisfied and therefore $\partial l/\partial S_j = \partial l/\partial S_{j+1} = 0$. An alternative is that the likelihood function reaches its maximum at point B in area 2 (figure 1b), where $S_j < S_{j+1}$. The solution in this case will be given by B' which satisfies the constraint with equality. At B', the slope is non-increasing in the S_j direction ($\partial l/\partial S_j \leq 0$) and non-decreasing in the S_{j+1} direction ($\partial l/\partial S_{j+1} \geq 0$).

To support this interpretation, we did a number of numerical examples and found that the algorithm always produces the constrained maximum likelihood estimates. In particular, when corner solutions were encountered, the first derivatives of the likelihood function with respect to the violators alternate in sign and are equal in magnitude as stated in the lemma. These empirical results are shown in section 6.

5 Application to Contingent Valuation

This section presents an application of the algorithm to a data grouping mechanism found in contingent valuation (contingent valuation) studies. One of the most fre-

---

[11] This is called *pooling* in Ayer *et al*'s terminology.

quently used elicitation methods in contingent valuation surveys is the *dichotomous choice* (DC) question. [12] A DC question asks the respondent whether he or she would be willing to pay a specified amount of money to obtain a change in the level of provision of a public good. If the respondent answers "yes," that person has indicated a willingness to pay (WTP) that is greater or equal to the specified amount. If the respondent answers "no," the offered amount is considered as an upper bound on the true willingness to pay. We refer to this format as a "single-bounded dichotomous choice" (SBDC) question.

When there is a follow-up question which depends on the response given to the first offered bid, the format is called a "double-bounded dichotomous choice" (DBDC) question. In this case, the follow-up question poses a higher amount if the respondent answered "yes" to the original bid and a lower amount if a "no" answer was elicited. [13]

A common practice in contingent valuation DBDC studies is to pre-specify a small number (K) of versions of the questionnaire. Each version differs by the dollar amount of each bid and is randomly assigned to the respondents. Let $k = 1, 2, ..., K << N$ represent each version of the questionnaire. Let A_1^k, A_2^k and A_3^k respectively be the lowest, original and highest offered bid amounts associated with version k, where $A_1^j > A_1^k$ for $j < k$.

We refer to the "overlapping bids" case (OBC) as the situation where $A_2^k = A_1^{k+1}$ and $A_3^k = A_2^{k+1}$ for $k = 1, 2, ..., K - 1$. [14] The opposite situation is the "distinct bids" case (DBC), which occurs when $A_2^k \neq A_1^{k+1}$ and $A_3^k \neq A_2^{k+1}$ for $k = 1, 2, ..., K - 1$. A combination of these two cases is called "mixed bids" case (MBC), i.e. when some bids overlap while others are distinct. Let $0 = t_0 \leq t_1 \leq t_2 \leq ... \leq t_W$ be the ordered permutation of the A_m^k's ($m = 1, 2, 3$), where $W = K + 2$ for the OBC, $W = 3 \times K$

---

[12] Bishop and Heberlein (1979) is the first study to apply this type of question. Hanemann (1984) and Cameron and James (1987) develop the economic theory that underlies the SBDC format.

[13] An and Ayala (1995) advocate asking a follow-up question of the type "are you willing to pay anything for the good" only to the "no-no" respondents. This "zero WTP" question allows the researcher to identify consumers who are indifferent to the good that is being valued. Alternatively, for a SBDC format, Kriström (1995) and Haab (1995) suggested asking this zero WTP question to all respondents as the first elicitation question.

[14] See Carson *et al* (1994a, 1994b, 1994c) and Imber *et al* (1991) for examples.

for the DBC, and $K+2<W<3\times K$ for the MBC.

The type of information provided by an OBC is *right-*, *left-*, and *interval-censored.* Therefore we can apply either Turnbull's (1974) or Turnbull's (1976) algorithms. Due to the nature of the OBC data, $\gamma_{j-1,j}=\delta_j>0$ for all j. As it was mentioned in the last section, this characteristic guarantees that the *strict inequality condition* be automatically satisfied.

For the DBC and MBC, neither of the algorithms suggested by Turnbull can be applied. For "overlapping" *across-interval-censored* observations, i.e. $\gamma_{ij}>0$, $\gamma_{kl}>0$ for $i<k<j<l$, the algorithm is directly applicable. The numerical results presented in the next section show that the algorithm produces the same results as the constrained maximum likelihood estimates.

6 Empirical Results

The data we used comes from the contingent valuation study conducted by Hanemann, Loomis, and Kanninen (1991) to elicit the WTP for protecting wetland habitats and wildlife in California's San Joaquin Valley. [15] This survey was a combination of mail and telephone media. Households were chosen based on random digit sampling. Each household was asked if it was willing to participate in a survey. If it accepted, a mail questionnaire was sent and a certain time was arranged for the interviewer to call the household back. The survey used DBDC questions. [16]

The interview process was conducted in May 1989 in three geographical areas: (1) the San Joaquin Valley, (2) the rest of the state of California, and (3) the states of Oregon, Washington and Nevada as representatives of the "out-of-state" population. Of the 1960 households originally contacted, 1239 (63.1%) agreed to participate. From these, only 1004 completed interviews were collected: 227 from the San Joaquin

---

[15] In fact, Hanemann *et al* evaluated five different environmental programs in their survey. We focus only on the first of these programs: protection of wetlands and wildlife.

[16] The sets of bids used in this study are: (25,40,80), (25,50,110), (30,65,125), (40,80,125), and (55,110,170) where the first element of each set corresponds to the lowest bid, the second element to the original bid, and the third element to the highest bid.

Valley, 576 from the rest of California, and 201 from “out-of-state”. In this empirical application we used the same data that Hanemann *et al* used for their paper: the 576 interviews from the rest of the California sample. [17]

The results presented in this section were calculated using two methods: the algorithm and the constrained maximum likelihood estimation. In every case, the application of both methods led us to the same results. [18]

In order to estimate the survivor function for data with “distinct bids,” we consider two versions of the San Joaquin study's sets of bids for which the condition $A_2^1 \neq A_1^2$ and $A_3^1 \neq A_2^2$ holds. The chosen sets were (25,50,110) and (30,65,125).

The estimates are presented in table 1. The first column presents the unconstrained maximum likelihood estimates. They were calculated analytically from the first order conditions obtained by maximizing the log-likelihood function (16) for $W = 2$. This estimation, however, does not provide a monotonic sequence since $1 > \tilde{S}_1 > \tilde{S}_2 > \tilde{S}_3 > \tilde{S}_4 > \tilde{S}_5 < \tilde{S}_6 > 0$. Ayer *et al*'s pooling idea might suggest that the constrained estimates would pool $\tilde{S}_5$ and $\tilde{S}_6$ into a new estimate. Nevertheless, as it is shown in the second column, the constrained maximum likelihood estimation pools not only $\hat{S}_5 = \hat{S}_6$, but also $\hat{S}_2 = \hat{S}_3$. The algorithm provides the same results as the constrained maximum likelihood estimation.

It is easy to show that for these two cases, the score of the log-likelihood function alternates in sign and is equal in absolute value, as stated in lemma 1. In particular, $\partial l/\partial S_2 = -\partial l/\partial S_3 = -4.67$ and $\partial l/\partial S_5 = -\partial l/\partial S_6 = -19.06$.

We refer as “mixed bids” to those cases where some bids overlap while others are distinct. The complete set of bids for the San Joaquin study are such an example. The data are summarized in Table 2.

Notice that Turnbull's *self-consistent* algorithm cannot be applied in this case since we have “overlapping” *across-interval-censored* observations. The only algorithm available to nonparametrically estimate the survivor function is the algorithm

---

[17] The valid number of observations for the program we used in this application is 569. We dropped 7 “don't know-don't know-don't know” observations which did not provide any information.

[18] We used the maximum likelihood application of GAUSS386i. The program for the algorithm was also written in GAUSS and it is available from the authors upon request.

proposed in this paper. The estimation results are presented in table 3.

7 Conclusions

Environmental economists who conduct contingent valuation studies rely heavily on the nonparametric estimation of a distribution function from grouped data. In this estimation, the *self-consistent* algorithm of Turnbull (1974, 1976) plays an instrumental role. However, there is a data grouping mechanism found in some contingent valuation surveys to which Turnbull's method does not apply. We refer to these cases as distinct bids and mixed bids, where *across-interval-censored* observations are encountered.

In this paper we present a *generalized self-consistent algorithm* that estimates the survivor function for *right-* and *across-interval-censored* data. It is based on Turnbull's (1974) *reallocation* idea that reduces this *multiply-censored* problem to a *singly-censored* one so that the Kaplan-Meier estimation method can be applied.

The main result of this paper is that the algorithm produces maximum likelihood estimates. We prove analytically that this is the case for interior solutions. Unlike Turnbull (1974, 1976), we provide a heuristic interpretation of the validity of the algorithm under corner solutions. The numerical applications support our hypothesis. The algorithm gives the same results as the constrained maximum likelihood estimation for both interior and corner solutions. Moreover, it is shown that the algorithm satisfies the Kuhn-Tucker conditions when the estimates are pooled, i.e. when $\hat{S}_j = \hat{S}_{j+1}$.

Computationally, it is easier to program the algorithm than the constrained maximization problem. This ease in programming saves both time and resources. A potential drawback of the proposed method, however, is that it does not provide the estimates for standard errors. Further research can be done in this area with the help of resampling or bootstrapping techniques.

This paper leaves out an important extension. Economic data are nonexperimental in nature. It is very important to control for the effects of observed personal characteristics. An (1996b) addresses the estimation of the willingness to pay distribution which depends on a vector of covariates. In that semiparametric setting, the *generalized self-consistent algorithm* proposed here plays an instrumental role.

8 Appendix

Proof of the Theorem. Part (1) follows directly from the property of Kaplan-Meier estimates. We now show that the solution to the system of equations given by (11) also satisfies the first order conditions for the log-likelihood function (16). In other words, $\partial l(\hat{S})/\partial S_l = 0$ for $l = 1, ..., W$. We do this first for $l = W$ and then proceed by induction to show it is true for $l = W-1, W-2, ..., 1$.

Assume $\lambda_W > 0$. By assumption, $1 > \hat{S}_1 > ... > \hat{S}_{W-1} > \hat{S}_W > 0$. From (11) we can write

$$\hat{S}_W = \hat{S}_{W-1} \frac{\lambda_W}{\lambda_W + \delta''_W}.$$

Substituting for δ''_W we obtain

$$\hat{S}_W = \hat{S}_{W-1} \frac{\lambda_W}{\lambda_W + \sum_{i=0}^{W-1} \frac{\gamma_{iW}}{\hat{S}_i - \hat{S}_W}(\hat{S}_{W-1} - \hat{S}_W)},$$

or

$$\frac{\lambda_W}{\hat{S}_W} - \sum_{i=0}^{W-1} \frac{\gamma_{iW}}{\hat{S}_i - \hat{S}_W} = 0. \tag{25}$$

Expression (25) is the first order condition for $l = W$, i.e. $\partial l(\hat{S})/\partial S_W = 0$.

Assume $\partial l(\hat{S})/\partial S_l = 0$ for $l = W, W-1, ..., j+1$. Then, from (11) we obtain

$$\hat{S}_j = \hat{S}_{j-1} \frac{n''_j - \delta''_j}{n''_j} = \hat{S}_{j-1} \frac{n''_{j+1} + \lambda_j}{\lambda_j + \delta''_j + n''_{j+1}}.$$

Substituting for δ''_j we obtain

$$(\lambda_j + n''_{j+1})(\hat{S}_{j-1} - \hat{S}_j) = \hat{S}_j(\hat{S}_{j-1} - \hat{S}_j) \sum_{i=0}^{j-1} \sum_{k=j}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k},$$

or

$$\frac{\lambda_j}{\hat{S}_j} - \sum_{i=0}^{j-1} \sum_{k=j}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k} + \frac{n''_{j+1}}{\hat{S}_j} = 0. \tag{26}$$

We can rewrite expression (26) as

$$\frac{\lambda_j}{\hat{S}_j} - \sum_{i=0}^{j-1} \frac{\gamma_{ij}}{\hat{S}_i - \hat{S}_j} - \sum_{i=0}^{j-1} \sum_{k=j+1}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k} + \frac{n''_{j+1}}{\hat{S}_j} = 0,$$

or

$$\frac{\lambda_j}{\hat{S}_j} - \sum_{i=0}^{j-1} \frac{\gamma_{ij}}{\hat{S}_i - \hat{S}_j} + \sum_{k=j+1}^{W} \frac{\gamma_{jk}}{\hat{S}_j - \hat{S}_k} - \sum_{i=0}^{j} \sum_{k=j+1}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k} + \frac{n''_{j+1}}{\hat{S}_j} = 0. \tag{27}$$

Notice that the first three terms of expression (27) correspond exactly to the first order condition for $l = j$, i.e. $\partial l(\hat{S})/\partial S_j = 0$. Therefore, we need to prove that the last two terms of (27) are identical to zero:

$$\sum_{i=0}^{j-1} \sum_{k=j+1}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k} - \frac{n''_{j+1}}{\hat{S}_j} = 0. \tag{28}$$

First, notice that from (11) and the relation $n''_j = n''_{j+1} + \lambda_j + \delta''_j$, we can write

$$\frac{n''_j}{\hat{S}_{j-1}} = \frac{\lambda_j + n''_{j+1}}{\hat{S}_j}.$$

Therefore,

$$\begin{aligned} \frac{n''_{j+1}}{\hat{S}_j} &= \frac{\lambda_{j+1} + n''_{j+2}}{\hat{S}_{j+1}} \\ &= \frac{\lambda_{j+1}}{\hat{S}_{j+1}} + \frac{\lambda_{j+2}}{\hat{S}_{j+2}} + \frac{n''_{j+3}}{\hat{S}_{j+2}} \\ &= \frac{\lambda_{j+1}}{\hat{S}_{j+1}} + \frac{\lambda_{j+2}}{\hat{S}_{j+2}} + \ldots + \frac{\lambda_{W-2}}{\hat{S}_{W-2}} + \frac{\lambda_{W-1}}{\hat{S}_{W-1}} + \frac{n''_W}{\hat{S}_{W-1}}. \end{aligned}$$

But as $n''_W/\hat{S}_{W-1} = \lambda_W/\hat{S}_W$, we obtain

$$\frac{n''_{j+1}}{\hat{S}_j} = \sum_{k=j+1}^{W} \frac{\lambda_k}{\hat{S}_k}. \tag{29}$$

Using this last result, expression (28) can now be written as

$$\sum_{k=j+1}^{W} \frac{\lambda_k}{\hat{S}_k} - \sum_{i=0}^{j-1} \sum_{k=j+1}^{W} \frac{\gamma_{ik}}{\hat{S}_i - \hat{S}_k} = 0. \tag{30}$$

It is not hard to show that expression (30) is exactly equivalent to the sum of the first order conditions for $l = W, W-1, ..., j+1$ which by the induction hypothesis equals zero, e.g. $\sum_{l=j+1}^{W} \partial l(\hat{S})/\partial S_l = 0$. This completes the proof. Q.E.D.

Proof of the Lemma. The objective is to maximize the likelihood function l given by expression (16) subject to $1 \geq S_1 \geq S_2 \geq ... \geq S_W \geq 0$. This constraint can be written as a set of $W+1$ inequalities as follows:

$$\begin{aligned} 1 - S_1 &\geq 0 \\ S_1 - S_2 &\geq 0 \\ &\vdots \\ S_{j-1} - S_j &\geq 0 \\ S_j - S_{j+1} &\geq 0 \\ &\vdots \\ S_{W-1} - S_W &\geq 0 \\ S_W &\geq 0 \end{aligned}$$

Let y_l for $l = 1, ..., W+1$ be the lagrangian multipliers. The Kuhn-Tucker sufficient conditions for j and $j+1$ are:

$$\begin{aligned} \frac{\partial l}{\partial S_j} - \hat{y}_j + \hat{y}_{j+1} &= 0 \\ \frac{\partial l}{\partial S_{j+1}} - \hat{y}_{j+1} + \hat{y}_{j+2} &= 0 \\ \hat{y}_j(\hat{S}_{j-1} - \hat{S}_j) &= 0 \\ \hat{y}_{j+1}(\hat{S}_j - \hat{S}_{j+1}) &= 0 \\ \hat{y}_{j+2}(\hat{S}_{j+1} - \hat{S}_{j+2}) &= 0 \\ \hat{y}_j, \hat{y}_{j+1} &\geq 0 \end{aligned}$$

Assume $\hat{S}_{j-1} > \hat{S}_j > 0$, which implies that $\hat{y}_j = 0$.

First, consider the case in which the constraint for $l = j$ is satisfied with strict inequality, i.e. $\hat{S}_j > \hat{S}_{j+1}$. This implies that $\hat{y}_{j+1} = 0$, and therefore, $\partial l/\partial S_j = 0$.

Consider now the case in which the constraint is satisfied with equality, i.e. $\hat{S}_j = \hat{S}_{j+1}$. This implies that $\hat{y}_{j+1} \geq 0$, and therefore, $\partial l/\partial S_j = -\hat{y}_{j+1} \leq 0$.

If, in addition, $\hat{S}_{j+1} > \hat{S}_{j+2}$, then $\hat{y}_{j+2} = 0$, and therefore, $\partial l/\partial S_{j+1} = \hat{y}_{j+1} \geq 0$. Q.E.D.

9 References

An, Mark Y. (1996a): "A Proportional Hazard Model for Bivariate Durations and Statistical Inference with Grouped Data," manuscript, Duke University.

An, Mark Y. (1996b): "Semiparametric Estimation of Willingness to Pay Distribution," manuscript, Duke University.

An, Mark Y. and Roberto A. Ayala (1995): "A Mixture Model of Willingness to Pay Distributions," *Working Paper No. 95-21*, Duke University.

Ayer, Miriam, H. D. Brunk, G. M. Ewing, W. T. Reid and Edward Silverman (1955): "An Empirical Distribution Function For Sampling with Incomplete Information," *Annals of Mathematical Statistics*, Vol. **26**, pp. 641-647.

Bishop, Richard C. and Thomas A. Heberlein (1979): "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?," *American Journal of Agricultural Economics*, Vol. **61**, No. 5, pp. 926-930.

Cameron, Trudy Ann and M. D. James (1987): "Efficient Estimation Methods for Use with 'Closed-Ended' Contingent Valuation Survey Data," *Review of Economics and Statistics*, Vol. **69**, pp. 269-276.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, J. A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud and V. Kerry Smith (1994a): "Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight," NOOAA contract No. 50-DGNC-1-00007.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser and Paul A. Ruud (1994b): "Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez," *Discussion Paper 94-18*, Resources for the Future.

Carson, Richard T., Leanne Wilks and David Imber (1994c): "Valuing the Preservation of Australia's Kakadu Conservation Zone," *Oxford Economic Papers*, Vol. **46**, pp. 727-749.

Cummings, R. G., D. S. Brookshire and W. D. Schulze (1986): *"Valuing Environmental Goods: An Assessment of the Contingent Valuation Method,"* Totowa NJ, Rowman & Allanheld.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977): "Maximum likelihood estimation from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, Vol. **39**, pp. 1-22.

Groeneboom, Piet and Jon A. Wellner (1992): *"Information Bounds and Nonparametric Maximum Likelihood Estimation,"* Basel-Boston-Berlin, Birkhäuser.

Haab, Timothy (1995): "The Distributional Impacts of Indifference on Dichotomous Choice Contingent Valuation," Working paper, East Carolina University, November 1995.

Hanemann, W. Michael (1984): "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, Vol. **66**, pp. 332-341.

Hanemann, W. Michael, John Loomis and Barbara Kanninen (1991): "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation," *American Journal of Agricultural Economics*, Vol. **73**, pp. 1255-1262.

Hausman, Jerry (1993): *"Contingent Valuation: a Critical Assessment,"* North-Holland, Amsterdam, New York.

Imber, David, Gay Stevenson and Leanne Wilks (1991): "A Contingent Valuation Survey of the Kakadu Conservation Zone," *Research Paper No. 3*, Resource Assessment Commission.

Kaplan, E. L. and P. Meier (1958): "Nonparametric Estimation for Incomplete Observations," *Journal of the American Statistical Association*, Vol. **53**, pp. 457-481.

Kiefer, Nicholas M. (1988): "Analysis of Grouped Duration Data," *Contemporary Mathematics*, Vol. **80**, pp. 107-137.

Kriström, Bengt (1995): "Spike Models in Contingent Valuation: Theory and Illustrations," invited paper to the First Toulouse Conference on Environmental and Resource Economics, Toulouse, France, March 30-31, 1995.

McFadden, Daniel (1994): "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, Vol. **76**, pp. 689-708.

Mitchell, Robert C. and Richard T. Carson (1989): *"Using Surveys to Value Public Goods: The Contingent Valuation Method,"* Washington DC, Resources for the Future.

Ryu, Keunkwan (1994): "Group Duration Analysis of the Proportional Hazard Model: Minimum Chi-squared Estimators and Specification Tests," *Journal of the American Statistical Association*, Vol. **89**, pp. 1386-1395.

Turnbull, Bruce (1974): "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data," *Journal of the American Statistical Association*, Vol. **69**, pp. 169-173.

Turnbull, Bruce (1976): "The Empirical Distribution Function with Arbitrary Grouped, Censored, and Truncated Data," *Journal of the Royal Statistical Society*, Series B, pp. 290-295.

Table 1: Nonparametric Estimates from two versions of a Double-Bounded Dichotomous Choice Question with Distinct Bids for the San Joaquin Study

Bid Amount t_j	**Unconstrained MLE** $(\tilde{S}_j)$	**Constrained MLE/algorithm** $(\hat{S}_j)$
25	**0.9077**	**0.9109** (0.0210)
30	**0.8571**	**0.8515** (0.0222)
50	**0.8462**	**0.8515** (0.0222)
65	**0.7539**	**0.7401** (0.0335)
110	**0.4231**	**0.4570** (0.0311)
125	**0.4921**	**0.4570** (0.0311)

- Standard deviations are in parenthesis

Table 2: Summary of the San Joaquin Study Data from a Double-Bounded Dichotomous Choice Setting with Mixed Bids

t_j	λ_j	γ_{0j}	γ_{1j}	γ_{2j}	γ_{3j}	γ_{4j}	γ_{5j}	γ_{6j}	γ_{7j}	γ_{8j}	γ_{9j}
25	0	28	-	-	-	-	-	-	-	-	-
30	0	18	0	-	-	-	-	-	-	-	-
40	0	18	5	0	-	-	-	-	-	-	-
50	0	0	8	0	0	-	-	-	-	-	-
55	0	15	0	0	0	0	-	-	-	-	-
65	0	0	0	13	0	0	0	-	-	-	-
80	73	0	0	0	42	0	0	0	-	-	-
110	55	0	0	0	0	55	12	0	0	-	-
125	106	0	0	0	0	0	0	33	22	0	-
170	48	0	0	0	0	0	0	0	0	19	0

Table 3: Nonparametric Estimates from a Double-Bounded Dichotomous Choice Setting with Mixed Bids for the San Joaquin Study

Bid Amount t_j	**Constrained MLE/algorithm** $(\hat{S}_j)$
25	**0.8984**
	(0.0148)
30	**0.8513**
	(0.0152)
40	**0.8513**
	(0.0152)
50	**0.8513**
	(0.0152)
55	**0.7410**
	(0.0249)
65	**0.7410**
	(0.0249)
80	**0.6613**
	(0.0256)
110	**0.5317**
	(0.0274)
125	**0.4625**
	(0.0155)
170	**0.3809**
	(0.0323)

- Standard deviations are in parenthesis