

Day, Brett

Working Paper

Distribution-free estimation with interval-censored contingent valuation data: Trobles with Turnbull?

CSERGE Working Paper EDM, No. 05-07

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Day, Brett (2005) : Distribution-free estimation with interval-censored contingent valuation data: Trobles with Turnbull?, CSERGE Working Paper EDM, No. 05-07, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/80251>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Distribution-free estimation with interval-
censored contingent valuation data:
Troubles with Turnbull?**

by

Brett Day

CSERGE Working Paper EDM 05-07

**Distribution-free estimation with interval-censored contingent valuation data:
Troubles with Turnbull?**

Brett Day

**Centre for Social and Economic Research on the Global Environment (CSERGE),
University of East Anglia, NR4 7TJ, UK**

Author contact details
email: brett.day@uea.ac.uk
tel: +44 (0)1603 591413

Acknowledgements

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the interdisciplinary research programme of the ESRC Research Centre for Social and Economic Research on the Global Environment (CSERGE).

The author is indebted to Professor Ian Bateman for his insightful comments on this manuscript and for providing one of the data sets used in the empirical investigations.

ISSN 0967-8875

Abstract

Contingent valuation (CV) surveys frequently employ elicitation procedures that return interval-censored data on respondents' willingness to pay (WTP). Almost exclusively, CV practitioners have applied Turnbull's self-consistent algorithm to such data in order to obtain nonparametric maximum likelihood (NPML) estimates of the WTP distribution. This paper documents two failings of Turnbull's algorithm; (1) that it may not converge to NPML estimates and (2) that it may be very slow to converge. With regards to (1) we propose new starting and stopping criteria for the algorithm that guarantee convergence to the NPML estimates. With regards to (2) we present a smorgasbord of alternative NPML estimators and demonstrate, through Monte Carlo simulations, their performance advantages over Turnbull's algorithm.

Key words: Interval-censored data, Turnbull's self-consistent algorithm, nonparametric maximum likelihood, contingent valuation

1. INTRODUCTION

Increasingly, practitioners of contingent valuation (CV) employ elicitation techniques that do not directly identify a respondent's willingness to pay (WTP) for the non-market good under consideration. Rather, many CV surveys collect WTP data in a format that only indicates a range of values within which a respondent's precise WTP is contained. For example, a single-bounded (SB) dichotomous choice question records whether a respondent's WTP is either above or below a particular value. Such data is said to demonstrate Type 1 interval-censoring. Moreover, when the SB question is followed-up by a second dichotomous choice question, as is the case with the one-and-one-half-bounded (OOHB) and double-bounded (DB) question formats, a respondent's WTP may also be recorded as falling in the interval between two particular values. Data of this kind demonstrates Type 2 interval-censoring.

This paper concerns itself with the problem of estimating the population distribution of WTP from a CV survey recording interval-censored WTP data. In general, this distribution is the primary object of concern in CV studies since it is from this distribution that policy-relevant quantities such as the population mean and median WTP can be estimated. In particular, we consider estimation techniques that do not impose *ad hoc* parametric assumptions on the distribution of WTP. Such non-parametric or distribution-free techniques make no assumptions concerning the population distribution of WTP and can be used to derive a lower-bound approximation to the population's expected WTP (Haab and McConnell, 1997; Boman et al., 1999).

When WTP data is not censored, the natural non-parametric estimator of the WTP distribution is provided by the empirical distribution function (EDF). The EDF is a step function that rises only at the WTP values reported by respondents in the sample. As will be shown subsequently, the equivalent function for censored data is the solution to a constrained maximum likelihood problem, generally referred to as the *non-parametric maximum likelihood* (NPML) problem. Like the EDF, NPML estimates of the distribution function return a step function that rises only (but not necessarily) at the values that bound respondents' WTP intervals.

Analysis of interval-censored WTP data requires an estimator that solves the NPML problem reliably and quickly; 'reliably' since we would like to be guaranteed that our estimator returns the unique NPML estimates of the distribution function and 'quickly' since certain estimation problems (e.g. bootstrapping confidence intervals for statistics derived from the non-parametric WTP distribution or the application of distribution-free semi-parametric estimators) may require the NPML problem to be solved very many times. The key message of this paper is that, in general, the standard estimation method employed by CV practitioners to derive NPML estimates from interval-censored data, Turnbull's self-consistent (SC) algorithm, is neither reliable nor fast.

In this journal, Haab and McConnell (1997) show that a closed-form solution to the NPML problem can be derived for the particular case of Type 1 interval-censoring resulting from SB dichotomous choice WTP data. Indeed, this closed-form solution is essentially the well-known pooled adjacent violators algorithm (PAVA) (Ayer et al. 1955). However, as Haab and McConnell emphasise (p. 257), the NPML problem does not have a closed-form solution when WTP data is Type 2 interval-censored (as per data from OOHB or DB dichotomous choice questions). To solve the NPML problem with data of this type requires the application of iterative numerical techniques.

As far as the author is aware, Carson and Steinberg (1990) were the first CV practitioners to calculate non-parametric estimates of the WTP distribution from Type 2 interval-censored data. The estimation procedure they applied was Turnbull's SC algorithm. In fact, there are a whole variety of estimation methods that might be used to derive NPML estimates from interval-censored data. All the same, a comprehensive review of the CV literature reveals

that subsequent authors have chosen, almost without exception, to follow the lead of Carson and Steinberg and employed Turnbull's SC algorithm to analyse their data (e.g. Carson, Wilks and Imber, 1994; Carson et al., 1994; Huhtala, 1994, 2000; Carson and Mitchell, 1995; An and Ayala, 1996; Hutchinson et al., 2001; Scarpa, Willis and Garrod, 2001; Nunes and van den Bergh, 2004). Reliance on Turnbull's SC algorithm appears to have been further entrenched by the fact that Hanemann and Kanninen (1999) direct practitioners to this algorithm in their influential review of CV data analysis. Unchallenged, Turnbull's SC algorithm has firmly established itself as the estimator of choice in the environmental economics literature.

One objective of this paper is to investigate whether Turnbull's SC algorithm is deserving of its position of pre-eminence. As shall become clear, Turnbull's algorithm is not without shortcomings. Indeed, it can be shown that Turnbull's algorithm does not completely solve the NPML problem. Accordingly, estimates of the WTP distribution derived from Turnbull's algorithm are not necessarily NPML estimates and consequently will not always possess all their desirable qualities.

This 'trouble with Turnbull' has long been recognised outside the field of environmental economics (e.g. Gu and Zhang, 1993; Gentleman and Geyer, 1994; Mykland and Ren, 1996; Wellner and Zhan, 1997). This paper builds on that previous research by investigating the conditions under which these problems with Turnbull's algorithm might be encountered in actual estimation problems. Furthermore, a set of convergence criteria are proposed that guarantee the convergence of Turnbull's algorithm to the NPML estimates.

Regrettably, there is a further well-documented problem with Turnbull's SC estimator; it has been shown that the estimator can be extremely slow to converge (Wellner and Zhan, 1997, Jongbloed, 1998, Zhang and Jamshidian, 2004). A second objective of this paper is to introduce a series of alternative estimators drawn from the mathematical programming literature. These include, sequential quadratic programming (Boggs and Tolle, 1995), an interior-point barrier method (Forsgren et al., 2002), a generalised Rosen estimator (Zhang and Jamshidian, 2004), the iterative convex minorant algorithm (Groenboom, 1991) and hybrid algorithms that combine the last two estimators with a self consistent step (Wellner and Zhan, 1997; Zhang and Jamshidian, 2004). As far as the author is aware none of these estimators have been previously applied to the analysis of interval-censored WTP data. Moreover, the paper reports on a series of Monte Carlo simulations that compare the performance of these various estimators when applied to standard CV survey datasets. Our investigation confirms that Turnbull's SC algorithm is frequently out-performed by these alternative estimators.

The paper is organised as follows; section 2 describes the NPML problem that might be used to derive a nonparameteric estimate of the distribution of WTP when CV data is Type 2 interval censored, section 3 describes Turnbull's SC algorithm for solving the NPML problem, section 4 and 5 describe failings of Turnbull's SC algorithm and introduce a series of alternative estimators, section 6 reports on a Monte Carlo experiment that compares these various estimators, section 7 summarises and provides some concluding remarks.

2. CV DATA AND THE NONPARAMETRIC MAXIMUM LIKELIHOOD PROBLEM

Broadly speaking the primary objective of a CV survey is to discover the value placed on changes in the provision of a non-market good. To this end, researchers survey a sample of individuals drawn from the population affected by the proposed changes. Using one of a number of elicitation methods, respondents are asked to reveal the value they would place on the changes described in the survey. In the main, respondents are asked their maximum WTP to avoid changes that are welfare decreasing or maximum WTP to achieve changes that are welfare increasing.

Naturally, individuals in the population do not hold identical WTP values for the changes under consideration. Indeed, the usual assumption is that WTP can be regarded as a random variable, denoted C , whose distribution in the population can be characterised by some continuous cumulative distribution function;

$$F(c) = \Pr(C \leq c) \quad (1)$$

The WTP values of the N respondents in the sample, (c_1, c_2, \dots, c_N) , are simply a selection of independent values drawn at random from this population distribution.

The primary objective of many CV surveys is to use data on the sample's WTP in order to recover an estimate of the population distribution F . Armed with such an estimate, the analyst can recover estimates of policy-relevant statistics such as the population's mean and median WTP.

When the CV survey returns each respondent's precise WTP, c_i ($i = 1, 2, \dots, N$), one particularly appealing estimate of F is provided by the empirical distribution function (EDF);

$$\hat{F}(c) = \frac{1}{N} \sum_{i=1}^N \delta(c)_i \quad \text{where } \delta(c)_i = \begin{cases} 1 & \text{if } c_i \leq c \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

As is well known, as the size of the random sample tends to infinity then the EDF constructed from the sample will tend, with probability one, to the underlying CDF of the population (a result known as the Glivenko-Cantelli Lemma).

Unfortunately, the procedure of simply asking respondents what they are willing to pay, the so-called open-ended (OE) question format, has been criticised on the grounds that it presents respondents with an overly complex cognitive task (Boyle and Bishop, 1988; NOAA, 1993) and may provide incentives for strategic responses (Hoehn and Randall, 1987; Carson, Groves and Machina, 1999). Accordingly, over recent years attention has focused on various question formats that elicit WTP responses by asking respondents one or more dichotomous choice questions. In designing such elicitation formats practitioners specify a number of money amounts, commonly termed bid amounts, that it is believed span the range of WTP to be found in the population. Assuming WTP to be non-negative, we can denote each of the M bid amounts B_m such that,

$$0 = B_0 < B_1 < B_2 < \dots < B_{M-1} < B_M < B_{M+1} = \infty \quad (3)$$

The single-bounded (SB) dichotomous choice elicitation format (Bishop and Heberlein, 1979) randomly assigns a single bid amount to each individual in the sample and respondents are asked to indicate whether they would or would not pay this amount. Accordingly, the SB dichotomous choice format identifies whether a respondent's WTP is in

the interval below that bid amount $[B_0, B_k)$ or above that bid amount $[B_k, B_{M+1}]$. Such data is said to be subject to Type 1 interval-censoring. The double-bounded (DB) and one-and-one-half bounded (OOHB) dichotomous choice elicitation formats (Hanemann, Loomis and Kanninen, 1991; Cooper, Hanemann and Signorello 2002) present respondents with up to two different and not necessarily adjacent bid amounts. Data from such techniques identifies whether a respondent's WTP is in the interval below the lowest bid amount that they have been offered $[B_0, B_l)$, above the highest bid amount that they have been offered $[B_h, B_{M+1}]$ or in the interval between the two bid amounts $[B_l, B_h)$. For convenience we shall indicate the lower bound of this interval for individual i as B_{L_i} and the higher bound B_{H_i} . Data of this type is said to be subject to Type 2 interval-censoring.

When CV surveys collect interval-censored WTP data, the EDF as defined in (2) cannot be calculated and other estimation strategies must be adopted. One possible approach is to assume that F can be described by some parametric distribution function. Of course, such assumptions are essentially *ad hoc* such that the asymptotic convergence properties that characterise the EDF will not hold for parametric specifications in which an incorrect distributional assumption is made. Moreover, estimates of mean and median WTP based on such assumptions may be highly sensitive to the particular choice of parametric distribution (Ready and Hu, 1995; Haab and McConnell, 1997; Kerr, 2000)¹. Accordingly, attention has focussed on distribution-free methods for estimating F . Since distribution-free estimates of F make no *ad hoc* assumptions concerning the population distribution of WTP, estimates of mean and median WTP derived from such approaches are to be preferred. Indeed, the central tendency measure from distribution-free estimates of F can be shown to be a lower-bound approximation to the population's expected WTP (Haab and McConnell, 1997; Boman et al., 1999).

For interval-censored data, the problem of deriving a distribution-free estimate of F can be reduced to one of maximising a likelihood function subject to certain constraints. Accordingly, the estimates of F derived from solving this problem are termed *nonparametric maximum likelihood* estimates.

Assume, for now, that we knew the form of the population distribution function. Then the probability of observing an individual from this population indicating that their WTP lay in the interval between B_{L_i} and B_{H_i} is given by;

$$\Pr[B_{L_i} \leq c_i < B_{H_i}] = F(B_{H_i}) - F(B_{L_i}) \quad i = 1, 2, \dots, N \quad (4)$$

Furthermore, the probability of observing a particular set of responses in a random sample of N individuals from the population is given by the likelihood function;

$$L = \prod_{i=1}^N F(B_{H_i}) - F(B_{L_i}) \quad (5)$$

To provide a more generic presentation of the likelihood function we require some notation. First we need to be able to refer to the intervals between the amounts in the bid design (3).

¹ In contrast, Giraud et al. (2001) fit four parametric models to their data; logit, probit, Weibull and log-logit. They are unable to calculate a mean WTP value from the log-logit model, but find that estimates of mean WTP from the other three distributions do not differ significantly provided they first truncate the estimated distributions at zero and the highest bid level.

As such, we label these various intervals $m = 1$ to M such that interval m is the interval defined by $[B_{m-1}, B_m)$. Following An and Ayala (1996) we call these *basic intervals* to distinguish them from the WTP intervals defined by an individual's upper and lower bounds and the equivalence classes that we shall define subsequently. Further we define an indicator variable d_{im} that records whether an individual's WTP interval includes the m^{th} basic interval. That is;

$$d_{im} = \begin{cases} 1 & \text{if } B_{L_i} \leq B_{m-1}, B_m \leq B_{H_i} \\ 0 & \text{otherwise} \end{cases} \quad m = 1, 2, \dots, M+1; i = 1, 2, \dots, N \quad (6)$$

With this notation the likelihood function (5) can be expressed;

$$L = \prod_{i=1}^N \left(\sum_{m=1}^{M+1} d_{im} [F(B_m) - F(B_{m-1})] \right) \quad (7)$$

Simplifying notation such that $F(B_m) = F_m$ ($m = 0, 1, \dots, M+1$) with $F_0 = 0$ and $F_{M+1} = 1$, and taking logs gives the *log likelihood function*;

$$\ln L = \sum_{i=1}^N \ln \left(\sum_{m=1}^{M+1} d_{im} [F_m - F_{m-1}] \right) \quad (8)$$

Of course, we do not know the probability distribution of *WTP*. Indeed, the challenge facing the analyst is to estimate this distribution function based on data provided by the CV survey.

It is clear from Equation (8) that for a particular set of data, the value of the log likelihood is determined solely by the values taken by the M probabilities, F_m ($m = 1, 2, \dots, M$). Consequently, the maximally flexible specification of the log likelihood is one in which each of the F_m ($m = 1, 2, \dots, M$) are treated as parameters. That is to say, the problem can be reduced to the finite dimensional problem of estimating the value of M probabilities.

As a matter of fact, the problem can be further reduced since the sample data may not provide information that allows us to independently estimate each of the F_m ($m = 1, 2, \dots, M$). For example, imagine that there are no respondents in our sample with WTP intervals spanning a particular basic interval of the bid design, let us say interval k . In that case $d_{ik} = 0$ for all respondents, and the data provides no evidence of probability mass falling into the k^{th} basic interval. As a result, the maximum likelihood estimate for F_k will be identical to that for F_{k-1} such that F_k need not be estimated. Likewise, if each respondent in the sample states a WTP interval that either spans or does not span both of two (or more) consecutive basic intervals then the distribution of probability mass within those basic intervals is irrelevant to the maximisation problem. Accordingly, the basic intervals can be combined into one interval and intervening values of F_m need not be estimated. Indeed as described by Goetghebeur and Ryan (2000) the only intervals that are relevant to the maximisation problem are those defined by the lower bound of one observation's WTP interval and the upper bound of another (possibly the same) observation's WTP interval, such that no other observation's lower or upper bound falls within that interval. Such intervals are termed *equivalence classes*. Equivalence classes are the basic building blocks of the maximum likelihood estimator. They determine the intervals within which the maximum likelihood estimate of the probability distribution can, but not necessarily does, attribute probability mass.

The $M + 1$ basic intervals defined by the bid design can be maintained, aggregated or discarded to form J equivalence classes. Notice that the construction of equivalence classes ensures that at least one observation falls entirely within the first equivalence class and that

at least one observation falls entirely within the J^{th} equivalence class. Denote the bid level at the start of the first equivalence class B_0 and that at the end of each of the J equivalence classes by B_j ($j = 1, 2, \dots, J$). Then the indicator variable can be reformulated accordingly;

$$d_{ij} = \begin{cases} 1 & \text{if } B_{L_i} \leq B_{j-1}, B_j \leq B_{H_i} \\ 0 & \text{otherwise} \end{cases} \quad j = 1, 2, \dots, J; i = 1, 2, \dots, N \quad (9)$$

Further let \mathbf{d}_i represent the J vector of ones and zeros indicating which equivalence classes are spanned by individual i 's WTP interval. Further, stacking the \mathbf{d}_i' ($i = 1, 2, \dots, N$), let \mathbf{d} be the N by J vector representing the data in the sample. If we denote the cumulative probability density function at the beginning of the first equivalence class by F_0 , that at the end of the J^{th} equivalence class F_J , and gather those at the end of the intervening $J - 1$ equivalence classes into the vector \mathbf{F} with elements F_j ($j = 1, 2, \dots, J - 1$), then the log likelihood of the maximum likelihood problem can be written;

$$\begin{aligned} \text{Max}_{\mathbf{F}} \ln L(\mathbf{F} | \mathbf{d}) &= \sum_{i=1}^N \ln \left(\sum_{j=1}^J d_{ij} [F_j - F_{j-1}] \right) \\ \text{Subject to :} & \\ 0 = F_0 &\leq F_1 \leq F_2 \leq \dots \leq F_J = 1 \end{aligned} \quad (10)$$

Here the constraints ensure that the NPML estimates define a valid cumulative probability distribution function; that is, they ensure that it is a positive, monotonically increasing function that has a maximal value of unity. Solving this problem provides the *nonparametric maximum likelihood* (NPML) estimates of the distribution function. We shall denote the solution by the $J - 1$ vector $\hat{\mathbf{F}}^{ML}$ with elements \hat{F}_j^{ML} ($j = 1, 2, \dots, J - 1$).

The NPML estimates will be unique provided the log likelihood $\ln L(\mathbf{F} | \mathbf{d})$ is strictly concave a property that follows if the Hessian of the log likelihood is negative definite. To form the Hessian we note that the $J - 1$ elements of the gradient vector, \mathbf{g} , are given by;

$$g_k = \frac{\partial \ln L}{\partial F_k} = \sum_{i=1}^N \frac{(d_{ik} - d_{ik+1})}{\sum_{j=1}^J d_{ij} (F_j - F_{j-1})} = \sum_{i=1}^N \frac{(d_{ik} - d_{ik+1})}{\eta_i} \quad (k = 1, 2, \dots, J - 1) \quad (11)$$

where $\eta_i = \sum_{j=1}^J d_{ij} (F_j - F_{j-1})$, and the elements of the $J - 1$ by $J - 1$ Hessian matrix, \mathbf{H} , are given by;

$$H_{kh} = \frac{\partial^2 \ln L}{\partial p_k \partial p_h} = - \sum_{i=1}^N \frac{(d_{ik} - d_{ik+1})(d_{ih} - d_{ih+1})}{\eta_i^2} \quad (k = 1, 2, \dots, J - 1; h = 1, 2, \dots, J - 1) \quad (12)$$

Following the development in Gentleman and Geyer (1994), the Hessian, can be decomposed into the two matrices, \mathbf{d}^- (the N by $J - 1$ matrix formed by first differencing \mathbf{d}) and \mathbf{A} , the N by N diagonal matrix with elements $-1 / \eta_i^2$, according to;

$$\mathbf{H} = \mathbf{d}' \mathbf{A} \mathbf{d}^- \quad (13)$$

which is a simple quadratic form. Since \mathbf{A} is negative definite, \mathbf{H} will be of full rank and the NPML estimates will be unique provided;

$$\text{rank}(\mathbf{d}^-) = J - 1 \quad (14)$$

which is guaranteed by the manner in which the J equivalence classes have been constructed.

Since the objective function is concave and the constraints are linear, (10) is an example of a concave program. Accordingly, we formulate the Langrangian function;

$$G = \sum_{i=1}^N \ln \left(\sum_{j=1}^J d_{ij} [F_j - F_{j-1}] \right) - \lambda_0 F_0 + \sum_{j=1}^J \mu_j (F_j - F_{j-1}) - \lambda_1 (F_J - 1) \quad (15)$$

where λ_0 , λ_1 and μ_j ($j = 1, 2, \dots, J$) are Lagrange multipliers, and derive the following Khun-Tucker conditions describing the first-order necessary conditions for a maximum;

1. *Primal Constraints:*

$$F_0 = 0, F_J = 1, F_j - F_{j-1} \geq 0 \quad (j = 1, 2, \dots, J) \quad (15a)$$

2. *Dual Constraints:*

$$\mu_j \geq 0 \quad (j = 1, 2, \dots, J) \quad (15b)$$

3. *Complementary Slackness:*

$$\mu_j (F_j - F_{j-1}) = 0 \quad (j = 1, 2, \dots, J) \quad (15c)$$

4. *Gradients of Lagrangian:*

$$\alpha_1 + \mu_1 = -\lambda_0 \quad (15d)$$

$$\alpha_j + \mu_j = \lambda_1 \quad (15e)$$

$$\alpha_j + \mu_j = \alpha_{j+1} + \mu_{j+1} \quad (j = 1, 2, \dots, J - 1) \quad (15f)$$

where;

$$\alpha_j = \frac{\sum_{i=1}^N d_{ij}}{\sum_{k=1}^J d_{ik} (F_k - F_{k-1})} \quad (j = 1, 2, \dots, J) \quad (16)$$

The unique NPML estimates $\hat{\mathbf{F}}^{ML}$ will simultaneously comply with conditions (15a to f).

3. NPML ESTIMATION USING TURNBULL'S SELF-CONSISTENT ALGORITHM

In the case of Type 1 interval censored data, the NPML problem in (15) can be shown to have a simple closed form solution. Indeed, Haab and McConnell (1997) demonstrate this solution and apply it to SB dichotomous choice WTP data. In contrast, no closed form solution exists for Type 2 interval-censored data. Accordingly, NPML estimates of the WTP distribution from OOHb or DB dichotomous choice data, requires the application of iterative numerical techniques. One such approach, and that which has seen almost blanket application in the CV literature, is that suggested by Turnbull (1974, 1976). For reasons that shall become clear shortly, Turnbull's approach is called a self-consistent (SC) algorithm (Efron, 1967).

To motivate this method of estimation consider the Kuhn-Tucker conditions (15a-f). The primal constraints, (15a), simply restate the monotonicity requirements. They ensure that the cumulative probability distribution function is monotonically increasing and that the probability masses within each equivalence class sum to one. That is to say;

$$\sum_{j=1}^J (F_j - F_{j-1}) = 1 \quad (17)$$

Moreover, the nullity of the gradient of the Lagrangean, (15d), (15e) and (15f), informs us that at a maximum;

$$\alpha_j + \mu_j = -\lambda_0 = \lambda_1 = \lambda \quad (j = 1, 2, \dots, J) \quad (18)$$

Multiplying (18) through by $(F_j - F_{j-1})$ gives;

$$\alpha_j (F_j - F_{j-1}) + \mu_j (F_j - F_{j-1}) = \lambda (F_j - F_{j-1}) \quad (j = 1, 2, \dots, J) \quad (19)$$

Enforcing the complimentary slackness condition (15c) requires $\mu_j (F_j - F_{j-1}) = 0$, such that summing the J equations of (19) and simplifying using (17) provides;

$$\sum_{j=1}^J \alpha_j (F_j - F_{j-1}) = \lambda \quad (20)$$

Notice that our manipulations have employed the Kuhn-Tucker conditions relating to the primal constraints, complementary slackness and vanishing gradients, but not those pertaining to the dual constraints (15b).

Before, completing the derivation of Turnbull's algorithm let us consider for a moment the term $\alpha_j (F_j - F_{j-1})$ that appears in (20). Expanding using (16) reveals;

$$\alpha_j (F_j - F_{j-1}) = \frac{\sum_{i=1}^N d_{ij} (F_j - F_{j-1})}{\sum_{k=1}^J d_{ik} (F_k - F_{k-1})} = \sum_{i=1}^N \bar{\delta}_{ij}(\mathbf{F}) \quad (j = 1, 2, \dots, J) \quad (21)$$

The ratio in the middle expression of (21) defines a conditional probability; the likelihood that individual i 's WTP lies in the j^{th} equivalence class, given estimates of the population distribution of WTP. These conditional probabilities, labelled $\bar{\delta}_{ij}(\mathbf{F})$ ($j = 1, 2, \dots, J$; $i = 1, 2, \dots, N$), have an intuitive interpretation. Imagine we wished to count the number of individuals

in the sample whose WTP was contained in each of the equivalence classes. Since an individual's WTP interval might span several equivalence classes, the contribution of that observation to the counts must somehow be divided between that set of equivalence classes. The $\bar{\delta}_{ij}(\mathbf{F})$ ($j = 1, 2, \dots, J; i = 1, 2, \dots, N$) represent just such a division. Indeed, in the absence of specific information, the fraction of each observation allotted to each particular equivalence class is determined by our estimates of the population distribution of WTP across that set of equivalence classes. The dependence of each fraction on our estimates of the population distribution of WTP is made explicit by expressing $\bar{\delta}_{ij}(\mathbf{F})$ ($j = 1, 2, \dots, J; i = 1, 2, \dots, N$) as a function of \mathbf{F} .

Summing these probabilities over all individuals in the sample, as in the right hand expression of (21), gives the expected number of individuals with WTP in equivalence class j . Since, all individuals must have a WTP that lies in at least one of the equivalence classes, summing these expectations over all j will simply give the number of individuals in the sample, N . Accordingly, we derive an explicit solution for (20);

$$\lambda = \sum_{j=1}^J \alpha_j (F_j - F_{j-1}) = \sum_{j=1}^J \sum_{i=1}^N \bar{\delta}_{ij}(\mathbf{F}) = N \quad (22)$$

Substituting this result into the rhs of equations (18), multiplying through by $(F_j - F_{j-1})$ and rearranging reveals;

$$F_j = \frac{\sum_{i=1}^N \bar{\delta}_{ij}(\mathbf{F})}{N} + F_{j-1} \quad (j = 1, 2, \dots, J) \quad (23)$$

Since (15a) dictates that $F_0 = 0$ we can solve for F_j ($j = 1, 2, \dots, J-1$) recursively deriving the general expression;

$$F_j = \frac{1}{N} \sum_{k=1}^j \sum_{i=1}^N \bar{\delta}_{ik}(\mathbf{F}) \quad (j = 1, 2, \dots, J-1) \quad (24)$$

These are *Turnbull's self-consistency equations* and any set of F_j satisfying (24) are termed self-consistent (SC) estimates of the NPML problem. We denote such a set of estimates by the $J-1$ vector $\hat{\mathbf{F}}^{SC}$ with elements \hat{F}_j^{SC} ($j = 1, 2, \dots, J-1$). In particular, notice from (24) that the F_j appear on both the left-hand and right-hand sides of the $J-1$ self consistency equations. For this set of equations to hold, therefore, it must be the case that, given the data, the vector of probabilities, \mathbf{F} , predict themselves. It is in this sense that the vector $\hat{\mathbf{F}}^{SC}$ is said to be a *self-consistent* estimate. Moreover, observe the similarity between the self-consistency equations (24) and the empirical distribution function (2). Rather than calculating the cumulative proportion of individuals with WTP less than or equal to B_j , the self-consistency equations calculate the cumulative proportion of individuals with *expected* WTP of that amount or less.

As expounded by Turnbull (1976) and, in the context of WTP data, by An and Ayala (1996), the format of (24) suggests a simple estimation method;

1. Select initial parameter estimates $\mathbf{F}^{(0)}$, conforming with the primal constraints, such that;

$$0 = F_0^{(0)} \leq F_1^{(0)} \leq F_2^{(0)} \leq \dots \leq F_J^{(0)} = 1$$

2. Use $\mathbf{F}^{(0)}$ to calculate the expected observation fractions in each equivalence class according to;

$$\bar{\delta}_{ij}(\mathbf{F}^{(0)}) = \frac{d_{ij}(F_j^{(0)} - F_{j-1}^{(0)})}{\sum_{k=1}^J d_{ik}(F_k^{(0)} - F_{k-1}^{(0)})} \quad (j = 1, 2, \dots, J; i = 1, 2, \dots, N). \quad (\text{E})$$

3. Obtain new parameter estimates using the standard EDF calculation according to;

$$F_j^{(1)} = \frac{1}{N} \sum_{k=1}^j \sum_{i=1}^N \bar{\delta}_{ik}(\mathbf{F}^{(0)}) \quad (j = 1, 2, \dots, J-1) \quad (\text{M})$$

4. Return to step 2 replacing $\mathbf{F}^{(0)}$ with $\mathbf{F}^{(1)}$
5. Stop when the absolute value of the change in parameters from one iteration to the next, falls below a given tolerance level, ε , for all the parameters. That is, the parameters are said to have converged when;

$$\max_{j=1,2,\dots,J-1} \left(\left| F_j^{(t+1)} - F_j^{(t)} \right| \right) < \varepsilon \quad (25)$$

where t is a counter denoting iterations of the algorithm.

6. The resulting vector of probabilities are self-consistent estimates, $\hat{\mathbf{F}}^{SC}$.

Notice, that initialising the algorithm from starting values, $\mathbf{F}^{(0)}$, that conform with the monotonicity constraints (15a), ensures that in Step 2 each $\bar{\delta}_{ij}(\mathbf{F}^{(0)})$ ($j = 1, 2, \dots, J; i = 1, 2, \dots, N$) is nonnegative. In turn, in step 3 $F_j^{(1)}$ ($j = 1, 2, \dots, J-1$), the updated parameter estimates, also conform with the monotonicity constraints. Indeed, we are guaranteed that the primal constraints will not be violated at each subsequent iteration of the algorithm.

As Turnbull (1976) points out, the SC algorithm can be viewed as an EM algorithm (Dempster et al., 1977). In this case, the observed data are incomplete insomuch as each observation indicates the range of equivalence classes but not the particular equivalence class within which a respondent's WTP is contained. The E-step of the algorithm (step 2 above) calculates the expected value of the complete data conditional on a set of initial parameter estimates and the M-step of the algorithm (step 3 above) maximises the log likelihood of this expected data so as to provide a new set of parameters estimates. The properties of the EM algorithm are well-established. In particular, Dempster et al. (1977) show that at each iteration of the EM algorithm, the objective function is monotonically increasing. That is to say, we can be certain that $\ln L(\mathbf{F}^{(t+1)} | \mathbf{d}) \geq \ln L(\mathbf{F}^{(t)} | \mathbf{d})$ for each step of Turnbull's algorithm. Moreover, Nettleton (1999) building on the work of Wu (1983), provide theorems that prove that the sequence of iterations of Turnbull's algorithm will converge to a fixed point; that is to a set of self-consistent estimates.

4. TROUBLES WITH TURNBULL: CONVERGENCE TO NPML ESTIMATES

A fundamental question remains to be addressed; are a set of self-consistent estimates, $\hat{\mathbf{F}}^{SC}$, resulting from the convergence of Turnbull's algorithm necessarily the unique set of NPML estimates, $\hat{\mathbf{F}}^{ML}$, that we seek?

Recall that the self-consistency equations (24) are derived directly from the Kuhn-Tucker conditions characterising the NPML problem (15a-f). As a matter of fact, in deriving the self-consistency equations we made use of only five of those six conditions (15a,c-f). Since, by definition, the unique NPML estimates, $\hat{\mathbf{F}}^{ML}$, must also conform to these five conditions, they must be self-consistent estimates. However, in general, the reverse is not true. Indeed, as illustrated by Gu and Zhang (1993), there may be several parameter vectors $\hat{\mathbf{F}}^{SC}$ fulfilling the self-consistency equations, only one of which is the NPML solution vector $\hat{\mathbf{F}}^{ML}$.

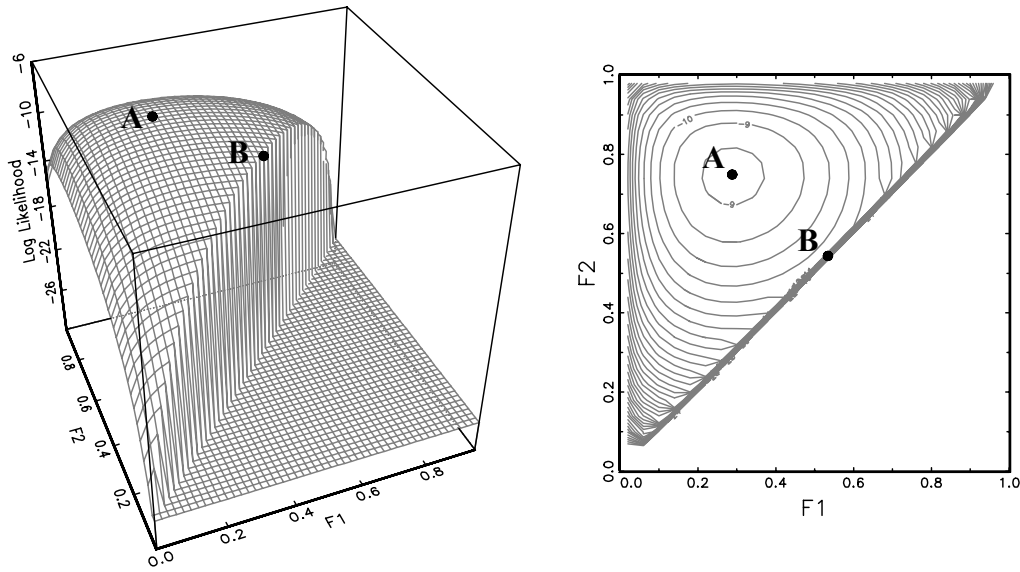
In particular, observe that the derivation of the self-consistency equations did not involve enforcing the dual constraints; that is, non-negativity of the Lagrange multipliers $\mu_j \geq 0$ ($j = 1, 2, \dots, J$) (15b). Accordingly, there may exist locations in the parameter space at which the parameter estimates fulfil the self-consistency equations but are characterised by one or more negative Lagrange multipliers. Such self-consistent estimates will not be the maximum likelihood estimates that we seek.

Of course, the complimentary slackness conditions (15c) intimate that non-zero values for Lagrange multipliers must be associated with locations where one of the monotonicity constraints are binding (i.e. where $\hat{F}_j^{SC} = \hat{F}_{j-1}^{SC}$ for some j). These alternative self-consistent locations must, therefore, be found on one of the boundaries of the feasible parameter space. As pointed out by Gentleman and Geyer (1994), such locations must represent maxima with respect to the log likelihood along that boundary, but are not stationary points of the log-likelihood in general. Moreover, since the Lagrange multiplier at this location is negative, moving away from such a location into the interior of the feasible parameter space must increase the log-likelihood.

Observe Figure 1 which displays the log-likelihood surface for a hypothetical dataset with fifteen observations defining three equivalence classes ($J = 3$). We wish to find the NPML estimates of the probability density function for this data. By construction of equivalence classes, it must be the case that the probability density function at the lower bound of the first equivalence class, F_0 , is equal to zero and that at the upper bound of the third equivalence class F_3 , is equal to one. Accordingly, our problem reduces to a two-dimensional problem of finding values of F_1 and F_2 (the probability density function at the upper bound of the first and second equivalence classes respectively) that maximise the log-likelihood for this data.

The left-hand panel of Figure 1 shows a three-dimensional relief of the log-likelihood surface for this data. A contour line representation of the same surface is shown in the right-hand panel. Notice that the monotonicity constraints on the probability density function ($0 = F_0 \leq F_1 \leq F_2 \leq F_3 = 1$) restrict the feasible space to an area above the off-diagonal in the unit square.

Figure 1: NPML and SC locations on the log-likelihood surface; three-dimensional and contour representations of a hypothetical two parameter problem



As foreseen, the log-likelihood is strictly concave with a maximum at the point labelled A where $F_1 = 0.2857$ and $F_2 = 0.75$. Indeed, this is a self-consistent location to which Turnbull's algorithm will converge when initiated from most starting parameters within the feasible space. However, observe location B on the likelihood surface where $F_1 = F_2 = 0.5333$. This location represents a maximum with respect to the boundary of the feasible region imposed by the constraint $F_1 \leq F_2$. Location B is also a self-consistent location and, as we shall see, Turnbull's algorithm will converge to this location when initiated from particular feasible starting parameters.

Table 1 documents some pertinent statistics regarding characteristics of the two self-consistent locations, A and B. The final column lists the values of the Lagrange multipliers which can be calculated easily by deriving the following expression from the manipulation of (18) and (22);

$$\mu_j = N - \alpha_j \quad (j = 1, 2, \dots, J) \quad (26)$$

Notice that location A is the unconstrained maximum of the log-likelihood function. Accordingly, the gradient vector is zero and, since none of the constraints are binding, each Lagrange multiplier is also zero. In contrast, at location B the gradients are not zero and the Lagrange multiplier associated with the second constraint ($F_1 \leq F_2$) is negative. Whilst it can easily be verified from Table 1 that location B satisfies each of the Kuhn-Tucker conditions from which the self-consistency equations are derived, it violates the dual constraints and, therefore, cannot be a constrained maximum of the log-likelihood function.

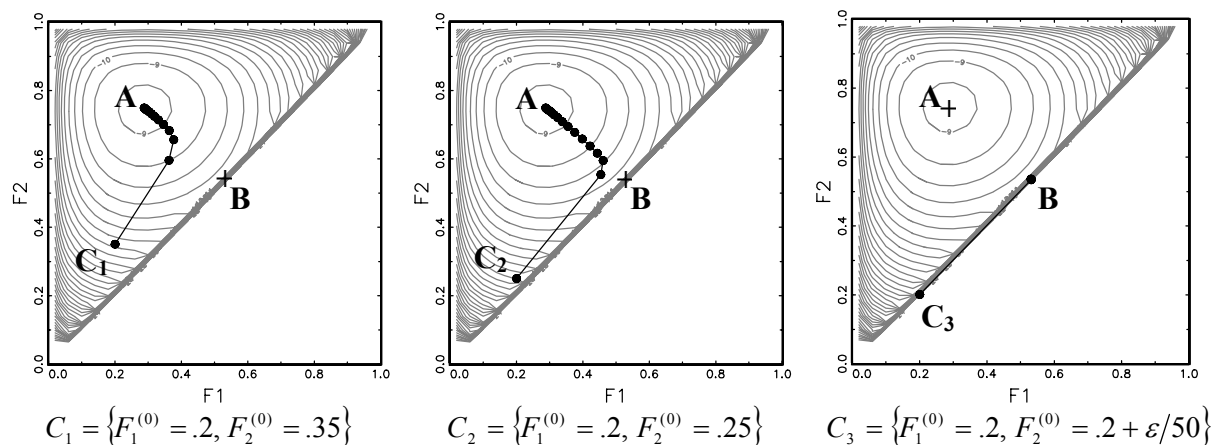
Table 1: Characteristics of self-consistent locations A and B

Equivalence Class (j)	Probability (F_j)	Gradient (g_j)	Alpha (α_j)	Lagrange Multiplier (μ_j)
<i>Location A:</i>				
1	0.2857	0	15	0
2	0.7500	0	15	0
3	1		15	0
<i>Location B:</i>				
1	0.5333	-6.9643	15	0
2	0.5333	6.9643	21.9643	-6.9643
3	1		15	0

The existence of multiple SC locations in the feasible parameter space has long been recognised by authors outside the field of environmental economics. (e.g. Gu and Zhang, 1993; Gentleman and Geyer, 1994; Mykland and Ren, 1996; Wellner and Zhan, 1997). Indeed, Mykland and Ren (1996) report the results of a simulation exercise that indicate that, in a moderate sample size, the NPML location is a better estimator of the true population probability distribution function than an arbitrarily selected SC location (that is not the NPML location).² Moreover, this ‘trouble’ with Turnbull’s algorithm has been advanced by certain authors as a reason to prefer alternative estimators that unfailingly converge to the NPML estimates (Wellner and Zhan, 1997).

As a matter of fact, the problem is not as troubling as it might at first appear. Consider Figure 2. Using the same data as that illustrated in Figure 1, the three panels show the steps taken by Turnbull’s SC algorithm from three different initial parameter vectors, C_1 , C_2 and C_3 .

Figure 2: Location of iterations of Turnbull’s SC algorithm from various starting parameters to convergence ($\varepsilon = 10^{-5}$)



² As a matter of fact Mykland and Ren (1996) consider data subject to a slightly different censoring scheme than that considered here, though their results will apply equally to interval-censored data.

Notice that when the starting parameters are well away from the boundary of the feasible parameter space (C_1) the algorithm finds its way reasonably directly to the NPML estimates (location A). When the algorithm is initiated from a location comparatively close to the boundary (C_2), the algorithm may require relatively more iterations but still converges to the NPML estimates. Indeed, in order to force Turnbull's algorithm to converge on a SC location along a particular boundary (location B), a set of starting parameters must be selected that are arbitrarily close to that boundary of the feasible space (C_3).

The reason for this observed pattern is clear. As we have shown, the log-likelihood function is concave and the feasible space is delineated by a set of monotonicity constraints that define a convex set (indeed, a finitely-generated convex cone). As a consequence, at every location in the feasible parameter space, barring the NPML location, there exists a direction in which the log-likelihood is increasing. Moreover, a property of Turnbull's algorithm (and more generally EM algorithms) is that the log-likelihood function is non-decreasing with each iteration. Accordingly, even at locations arbitrarily close to a SC location such as B, the algorithm will continue to clamber uphill to the NPML estimates at the summit.³

Be that as it may, around SC locations, Turnbull's algorithm takes arbitrarily small steps. Accordingly, if the algorithm arrives close to such a location then the step length may fall below the tolerance limit defining convergence, ε , and the iterations may cease. It turns out, that the algorithm will only arrive at an SC location that is not the NPML location under two circumstances. First, if the iterations are initiated from very close to that SC location itself. Second, if the iterations are initiated from any other location that is very close to the boundary upon which the SC location is situated.⁴

This characterisation suggests two possible solutions to this problem with Turnbull's SC algorithm. First, one should avoid choosing starting parameters close to the boundary of the feasible parameter space. Indeed, Turnbull (1976) himself indicates that starting parameters should be strictly in the interior of the parameter space, that is;

$$0 = F_0^{(0)} < F_1^{(0)} < F_2^{(0)} < \dots < F_j^{(0)} = 1 \quad (27)$$

Second, one might adopt a convergence criterion that does not depend on step length. In particular, we would like to confirm that a candidate location conforms to the Kuhn-Tucker conditions (15a-f). Wellner and Zhan (1997) propose a set of stopping criteria that check compliance of the candidate solution with the complementary slackness and gradient conditions. Of course, such criteria do not protect against SC locations that are not the NPML estimates. Accordingly, we propose broadening Wellner and Zhan's stopping criteria to ensure that a candidate solution also complies with the dual constraint conditions

³ More formally, this property follows from Corollary 1 of Theorem 2 proved by Nettleton (1999) for the specific case in which the objective function is globally concave.

⁴ In the latter case it matters how many observations lie wholly within the equivalence class defined by the constraint boundary (i.e. equivalence class j for the boundary defined by the constraint $F_{j-1} \leq F_j$). When there are very few observations wholly within equivalence class j and the starting parameters are such that the difference between $F_j^{(0)}$ and $F_{j-1}^{(0)}$ is close to zero, then the expected number of observations in equivalence class j , calculated in the E-step of Turnbull's algorithm, will also be close to zero. In turn, the absolute difference in the values of $F_j^{(1)}$ and $F_{j-1}^{(1)}$, estimated in the M-step of Turnbull's algorithm, will itself be very small. In such situations, the algorithm may take numerous iterations to struggle away from that boundary, in which time it may skirt close to the SC location at the maximal value of the log likelihood along that boundary.

stipulating non-negativity of the Lagrange multipliers.⁵ Our proposed stopping criteria are as follows;

$$\left| \sum_{j=1}^{J-1} \hat{F}_j \hat{g}_j \right| < \varepsilon \quad (28a) \quad \left| \sum_{j=1}^{J-1} \hat{g}_j \right| < \varepsilon \quad (28b)$$

$$\left| \hat{\mu}_j \right| < \varepsilon \quad \text{if } \hat{\mu}_j < 0 \quad (j = 1, 2, \dots, J) \quad (28c)$$

where \hat{F}_j ($j = 1, 2, \dots, J-1$) is a candidate solution, \hat{g}_j ($j = 1, 2, \dots, J-1$) and $\hat{\mu}_j$ ($j = 1, 2, \dots, J$) are respectively the gradients (11) and Lagrange multipliers (26) calculated at that location and ε is some small value. The first criteria (28a) is derived by subtracting the J complementary slackness equations (15c) from one another, observing from (11), (16) and (15f) that $g_j = \alpha_j - \alpha_{j+1} = \mu_{j+1} - \mu_j$ ($j = 1, 2, \dots, J-1$) and noting that at the NPML location $\mu_1 = 0$ and $\mu_J = 0$.⁶ The second criteria (28b) is derived by observing that the gradient conditions (15f) indicate that at a maximum, $\mu_{j+1} = g_{j+1} - \mu_{j+2}$ ($j = 1, 2, \dots, J-1$) and that this relationship can be recursively replaced in the first of the gradient conditions and the resulting expression simplified by setting $\mu_1 = 0$, $\mu_J = 0$ and $\alpha_1 - \alpha_2 = g_1$. The third criteria (28c) simply checks that the dual constraints are not contravened.

As pointed out by Gentleman and Geyer (1994) all of the quantities in (28) are calculated during the self-consistent step. Moreover, at each iteration, the three criteria may be evaluated sequentially, if one of the criteria is not met then the candidate solution can be rejected without the other criteria being tested. Accordingly, the stopping criteria in (28) add little computational effort to each self-consistent step and, provided the algorithm is initiated from a strictly interior location in the parameter space, guarantee convergence of Turnbull's SC to NPML estimates.

⁵ In essence, a comprehensive set of convergence criteria such as these is anticipated by Gentleman and Geyer (1994) and underpins Mykland and Ren's (1996) proposal that the algorithm should be restarted from alternative initial values if non-negativity of the Lagrange multipliers cannot be confirmed for a candidate solution.

⁶ The latter two equalities following from the observation that the construction of equivalence classes dictates that at least one observation falls entirely within the first equivalence class and that at least one observation falls entirely within the J^{th} equivalence class. Consequently, the NPML estimates must attribute positive probability mass to the first and J^{th} equivalence classes such that complementary slackness requires $\mu_1 = \mu_J = 0$.

5. TROUBLES WITH TURNBULL: SPEED OF CONVERGENCE AND ALTERNATIVE ESTIMATORS

A number of estimation problems require that NPML estimates are calculated very many times. For example, one might wish to employ bootstrapping methods in order to determine nonparametric confidence intervals for the NPML estimates themselves or for the lower-bound mean WTP estimate calculated from them. Similarly, certain semiparametric estimators (e.g. that proposed by Huang and Wellner, 1996), that investigate the influence of covariates on the nonparameteric distribution of WTP, require repeated calculation of NPML estimates. The time demands of such procedures are often prohibitive such that great stock is placed on the use of estimators that converges rapidly. Regrettably, there is extensive evidence indicating that Turnbull's algorithm can be extremely slow to converge (e.g. Wellner and Zhan, 1997, Jongbloed, 1998, Zhang and Jamshidian, 2004).

Whilst CV practitioners have relied almost exclusively on Turnbull's SC algorithm in order to solve the NPML problem for Type 2 interval censored data, an abundance of alternative estimators have been developed in other fields of research to solve the same problem. A further objective of this paper is to introduce to practitioners of contingent valuation a series of alternative NPML estimators drawn from the extensive mathematical programming literature. As we shall detail in the following section, certain of these estimators converge much more rapidly than Turnbull's algorithm when applied to typical contingent valuation data sets.

5.1 Sequential Quadratic Programming

Sequential Quadratic Programming (SQP) is perhaps the most widely applied method for solving constrained optimisation problems like the NPML problem (see Boggs and Tolle, 1995 for a discussion of the history and development of SQP). SQP consists of a family of iterative algorithms in which a simpler sub-problem involving a quadratic approximation to the objective function is solved at each iteration. Indeed, SQP methods can be regarded as the extension of Newton and Newton-like methods to constrained optimisation.

To illustrate the application of a SQP method to the NPML problem, consider taking a Taylor series expansion of the log-likelihood function around the parameter estimates at iteration t , $\mathbf{F}^{(t)}$, in order to obtain the local quadratic approximation;

$$\ln L(\mathbf{F}^{(t+1)}) \approx \ln L(\mathbf{F}^{(t)}) + (\mathbf{F}^{(t+1)} - \mathbf{F}^{(t)})' \mathbf{g}^{(t)} + \frac{1}{2} (\mathbf{F}^{(t+1)} - \mathbf{F}^{(t)})' \mathbf{H}^{(t)} (\mathbf{F}^{(t+1)} - \mathbf{F}^{(t)}) \quad (29)$$

where $\mathbf{g}^{(t)}$ and $\mathbf{H}^{(t)}$ are respectively the vector of gradients (11) and the Hessian matrix (12) evaluated at $\mathbf{F}^{(t)}$. The sub-problem at this iteration is to find $\mathbf{F}^{(t+1)}$ that maximises (29) subject to the monotonicity constraints; $0 = F_0^{(t+1)} \leq F_1^{(t+1)} \leq \dots \leq F_J^{(t+1)} = 1$. In this context it is usual to reformulate those constraints as follows;

$$\mathbf{A}\mathbf{F}^{(t+1)} \geq \mathbf{b} \quad (30a) \quad \mathbf{F}^{(t+1)} \geq \mathbf{0} \quad (30b)$$

where \mathbf{A} is a $J - 1$ square matrix with each element of the main diagonal being -1, each element of the first upper diagonal being 1, and all other elements 0, whilst \mathbf{b} is a $J - 1$ vector of zeros with the value of the $J - 1$ th element set to -1. The Khun-Tucker conditions characterising a maximum of this problem take the form;

$$\mathbf{v} = \mathbf{H}^{(t)} \mathbf{F}^{(t+1)} - \mathbf{A}\boldsymbol{\gamma} + \mathbf{g}^{(t)} - \mathbf{H}^{(t)} \mathbf{F}^{(t)} \quad (31a)$$

$$\mathbf{s} = \mathbf{A}\mathbf{F}^{(t+1)} - \mathbf{b} \quad (31b)$$

$$F_j^{(t+1)}, \gamma_j, v_j, s_j \geq 0 \quad (j = 1, 2, \dots, J-1) \quad (31c)$$

$$F_j^{(t+1)}v_j = 0, s_j\gamma_j = 0 \quad (j = 1, 2, \dots, J-1) \quad (31d)$$

where $\boldsymbol{\gamma}$ and \boldsymbol{v} are both $J-1$ vectors representing the Lagrange multipliers on the inequality constraints (30a) and non-negativity constraints (30b) respectively. Solving these first order necessary conditions for $\boldsymbol{F}^{(t+1)}$ is made easy by observing that they are in the form of a linear complementarity problem (LCP) which seeks to find vectors \boldsymbol{w} and \boldsymbol{z} satisfying;

$$\boldsymbol{w} = \boldsymbol{M}\boldsymbol{z} + \boldsymbol{q} \quad (32a)$$

$$w_j \geq 0, z_j \geq 0 \quad (j = 1, 2, \dots, J-1) \quad (32b)$$

$$w_j z_j = 0 \quad (j = 1, 2, \dots, J-1) \quad (32c)$$

where \boldsymbol{w} , \boldsymbol{z} and \boldsymbol{q} are all $2(J-1)$ vectors with $\boldsymbol{w} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{s} \end{bmatrix}$, $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{F}^{(t+1)} \\ \boldsymbol{\gamma} \end{bmatrix}$ and

$\boldsymbol{q} = \begin{bmatrix} \boldsymbol{g}^{(t)} - \boldsymbol{H}^{(t)}\boldsymbol{F}^{(t)} \\ -\boldsymbol{b} \end{bmatrix}$ whilst $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{H}^{(t)} & -\boldsymbol{A}' \\ \boldsymbol{A} & 0 \end{bmatrix}$ is a $2(J-1)$ square vector. Numerous fast

algorithms are available for solving the LCP (Cottle et al. 1992). In the empirical applications described in the next section we adopt a variant of the simplex algorithm.

Typically, each iteration is augmented by a line search between $\boldsymbol{F}^{(t)}$ and $\widehat{\boldsymbol{F}}^{(t+1)}$ (the vector of parameter estimates derived from solution of the LCP) in which a value for $\alpha^{(t)} \in [0, 1]$, the step size, is chosen such that the objective function (or more generally some *merit function*) falls sufficiently from one iteration to the next. In the empirical applications in the next section, we employ a line search procedure suggested by Jongbloed (1998). The parameters are updated according to;

$$\boldsymbol{F}^{(t+1)} = \boldsymbol{F}^{(t)} + \alpha^{(t)}(\widehat{\boldsymbol{F}}^{(t+1)} - \boldsymbol{F}^{(t)}) \quad (33)$$

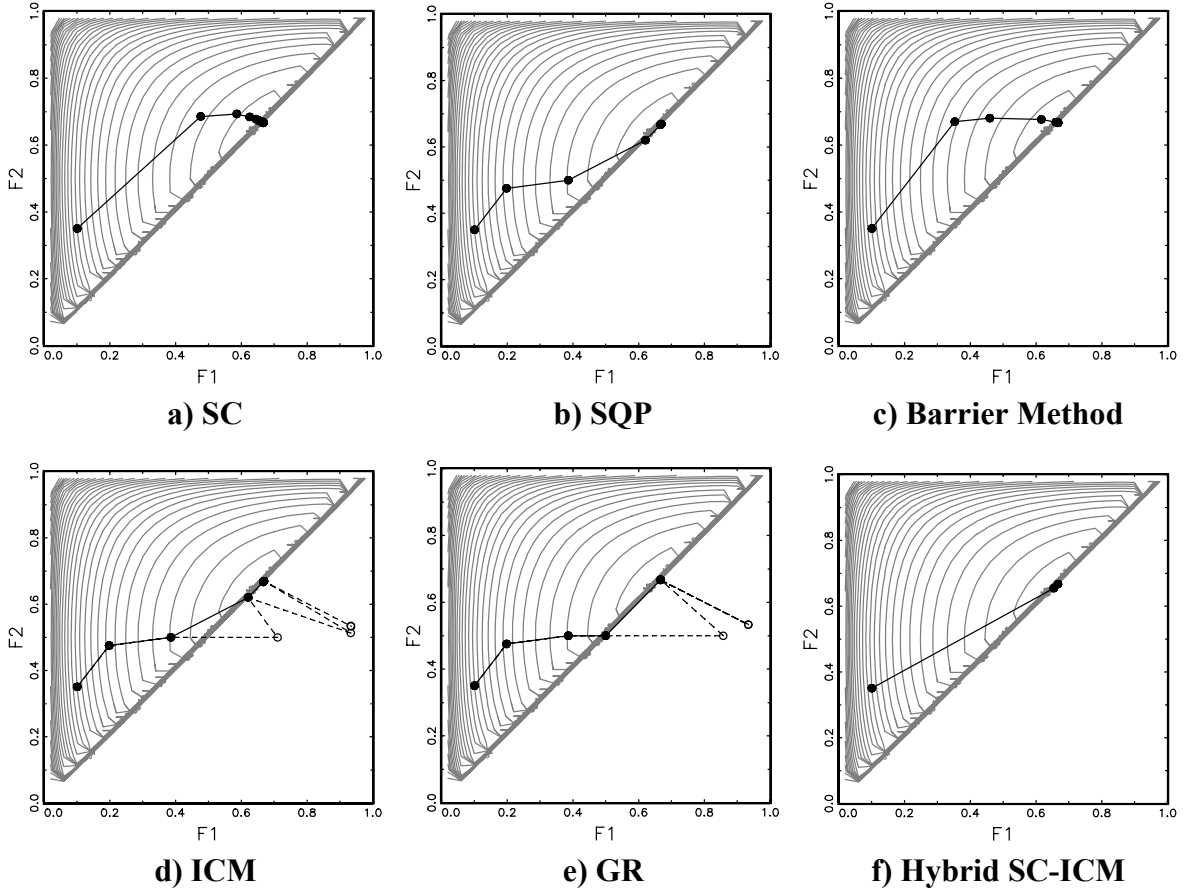
and iterations are continued until convergence.

Convergence results for SQP methods are well-established. Indeed, under appropriate assumptions, the iterative process should converge quadratically on a local maximum (see, for example, Theorem 3.1 in Boggs and Tolle, 1995).

To provide a comparison of the solution paths followed by the SQP algorithm and Turnbull's SC algorithm observe Figure 3. Once again, the log-likelihood surface illustrated in the various panels of Figure 3 is based on a hypothetical dataset consisting of fifteen observations defining three equivalence classes. In contrast to the data generating Figures 1 and 2, however, the log-likelihood function reaches a maximum on the boundary of the feasible parameter space at a location at which $F_1 = F_2$. Panel (a) of Figure 3 shows the vector of parameter values at each iteration of Turnbull's SC algorithm. Observe that Turnbull's algorithm remains strictly in the interior of the feasible parameter space and takes progressively shorter steps as it approaches the maximum. In contrast, Panel (b) shows the parameter values at each iteration of the SQP algorithm. Notice that the SQP algorithm hits

the boundary of the feasible parameter space after 3 iterations and makes its way to the maximum by working along this boundary.

Figure 3: Convergence paths for iterations of various NPML estimators



5.2 The Barrier Method

Another solution method for optimising a constrained problem such as (10) is provided by the barrier method (one of a large class of solution methods known as interior point methods). As a result of theoretical developments which showed their potential speed (Karmarkar, 1984), this class of solution methods has been the focus of considerable attention in the optimisation literature over recent years (see Forsgren et al., 2002, for a discussion of the history and development of interior point methods).

Whilst SQP progresses through simplifying the objective function and solving a sequence of constrained problems, the barrier method solves a sequence of unconstrained problems by modifying the objective function in such a way that the constraints cannot be violated. In particular, the demands of the constraints are captured by specifying a barrier function whose value tends to negative infinity as the parameter estimates approach the boundary of the feasible parameter space. At each iteration the barrier function is added to the objective function and the composite function maximised using standard methods of unconstrained optimisation.

Adopting the popular logarithmic barrier specification results in a composite function for the NPML problem as follows;

$$B(\mathbf{F}, \mu^{(t)}) = \ln L(\mathbf{F}) + \mu^{(t)} \sum_{j=1}^J \ln(F_j - F_{j-1}) \quad (34)$$

where $\mu^{(t)}$ is a weighting parameter for iteration t .

Iterations are initiated from a strictly interior point of the feasible parameter space. Choosing a ‘large’ initial value for μ ensures that the first iterate returns parameters that are well away from the boundary of the feasible space. At each iteration, the value for μ is systematically reduced such that the composite function $B(\mathbf{F}, \mu)$ increasingly resembles the original objective function. Indeed, intuition suggests that estimates of \mathbf{F} that maximise $B(\mathbf{F}, \mu)$ for a sequence of positive μ values converging on zero, will be the NPML parameter estimates themselves. Indeed, given appropriate assumptions, the limit point of the parameter estimates from such a sequence can be shown to converge on the global solution to the constrained problem (see, for example, Theorem 3.8 in Forsgren et al., 2002).

The convergence path for iterates of the barrier method are shown in Panel (c) of Figure 3. Notice how the iterates remain strictly within the feasible parameter space such that the algorithm approaches the maximising location on the boundary from the interior.

5.3 Iterative Convex Minorant Algorithm

Maximising the quadratic approximation to the NPML problem (29) indicates that a possible update at each iteration, t , is provided by;

$$\tilde{\mathbf{F}}^{(t+1)} = \mathbf{F}^{(t)} + \mathbf{W}^{(t)-1} \mathbf{g}^{(t)} \quad (35)$$

When $\mathbf{W}^{(t)}$ is set equal to $-\mathbf{H}^{(t)}$ this is simply a standard step in Newton’s method for unconstrained optimisation. As is well known, a whole class of Newton-like methods can be derived by setting $\mathbf{W}^{(t)}$ equal to some other positive definite matrix. For example, if $\mathbf{W}^{(t)} = \mathbf{W} = \mathbf{I}$, where \mathbf{I} is the identity matrix, then we have the method of steepest ascent.

Of course, Newton-like methods do not guarantee that the parameters at each iterate will fall within the feasible parameter space. One solution to this problem is provided by the generalised gradient projection (GGP) method (Bertsekas, 1982). At each iteration, the parameter values indicated by a Newton-like step are projected back onto the feasible parameter space, denoted Ω , in the metric of \mathbf{W} . That is to say, at each iteration the parameter updates are given by;

$$\mathbf{F}^{(t+1)} = \arg \min_{\mathbf{F}} \left\{ \left(\mathbf{F} - \tilde{\mathbf{F}}^{(t+1)} \right)' \mathbf{W} \left(\mathbf{F} - \tilde{\mathbf{F}}^{(t+1)} \right) \mid \mathbf{F} \in \Omega \right\} \quad (36)$$

Obviously, if $\tilde{\mathbf{F}}^{(t+1)}$ is itself in the feasible space, then the GGP method will take a full Newton-like step.

Pan (1999), showed that the iterative convex minorant (ICM) algorithm originally proposed by Groenboom (1991) and Groenboom and Wellner (1992) to solve the NPML problem was actually a GGP method. Specifically, the ICM algorithm is a GGP method in which $\mathbf{W}^{(t)} = -\mathbf{D}^{(t)}$, the diagonal matrix with the same diagonal elements as $-\mathbf{H}^{(t)}$. In this special

case, the update in (35) and the projection in (36) reduce to a rather simple sequential calculation⁷.

- To begin with, set the parameter updates to their unconstrained Newton-like updates; that is $\mathbf{F}^{(t+1)} = \tilde{\mathbf{F}}^{(t+1)}$. Then, beginning with $j = 1$ check to see if these updates comply with the monotonicity constraints.
- If $F_j^{(t+1)} < 0$ then update $F_j^{(t+1)} = 0$ and move on to consider the next parameter.
- If $F_j^{(t+1)} \leq F_k^{(t+1)}$ for $k = j+1, j+2, \dots, J-1$, then the Newton-like update does not violate the monotonicity constraints for this parameter and we can move on to consider the next parameter.
- If the monotonicity constraints are violated for $F_j^{(t+1)}$ then beginning with $k=1$ we create a new, 'pooled' update according to;

$$F_j^{(t+1)} = \frac{\sum_{l=j}^{j+k} (D_{ll}^{(t)} F_l^{(t)} - g_l^{(t)})}{\sum_{l=j}^{j+k} D_{ll}^{(t)}} \quad (37)$$

where $D_{ll}^{(t)}$ is the (l,l) th element of $\mathbf{D}^{(t)}$, and set $F_l^{(t+1)} = F_j^{(t+1)}$ ($l = j+1, \dots, j+k$). After each pooling we check to see if the monotonicity constraints now hold for $F_j^{(t+1)}$. If the constraints are still violated then the pool is expanded by incrementing k until monotonicity with respect to $F_j^{(t+1)}$ is achieved and we can move on to consider the next parameter.

- Finally, any $F_j^{(t+1)}$ ($j=1,2,\dots,J-1$) greater than unity is set equal to 1.

Once again, the algorithm may be supplemented at each iteration with a line search between $\mathbf{F}^{(t)}$ and $\mathbf{F}^{(t+1)}$. Indeed, Jongbloed (1998) proposes a particular criterion by which the step size might be selected and proves that the ICM algorithm when augmented with such a line search is globally convergent. We employ Jongbloed's line search in our empirical application in the next section.

Panel (d) of Figure 3 illustrates iterates of the ICM algorithm. The hollow circles illustrate how certain iterations result in Newton-like steps that fall outside the feasible parameter space. The dashed lines indicate the projection of such steps back onto the feasible space by the ICM algorithm. Notice from the final iteration that once at the maximising location the algorithm projects directly back onto itself.

5.4 Generalised Rosen Algorithm

The estimator proposed by Rosen (1960) for solving constrained optimisation problems is an early example of a class of estimators referred to as active set methods. Rosen's algorithm

⁷ Groeneboom and Wellner (1992) describe how this computation has a graphical interpretation as the left derivatives of the greatest convex minorant of a cumulative sum diagram consisting of the points $P_0 = (0,0)$ and

$$P_j = \left(\sum_{k=1}^j D_{kk}^{(t)}, \sum_{k=1}^j (D_{kk}^{(t)} F_k^{(t)} - g_k^{(t)}) \right) (j = 1, 2, \dots, J-1).$$

begins in the interior of the feasible parameter space and progresses in Newton-like steps. If a Newton-like step traverses the boundary of the feasible parameter space then the step is curtailed at the boundary. A subsequent line search between this location on the boundary and the iterate's starting parameter values establishes whether the iterate makes the full step to the boundary or terminates somewhere in the interior of the parameter space. In the former case, the constraint defined by that boundary is added to the *active set*, Ψ . Since the constraints in Ψ are assumed to hold at a maximum, all subsequent iterates must move along, but not away from, the boundaries of the feasible region defined by the constraints in the active set. Similar to GGP methods, this is achieved by projecting each Newton-like step on to the null space of the set of active constraints.

Moreover, if other constraints are violated in ensuing iterations, these too are added to the active set. Iterations continue until the change in the absolute values of the parameters from one iteration to the next falls below some critical level (as per 25). At that stage, the algorithm checks the Lagrange multipliers associated with the active constraints. If each the Lagrange multiplier is non-negative then the parameter estimates maximise the NPML problem. Alternatively, if one or more of the Lagrange multipliers are negative then the constraint associated with the smallest Lagrange multiplier is dropped from the active set and iterations are resumed. Global convergence of Rosen's algorithm was established by Du and Zhang (1989).

Rosen's original algorithm assumed \mathbf{W} equal to the identity matrix. In this case, the Newton-like step reduces to the method of steepest ascent and at each iteration such steps are projected orthogonally on to the null-space of the constraints in the active set. Jamshidian (2004) proposed a generalisation of Rosen's algorithm in which \mathbf{W} is any positive definite matrix that can differ from iteration to iteration. In this case, projection of each Newton-like step onto the null-space of the active constraints is in the metric of \mathbf{W} . Jamshidian (2002) proved that this algorithm was globally convergent.

Zhang and Jamshidian (2004) have applied this generalised Rosen algorithm to the NPML problem. Moreover, by selecting $\mathbf{W}^{(t)} = -\mathbf{D}^{(t)}$, as in the ICM algorithm, they ensure that the projection is relatively simple to calculate.⁸ In particular, the NPML problem obliges parameter estimates to conform to the sequence of J constraints, $0 \leq F_1^{(t)} \leq F_2^{(t)} \leq \dots \leq F_{j-2}^{(t)} \leq F_{j-1}^{(t)} \leq 1$, then;

- if the first m constraints are in the active set then $F_j^{(t+1)} = 0$ ($j = 1, \dots, m$).
- if the last m constraints are in the active set then $F_j^{(t+1)} = 1$ ($j = J - m, \dots, J - 1$).
- if the active set does not constrain F_j (i.e. the j^{th} and $(j + 1)^{\text{th}}$ constraints do not hold) then its value is updated with a standard Newton-like step; $F_j^{(t+1)} = F_j^{(t)} - g_j^{(t)} / D_{jj}^{(t)}$.
- if the k^{th} to $(k + m)^{\text{th}}$ constraints are in the active set, whilst the $(k - 1)^{\text{th}}$ and $(k + m + 1)^{\text{th}}$ constraints are not, then the $m + 1$ parameters constrained by this sequence of constraints are updated according to;

⁸ Zhang and Jamshidian (2004) present their GR algorithm in terms of the interval probabilities $p_j = F_j - F_{j-1}$ ($j = 1, 2, \dots, J$). For continuity of exposition we present the same analysis in terms of F_j .

$$F_j^{(t+1)} = F_j^{(t)} + \frac{\sum_{l=k-1}^{k+m} g_l^{(t)}}{\sum_{l=k-1}^{k+m} D_{ll}^{(t)}} \quad j = (k-1, k, \dots, k+m) \quad (38)$$

Panel (e) of Figure 3 illustrates iterates of the GR algorithm with the hollow circles and dotted lines indicating projections. Notice how the first two steps of the GR algorithm are the same Newton-like steps as taken by the ICM algorithm⁹. The third iterate, however, steps outside the feasible parameter space and is curtailed at the boundary. The constraint defining this boundary is added to the active set and subsequent iterations are projected back onto this boundary until the algorithm converges.

5.5 Hybrid Algorithms

A final set of estimators comprise hybrid algorithms in which each step of the algorithm is augmented with a single self-consistency step. Wellner and Zhan (1997) proposed a hybrid ICM-SC algorithm and proved its global convergence. Panel (f) of Figure 3 illustrates iterates of the ICM-SC algorithm where each iteration is taken as a SC step followed by an ICM step. Notice the apparently rapid convergence of the hybrid algorithm. Likewise, Zhang and Jamshidian proposed a hybrid GR-SC algorithm in which each iteration involves a SC step followed by a GR step. In a similar vein, we commence iterations of our BM and SQP algorithms with a single SC step. Moreover, very occasionally, we observe that problems of machine precision prevent the SQP algorithm from finding an improving step when very close to a maximum.¹⁰ In such circumstances we find that nudging the algorithm with a single SC step precipitates rapid convergence.

⁹ Moreover, both algorithm's initial iterates closely resemble the full Newton steps made by the SQP algorithm. This will likely be the case when each observations's interval spans relatively few equivalence classes such that the Hessian matrix of the NPML problem contains relatively few non-zero off-diagonal elements

¹⁰ In our simulation studies, we encounter this problem only in our across-interval censored data set and then in less than 1% of our bootstrap samples.

6. MONTE CARLO COMPARISON OF NPML ESTIMATORS

Various authors have carried out simulation studies to compare the performance of different algorithms in solving the NPML problem for Type 2 interval-censored data. Jongbloed (1998) compares the ICM and SC algorithm and reports that the ICM algorithm massively outperforms Turnbull's SC algorithm both in terms of the number of iterations and the computing time needed to attain convergence. Zhang and Jamshidian (2004) compare Turnbull's SC algorithm with both naïve GR and ICM algorithms and hybrid GR-SC and ICM-SC algorithms. In their simulations they find that Turnbull's SC algorithm is by far the slowest to converge whilst the ICM and ICM-SC algorithms outperform the GR and GR-SC algorithms. Similarly, Wellner and Zhan (1997) compare the performance of Turnbull's algorithm and the hybrid ICM-SC algorithm but in the context of doubly-censored data (a censoring pattern in which the data contains both Type 1 interval-censored and point observations). Again the hybrid algorithm is found to converge hugely faster than Turnbull's SC algorithm. Moreover, the performance advantages of the hybrid algorithm are shown to increase with sample size.

The simulated data sets used in these previous studies differ significantly from that returned by a typical dichotomous-choice CV survey. In particular, in the simulated data sets, a unique interval is randomly generated for each observation such that the NPML problem will tend to have a very large number of equivalence classes (i.e J will be very large). In contrast, each observation in a dichotomous-choice CV survey will indicate an interval whose upper and lower bounds will be drawn from the small number of bid amounts (typically less than 10) stipulated in the survey design (see Section 2). Accordingly, the NPML problem for data from CV surveys will tend to have relatively few equivalence classes (i.e J will be small).

The Monte Carlo experiments we describe here, therefore, represent the first comparison of the performance of the SC, ICM, GR and hybrid algorithms to a data generating process that resembles that for dichotomous choice CV data. Moreover, as far as we are aware, our experiments are the first to compare the performance of these estimators to that of the SQP and BM algorithms in the context of the NPML problem for Type 2 interval-censored data.

Our experiments differ from previous studies in one further respect. Rather than randomly generating data from some known parametric distribution we choose to generate bootstrap samples from actual CV data sets. As described subsequently we make use of two different data sets, one from a DB dichotomous choice CV survey, the other from a OOHB dichotomous choice CV survey. In both cases we generate 1,000 bootstrap samples each containing 1,000 observations and record for each sample the number of iterations and computing time required to achieve convergence for each of the seven estimators described in the previous section.

For each bootstrap sample the problem is first reduced by identifying equivalence classes. Moreover, since both surveys specify less than 10 bid amounts, the same sets of equivalence classes are spanned by numerous observations in each bootstrap sample. We take advantage of this by first aggregating the data into a series of counts recording the number of observations spanning each particular sequence of equivalence classes. All subsequent calculations use the aggregated data, reducing their computational requirements from order N to order s ($s \leq \sum_{j=2}^J j$) floating operations.¹¹ Unlike previous simulation studies, therefore, the relative performance of the various algorithms should be independent of the sample size.

¹¹ Zhang and Jamshidian (2004) exploit similar computational gains in the somewhat simpler case of doubly censored data.

For each sample and each algorithm we use the same starting values, namely $F_j^{(0)} = j/J$ ($j = 1, 2, \dots, J-1$) which conform to (27), and the same convergence criteria, namely (28a-c) with $\varepsilon = 10^{-4}$. The algorithms and bootstrap procedure were written in the GAUSS matrix programming language¹² and the computations carried out on a Dell workstation with a 2.8GHz Intel Pentium 4 processor and 512Mb of RAM.

The first of the two data sets used in the analysis comes from a CV survey carried out in the Beijing metropolitan area of China in 1997 (Day and Mourato, 1998). The data consists of a sample of 829 responses to DB dichotomous choice questions eliciting WTP for surface water quality improvement in the Beijing region. The bid design used in this study is one commonly applied in DB dichotomous choice surveys. Each respondent is randomly allocated an initial bid amount, B_m ($1 < m < M$) (in this survey there were 7 positive bid amounts such that $M = 7$). A “no” response to the initial bid amount precipitates a second dichotomous choice question asking the respondent to consider the next lowest bid amount B_{m-1} , whilst a “yes” responses leads to a second question offering the next highest bid amount B_{m+1} . The consecutive nature of the bids used in this design leads to a particularly simple form of Type 2 interval-censored data. Observations are either below the lowest bid amount offered (left-censored), above the highest bid amount offered (right-censored) or contained entirely within an interval described by a single equivalence class. An and Ayala, (1996) describe this data pattern as *regular interval-censoring*. Moreover, the NPML estimates for this data describe an interior solution, though this will not necessarily be the case for each bootstrap sample. The results of the Monte Carlo experiments based on this data set are reported in Table 2.

Table 2: Monte Carlo comparisons of NPML estimators applied to a contingent valuation dataset with regular-interval censoring (based on 1,000 bootstrap samples)

Algorithm	No. of Iterations		Computing time (secs)	
	Mean	s.d.	Mean	s.d.
SC	44.6	4.65	0.0100	0.0007
ICM	128.1	44.25	0.0390	0.0125
ICM-SC	21.2	2.74	0.0117	0.0012
GRF*	129.3	46.91	0.0480	0.0172
GRF-SC	21.2	2.69	0.0130	0.0014
SQP	5.8	0.84	0.0075	0.0007
BM	5.2	0.39	0.0919	0.0216

* In one case the GRF algorithm failed to converge after 10,000 iterations. This iterate is not included in the calculation of the bootstrap statistics.

It is apparent from Table 2 that Turnbull’s SC algorithm performs rather well for this particular data configuration; a finding that is in sharp contrast to the simulation studies

¹² The data reduction procedures, the SC algorithm and parts of the ICM algorithm are adapted from the “Nonparametric density estimator for CVM data” written by Olvar Bergland of the Agricultural University of Norway. The SQP algorithm contains a solver for the linear complementarity problem written by Rob Dittmar of the Federal Reserve Bank of St. Louis. The rest of the code was written by the author and is available on request.

reported in the literature. Indeed, of the various algorithms, only the SQP algorithm converges faster. Furthermore, it is apparent that naïve applications of the ICM and GRF algorithms perform considerably worse than their hybrid counterparts converging over 3 times slower and requiring over 6 times more iterations. The poorest performance is recorded for the BM algorithm with convergence taking, on average, almost a tenth of a second.

Overall, however, the algorithms converge pleasingly quickly. Indeed, whilst comparisons are difficult because of differences in convergence criteria, the convergence times reported in Table 2 are several orders of magnitude smaller than those reported for equivalent-sized data sets in Jongbloed (1998) and Zhang and Jamshidian (2004). It appears that the comparatively simple censoring mechanism generating this type of CV data coupled with the application of algorithms that take advantage of a relatively small number of equivalence classes, yields very considerable performance improvements in the solving of NPML problems.

To put these results into perspective, applying the SQP algorithm to this data set would allow one to run a 1,000 iteration bootstrap procedure in just 7½ seconds. The comparative figure for application of the BM algorithm would be a little over 1½ minutes.

The second data set used in this study, comes from a CV survey carried out in Norfolk in the UK in 2003 (Bateman et al., 2004). The data comprise a sample of 1,254 responses to a OOH dichotomous choice question eliciting WTP for water quality improvements in the Norfolk Broads. The bid design for this survey paired together various combinations of 9 positive bid amounts ($M = 9$). Importantly the two bid amounts in any pair were not necessarily consecutive amounts in the bid design. Each respondent was randomly allocated a particular pair and asked one or more dichotomous choice questions which identified their WTP interval as being below the lower bid amount (left-censored), above the higher bid amount (right-censored) or contained in the interval between the two bid amounts. In this case, individual WTP intervals could span more than one equivalence class. An and Ayala, (1996) describe this data pattern as *across-interval censoring*. What is more, the NPML estimates for this data set describe a corner solution in which $\hat{F}_5^{ML} = \hat{F}_6^{ML}$, though again, this will not necessarily be the case for each bootstrap sample. The results of the Monte Carlo experiments based on this data set are shown in Table 3.

Table 3: Monte Carlo comparisons of NPML estimators applied to a contingent valuation dataset with across-interval censoring (based on 1,000 bootstrap samples)

Algorithm	No. of Iterations		Computing time (secs)	
	Mean	s.d.	Mean	s.d.
SC	1120.2	1250.01	0.3309	0.4137
ICM	42.7	5.31	0.0192	0.0019
ICM-SC	15.7	1.52	0.0126	0.0009
GRF	42.7	6.85	0.0229	0.0028
GRF-SC	15.0	3.41	0.0136	0.0025
SQP	5.2	1.03	0.0102	0.0015
BM	5.2	0.36	0.0338	0.0041

Whilst the SQP and hybrid algorithms still perform extremely well in solving the NPML problem for this second data set, the performance of Turnbull’s SC algorithm is much more

variable. On average Turnbull's SC algorithm takes 32 times longer than the SQP algorithm to converge. In practical terms, that means waiting 5½ minutes to run a 1,000 iteration bootstrap procedure, a procedure that the SQP algorithm could complete in just 10 seconds. Of course, averaging the performance statistics conceals considerable underlying variability. All the same, closer inspection of the bootstrap statistics reveals that the fastest convergence time achieved by the SC algorithm (.0847 seconds) is still some 8 times slower than the average SQP convergence time. Moreover for over 5% of the bootstrap iterations, Turnbull's SC algorithm took in excess of 1 second to converge. It appears that the comparatively poor performance of Turnbull's algorithm reported in previous simulation studies is replicated for this particular across-interval censored CV data set.

7. SUMMARY AND CONCLUDING REMARKS

A problem faced by CV analysts employing elicitation formats that return Type 2 interval-censored data is how to estimate NPML estimates of the WTP distribution. One method for deriving such estimates that has received particular attention in the environmental economics literature is the SC algorithm proposed by Turnbull. This paper has set out to document the NPML estimation problem and to study in detail the properties of Turnbull's estimator. Drawing on results well-documented in other fields of research we have shown that Turnbull's estimator does not completely solve the constrained maximum likelihood problem that defines the NPML estimates. Accordingly, it is not guaranteed that self-consistent estimates derived from Turnbull's algorithm will also be the unique NPML estimates. Nevertheless, in this paper we have characterised the problem and used our results to motivate a set of starting and stopping criteria (27 and 28 respectively) that ensure that Turnbull's SC algorithm converges to the unique NPML estimates.

Frequently, analysts may wish to calculate NPML estimates very many times, for example, in employing bootstrap procedure in order to determine nonparametric confidence intervals for the lower-bound mean WTP estimate calculated from NPML estimates. Unfortunately, numerous studies have reported that Turnbull's SC algorithm converges very slowly for certain configurations of Type 2 interval-censored data. Accordingly, we describe a series of alternative estimators that might be used to solve the NPML problem and compare their performance to that of Turnbull's SC algorithm using Monte Carlo experiments. Data for this purpose are drawn from two different CV surveys; one returning regular-interval censored data, the other returning across-interval censored data. As far as the author is aware, this is the first time these alternative estimators have been applied to CV data.

The Monte Carlo experiments reveal that Turnbull's SC algorithm performs reasonably well for the regular-interval censored data set but very poorly for the across-interval censored data set. In both cases, however, the SQP algorithm outperforms all the other algorithms, converging, on average, 30 times quicker than Turnbull's SC algorithm in the case of the across-interval censored data set. The take home message from this paper is that the established reliance on Turnbull's SC algorithm for estimating NPML estimates from Type 2 interval-censored CV data is ill-founded. The investigations reported in this paper suggest that a simple SQP algorithm may prove a more reliable and faster alternative.

References

- AN, M. Y., AND R. A. AYALA (1996). "Nonparametric estimation of a survivor function with across-interval-censored data", *Working Paper 96-02*, Duke University, Durham, NC.
- AYER, M., H. D. BRUNK, G. M. EWING, W. T. REID, E. SILVERMAN (1955). "An empirical distribution function for sampling with incomplete information", *Annals of Mathematical Statistics*, 26, pp. 641–647.
- BATEMAN, I., B. H. DAY, D. DUPONT AND S. GEORGIU (2004). "Ohh la la! Testing the one-and-one-half bound dichotomous choice elicitation method for robustness to anomalies", *CSEERGE Working Paper*, EDM-2004-06, University of East Anglia: UK.
- BERTSEKAS, D. P., (1982). "Projected Newton methods for optimization problems with simple constraints", *SIAM Control and Optimization*, 20, pp. 221-246.
- BISHOP, R. C. AND T. A. HEBERLEIN (1979). "Measuring values of extramarket goods: Are Indirect Measures Biased?", *American Journal of Agricultural Economics*, 61, pp. 926-930.
- BOGGS, P. T., AND J. W. TOLLE (1995). "Sequential quadratic programming", *Acta Numerica*, 4, pp. 1-51.
- BOMAN, M., G. BOSTEDT AND B. KRISTROM (1999). "Obtaining welfare bounds in discrete-response valuation studies: A nonparametric approach", *Land Economics*, 75, pp. 284-294.
- BOYLE, K. J., AND R. C. BISHOP (1988). "Welfare measurements using contingent valuation: A comparison of techniques", *American Journal of Agricultural Economics*, 70, pp. 20-28.
- CARSON, R. T., T. GROVES AND M. MACHINA (1999). "Incentive and informational properties of preference questions", Plenary Address, European Association of Environmental and Resource Economists, Compatibility Issues in Stated Preference Surveys. Oslo, Norway, June.
- CARSON, R. T., W. M. HANEMANN, R. J. KOPP, J. A. KROSNICK, R. C. MITCHELL, S. PRESSER, P. A. RUUD AND V. K. SMITH (1994). "Prospective interim lost use value due to DDT and PCB contamination in the southern California Bight", NOAA Contract No.-DGNC-1-00007.
- CARSON, R. T., AND R. C. MITCHELL (1995). "Sequencing and nesting in contingent valuation surveys", *Journal of Environmental Economics and Management*, 28(2), pp.155-173.
- CARSON, R. T., AND D. STEINBERG (1990), 'Experimental design for discrete choice voter preference surveys', in *1989 Proceedings of the Survey Methodology Section of the American Statistical Association*. Washington, DC: American Statistical Association.
- CARSON, R. T., L. WILKS, AND D. IMBER (1994). "Valuing the preservation of Australia's Kakadu conservation zone", *Oxford Economic Papers*, 46, pp. 727-749.
- COOPER, J., W.M. HANEMANN, AND G. SIGNORELLO (2002) "One-and-One-Half-Bound Dichotomous Choice Contingent Valuation", *Review of Economics and Statistics*, 84(4), pp. 742-750.
- COTTLE, R.W., J.-S. PANG, AND R.E. STONE (1992), *The Linear Complementarity Problem*, Academic Press, Boston.
- DAY, B.H., AND S. MOURATO, (1998). "Willingness to pay for water quality maintenance in Chinese rivers", *CSEERGE Working Paper*, WM 98-02, University College London: UK.

- DEMPSTER, A. P., N. M. LAIRD AND D. B. RUBIN (1977). "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society, Series B*, 39, pp. 1-38.
- DU, D-Z, AND X-S, ZHANG (1989). "Global convergence of Rosen's gradient projection method", *Mathematical Programming*, 44(3), pp. 357-366.
- EFRON, B., (1967). "The two-sample problem with censored data", *Proceedings of Fifth Berkley Symposium on Mathematical Statistics and Probability*, University of California Press, 4, pp. 831-853.
- FORSYTH, A., P. E. GILL AND M. H. WRIGHT (2002). "Interior methods for nonlinear optimization", *SIAM Review*, 44(4), pp.525-597.
- GENTLEMAN, R., AND C. J. GEYER (1994). "Maximum likelihood for interval censored data: Consistency and computation", *Biometrika*, 81, pp. 618-623.
- GIRAUD, K. L., J. B. LOOMIS AND J. C. COOPER (2001). "A comparison of willingness to pay estimation techniques from referendum questions", *Environmental and Resource Economics*, 20, pp331-346.
- GOETGHEBEUR, E. AND L. RYAN. (2000). "Semiparametric regression analysis of interval-censored Data", *Biometrics*, 56, pp. 1139-1144.
- GROENEBOOM, P., (1991). "Nonparametric maximum likelihood estimators for interval censoring and deconvolution", *Department of Statistics Technical Report*, 378, Stanford University.
- GROENEBOOM, P., AND WELLNER, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Basel: Birkhäuser.
- GU, M.G., AND C.H. ZHANG (1993). "Asymptotic Convergence of the Self-Consistent Estimator of the Distribution Function Based On Doubly Censored Data", *Annals of Statistics*, 21, (2), pp. 611-624.
- HAAB, T. C., AND K. E. MCCONNELL (1997). "Referendum models and negative willingness to pay: Alternative solutions", *Journal of Environmental Economics and Management*, 32, PP. 251-270.
- HANEMANN, W. M., AND B. KANNINEN (1999). 'The statistical analysis of discrete response CV data', in *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the U.S., EU, and Developing Countries*, (eds) I. J. Bateman and K. G. Willis, Oxford University Press, Oxford.
- HANEMANN, W.M., J. LOOMIS, AND B. KANNINEN (1991) "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation", *American Journal of Agricultural Economics* 73(4), 1255-1263.
- HOEHN, J. P. AND A. RANDALL (1987), "A satisfactory benefit cost indicator from contingent valuation", *Journal of Environmental Economics and Management*, vol. 14, pp. 226-247.
- HUANG, J., AND J. A. WELLNER (1996). *Interval censored survival data: A review of recent progress*, Proceedings of the first Seattle Symposium in Biostatistics: Survival Analysis. Lecture notes in Statistics, vol. 123, pp. 123-169. Springer-Verlag, Berlin.
- HUHTALA, A. H., (1994). *Is Environmental Guilt a Driving Force? An Economic Study of Recycling*, PhD dissertation, Department of Agricultural and Resource Economics, University of California, Berkeley.

- HUHTALA, A. H., (2000). "Binary choice valuation studies with heterogeneous preferences regarding the program being valued", *Environmental and Resource Economics*, 16, pp.263–279.
- HUTCHINSON, G., R. SCARPA., S. CHILTON AND T. MCCALLION (2001). "Parametric and non-parametric estimates of willingness to pay for forest recreation in Northern Ireland: a multi-site analysis using discrete choice contingent valuation with follow-ups", *Journal of Agricultural Economics.*, 52(1), pp.104-122.
- JAMSHIDIAN, M., (2004). " On algorithms for restricted maximum likelihood estimation", *Computational Statistics and Data Analysis*, 45, pp. 137-157.
- JONGBLOED, G., (1998). "The iterative convex minorant algorithm for nonparametric estimation", *Journal of Computational and Graphical Statistics*, 7, pp. 310-321.
- KARMAKAR, N., (1984). "A new polynomial-time algorithm for linear programming", *Combinatorica*, 4, pp. 373-395.
- KERR, G. N., (2000). "Dichotomous choice contingent valuation probability distributions", *The Australian Journal of Agricultural and Resource Economics*, 44(2), pp. 233-252.
- KRISTRÖM, B., (1990). "A non-parametric approach to the estimation of welfare measures in discrete response valuation studies", *Land Economics*, 66, pp. 135-139.
- MCFADDEN, D., (1994). "Contingent valuation and social choice", *American Journal of Agricultural Economics*, 76, pp. 689-708.
- MYKLAND, D. A., AND J.-J. REN (1996). "Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data", *Annals of Statistics*, 24, pp. 1740-1764.
- NETTLETON, D., (1999). "Convergence properties of the EM algorithm in constrained parameter spaces", *The Canadian Journal of Statistics*, 27(3), pp. 639-648.
- NOAA (1993). "Appendix 1: Report of the NOAA panel on contingent valuation", *Federal Register*, 58, pp. 4602-2614.
- NUNES, P., AND J. VAN DEN BERGH (2004). "Can people value protection against invasive marine species? Evidence from a joint TC-CV survey in the Netherlands", *Environmental and Resource Economics*, 28(4), pp.517-532.
- PAN, W., (1999) "Extending the iterative convex minorant algorithm to the Cox model for interval censored data", *Journal of Computational and Graphical Statistics*, 8, pp. 191-200. (Also Report 1997-013, Division of Biostatistics, University of Minnesota, 1997)
- READY, R.C., AND D. HU (1995). "Statistical approaches to the fat tail problem for dichotomous choice contingent valuation", *Land Economics*, 71(4), pp. 491-499.
- ROSEN, J. B., (1960). "The gradient projection method for nonlinear programming", *Journal of the Society for Industrial and Applied Mathematics*, 8, pp. 181-217.
- SCARPA, R., K.WILLIS AND G. GARROD (2001). "Estimating benefits for effective enforcement of speed reduction from dichotomous-choice CV: The case of rural trunk roads", *Environmental and Resource Economics*, 20(4), pp 281-304.
- TURNBULL, B., (1974). "Nonparametric estimation of a survivorship function with doubly censored data", *Journal of the American Statistical Association*, 69, pp. 169-173.

- TURNBULL, B., (1976). "The empirical distribution function with arbitrary grouped, censored and truncated data", *Journal of Royal Statistical Society, Series B*, 38, pp. 290-295.
- WELLNER, J. A., AND Y. ZHAN (1997). "A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data", *Journal of the American Statistical Association*, 92, pp. 945-959.
- WU, C. F. J., (1983). "On the convergence properties of the EM algorithm". *Annals of Statistics*, 11(1), pp.95--103.
- ZHANG, Y., AND M. JAMSHIDIAN (2004). "On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data", *Journal of Computational and Graphical Statistics*, 13(1), pp. 123-140.