



Centre for Environmental and Resource Economics

# Approach to Analysis of Self-Selected Interval Data

Yuri Belyaev

Centre of Biostochastics, SLU-Umeå, [Yuri.Belyaev@sekon.slu.se](mailto:Yuri.Belyaev@sekon.slu.se)

Bengt Kriström

CERE, SLU-Umeå and Umeå University, [Bengt.Kristrom@sekon.slu.se](mailto:Bengt.Kristrom@sekon.slu.se)

**Centre for Environmental and Resource Economics (CERE)** is an inter-disciplinary and inter-university research centre at the Umeå Campus, Umeå University and the Swedish University of Agricultural Sciences. The main objectives with the Centre is to tie together research groups at the different departments and universities; provide seminars and workshops within the field of environmental & resource economics and management; constitute a platform for a creative strong research environment within the field.

CERE Working Paper, 2010:2

[www.cere.se](http://www.cere.se)

Department of Economics, Umeå Universitet  
S-901 87, Umeå, Sweden

# Approach to Analysis of Self-Selected Interval Data

Yuri Belyaev and Bengt Kriström

February 12, 2010

# APPROACH TO ANALYSIS OF SELF-SELECTED INTERVAL DATA<sup>1</sup>

YURI BELYAEV<sup>a</sup> AND BENGT KRISTRÖM<sup>b</sup>

We analyze an approach to quantitative information elicitation in surveys that includes many currently popular variants as special cases. Rather than asking the individual to state a point estimate or select between given brackets, the individual can self-select any interval of choice. We propose a new estimator for such interval censored data. It can be viewed as an extension of Turnbull's estimator (Turnbull (1976)) for interval censored data. A detailed empirical example is provided, using a survey on the valuation of a public good. We estimate survival functions based on a Weibull and a mixed Weibull/exponential distribution and prove that a consistent maximum likelihood estimator exists and that its accuracy can be consistently estimated by re-sampling methods in these two families of distributions.

KEYWORDS: Interval data, Maximum Likelihood, Turnbull estimator, willing-

---

<sup>1</sup>Department of Forest Economics, 901 83 Umeå, Sweden. The research presented in this paper was carried out as a part of the R&D programme "Hydropower - Environmental impacts, mitigation measures and costs in regulated waters". It has been established and financed by Elforsk, the Swedish Energy Agency, the National Board of Fisheries and the Swedish Environmental Protection Agency. Partial funding from the "PlusMinus Program" (<http://www.plusminus.slu.se>) is also gratefully acknowledged. We thank the participants at the 15th Ulvön Conference on Environmental Economics, Ulvön, Sweden, June 17-19, 2008, 16th Annual Conference of the European Association for Environmental and Resource Economics, Göteborg, Sweden, June 25-28, 2008, IISA Conference Jan 4-8, 2010, Visakhapatnam, (Andhra U., Waltair), India, and seminar participants at the Department of Mathematics & Mathematical Statistics, University of Umeå and Department of Forest Economics, SLU-Umeå, for useful input on earlier versions. Special thanks to Cecilia Håkansson for contributing the data and providing other useful input to the paper. Magnus Ekström and Jun Yu provided detailed and very useful comments. The usual disclaimer applies.

<sup>a</sup>Department of Forest Economics SLU, 901 83 Umeå, Sweden.  
[yuri.belyaev@sekon.slu.se](mailto:yuri.belyaev@sekon.slu.se); <http://www.sekon.slu.se>

<sup>b</sup>Centre for Environmental & Resource Economics (CERE), 901 83 Umeå, Sweden.  
Sweden. [bengt.kristrom@sekon.slu.se](mailto:bengt.kristrom@sekon.slu.se); <http://www.cere.se>

ness-to-pay, quantitative elicitation.

## 1. BACKGROUND

Surveys make up the life-blood of empirical research in the social sciences in general and in economics in particular. Employment surveys, investment surveys, inflation surveys are traditional examples, to which we can add the growing recent literature on the valuation of public goods. A key issue in any survey is the elicitation architecture, i.e. the way of eliciting information from the respondent.<sup>1</sup> The choice is essentially between two types of survey questions, the open-ended and the closed-ended.<sup>2</sup> We analyze an interval type of question in surveys that includes many currently popular variants as special cases. Rather than asking the individual to state a point estimate or select between given brackets, the individual can self-select any interval of choice. This paper is a part of our research program on self-selected interval questions, [Håkansson \(2008\)](#), [Belyaev, Håkansson and Kriström \(2008\)](#), present background empirical analysis. This paper takes the next step, by proposing statistical (and economic) theory to support

---

<sup>1</sup>Perhaps the closest literature to our general approach is significant body of literature in psychology, statistics and survey research that provides approaches to elicit probability distributions, for a survey, see e.g. [Garthwaite, Kadane and O'Hagan \(2004\)](#). A compact survey of many issues in survey research, in particular regarding response errors and biases across formats is given in [McFadden et al. \(2005\)](#).

<sup>2</sup>Each of these can be further sub-divided into several categories. For example, closed-ended questions can be based on a Likert-scale (e.g. from "strongly disagree" to "strongly agree"), a multiple choice (e.g. 'circle one of the following alternatives'), an ordinal question (e.g. 'rank the following items from 1 to 5'), and a binary question ("are you willing to pay x USD for this public good" (yes, no)). There are also several variants of the open-ended questions, such as "How much are you willing to pay for this public good?" or "How did you make that choice?". The choice between the open and closed-ended questions is not straightforward, because they have advantages and disadvantages in different situations, see [Fink \(1985\)](#).

1 the use of such interval questions. 1

2 As argued below, this version could reduce a number of biases, provides 2  
 3 a richer picture of response uncertainty, potentially increase response-rates 3  
 4 and maintains a link to recent ideas on coherent arbitrariness. These argu- 4  
 5 ments are necessarily heuristical, because the empirical evidence is scant, a 5  
 6 point we will return to below. 6

7 Potential applications include, but are not limited to: recall situations 7  
 8 ("How many days were you unemployed the first quarter of last year?", 8  
 9 "What was your net income the previous year of taxation"?), projections 9  
 10 ("What is your best forecast of the next year's interest rate?") or contin- 10  
 11 gent valuation studies ("How much are you maximally willing to pay for the 11  
 12 suggested change?"). As [Manski and Molinari \(2010\)](#) points out intervals 12  
 13 are more common in daily communication than we ordinarily think. Thus, 13  
 14 weather reports and pilot communications include a form of implicit inter- 14  
 15 val, e.g. when meteorologists report that the wind blows from the "north 15  
 16 means that the wind direction lies in the interval [337.5., 22.5]." This kind 16  
 17 of rounding is also prevalent in many types of surveys; it is well-documented 17  
 18 that individuals often round their answers to open-ended survey questions, 18  
 19 see [Rosch \(1975\)](#), [Schaeffer and Bradburn \(1989\)](#), [Huttenlocher, Hedges and](#)  
 20 [Bradburn \(2008\)](#), [Hurd et al. \(1998\)](#), [Hobbs \(2004\)](#), and [van Exel et al.](#)  
 21 [\(2006\)](#). [Manski and Molinari \(2010\)](#) claims the The University of Michigan 21  
 22 Health and Retirement Study (HRS) surveys will, in the 2008 and 2009 22  
 23 rounds, add an interval option to the current point-type question about a 23  
 24 certain probability. The basic problem is that when a respondent is to re- 24  
 25 port a point, he sometimes round it to describe a sentiment that really is 25  
 26 an interval. [Manski and Molinari \(2010\)](#) present an approach to deal with 26  
 27 such intervals, which is different from the one suggested here. The differ- 27  
 28 ence arises partly because we ask the respondent to state a point or an 28  
 29 interval, not both. In addition, we take the view that the individual chooses 29

a particular interval (from unobserved personal set of admissible intervals) and state it. We try also to find out more probable individuals' behavior in selection of stated intervals and the corresponding WTP-distribution.

Consider, as an example of the standard interval type approach ("bracketing") the Quarterly Survey of Professional Forecasters (SPF) used by the European Central Bank (<http://www.ecb.int/stats/pdf/spfquestionnaire.pdf>). The 2007 version of the questionnaire asks designated experts for estimates of the future inflation, Gross Domestic Product growth and unemployment rates, coupled with probabilities for different outcomes. Thus, for example, intervals for inflation are given as ( $< 0\%$ ,  $0.0 - 0.4\%$ ,  $0.5 - 0.9\%$ ,  $1.0 - 1.4\%$ ,  $1.5 - 1.9\%$ ,  $2.0 - 2.4\%$ ,  $2.5 - 2.9\%$ ,  $3.0 - 3.4\%$ ,  $\geq 3.5\%$ ) and the respondents have to state their expectations for 2008 and 2009, by assigning probabilities for each outcome. A disadvantage with this type of bracketing approach is the possibility of starting point bias, or "bracketing effect", a phenomena that has been extensively studied and documented (see [McFadden et al. \(2005\)](#)). Briefly, in split-samples one often finds significant differences between responses depending on the chosen bracket structure.<sup>3</sup> An advantage with the interval approach suggested here is that we avoid such effects. In addition, we avoid the tendency of choosing a bracket "somewhere in the middle".<sup>4</sup>

Another advantage with the self-selected intervals is that they arguably

---

<sup>3</sup>For a recent example in an economic context, see [Winter \(2004\)](#)

<sup>4</sup>There is a similar method in the related literature referred to above. [Garthwaite, Kadane and O'Hagan \(2004\)](#) distinguishes between the *fixed interval method* and the *variable interval method*. In the first case, the respondent is asked to assess the probability that X is within a set of intervals proposed by the investigator (the constraint that probabilities sum to one is imposed). In the second, the respondent is asked to state the upper and lower quartiles for X (the maximum amount of tomorrow's precipitation, for example). The interval has a specified probability (e.g. 50% chance that the interval will cover the true value).

1 provide a richer picture of any underlying response uncertainty, compared 1  
 2 to bracketing and some recent approaches to cater for respondent uncer- 2  
 3 tainty in contingent valuation.<sup>5</sup> The currently most popular approach in 3  
 4 contingent valuation is a payment card containing several different costs 4  
 5 for a public good, combined with a question about how certain the respon- 5  
 6 dent feels about paying a certain cost (e.g. "definite yes", "probably yes", 6  
 7 "probably no" and "definite no"). Recent analysis shows that including such 7  
 8 uncertainty-assessments in the survey instrument may affect the estimate 8  
 9 of valuations. For a review of this literature in this area see [Broberg and](#) 9  
 10 [Brännlund \(2008\)](#). The intervals do not burden the respondent with the task 10  
 11 of categorizing his uncertainty about a certain quantity and no issues arise 11  
 12 as to how such categories should be represented in an econometric model.<sup>6</sup> 12

13 Furthermore, there is an interesting connection to recent ideas in eco- 13  
 14 nomic psychology on coherent arbitrariness. In [Ariely, Loewenstein and Pr-](#) 14  
 15 [elec \(2003\)](#), individual's valuations of private goods are shown in a set of 15  
 16 experiments to be anchored on some arbitrary initial price, but the values 16  
 17 change coherently with conditions. In our application, we do not suggest a 17  
 18 price for the public good under consideration, so in this sense there is no 18  
 19 direct connection to the coherent arbitrariness hypothesis. Still, a person 19  
 20 might have difficulty stating a precise value for a public good and prefer to 20  
 21 state, as in [Ariely, Loewenstein and Prelec \(2003\)](#), a "range of acceptable 21  
 22 values". As the provision of the public good increases, we might see the 22  
 23 interval "shifted to the right", which arguably is a behavior consistent with 23  
 24 24

---

25 <sup>5</sup>As noted, the SPF-survey does include a way for the respondent to submit his un- 25  
 26 certainty about certain outcomes in a more direct manner compared to the self-selected 26  
 27 interval approach. The fixed and variable interval methods are similar in this regard. 27

28 <sup>6</sup>Each category of uncertainty can be considered in terms of a threshold-parameter in 28  
 29 an ordered logit(probit) model. Assumptions are usually imposed on those parameters, 29  
 an issue avoided here.

the coherent arbitrariness hypothesis.<sup>7</sup>

As a final point not directly related to our statistical modelling, we note that a standard open-ended question tend to give lower response rates [McFadden et al. \(2005\)](#). The question then arises if the self-selected intervals give higher response rates. There is scant evidence on this issue, but some encouraging results are provided in [Håkansson \(2008\)](#). Further research on the issue of response rates across elicitation methods is needed.

A number of disadvantages have been demonstrated with the standard elicitation methods. For example, open-ended questions tend to depress response rates, while the close-ended question necessarily gives much more limited information. In addition, when an individual is offered a price to reject or accept (as in contingent valuation), there is a tendency for anchoring around the price; while there is no intended information content the individual may well anchor his valuation around the suggested price. In the psychological literature, this phenomena was documented in the beginning of the 1970s<sup>8</sup>. A useful overview on psychologists' research about "How the question shapes the answers", is given by [Schwartz \(1999\)](#). We do concede the point that our arguments in favor of the self-selected interval question are partially heuristical. There simply is not enough empirical evidence that allows any definite statement about the "best" elicitation approach. Even so, all existing elicitation methods are special cases of the self-selected intervals, and we think the approach merit further analysis.

Turning now to the specific statistical problem analyzed in this paper, we need to develop an estimator that allows us to estimate the distribution of

---

<sup>7</sup>Using a different type of valuation question, a payment ladder, [Hanley, Kriström and Shogren \(2008\)](#) empirically explores the concept in the context of a public good.

<sup>8</sup> [Slovic \(1972\)](#) and [Tversky and Kahnemann \(1974\)](#). See, however, [Kynn \(2007\)](#) for a candid review of these early, and very influential, papers. In particular, Kyle's review suggests that the extent to which individual's use heuristics for substantive decisions is unclear.



1 the variable of interest, given that the data are censored in a non-standard  
2 manner. Thus, in our data, we either observe a (potentially rounded) data-  
3 point exactly, or that the data is coarsened and the point is "hidden" by a  
4 self-selected interval. At first glance, such interval censored data would seem  
5 to be easily handled by standard methods developed in the statistical litera-  
6 ture. A general solution to the problem of (non-parametric) maximum like-  
7 lihood (ML-)estimation for censored data was obtained by [Turnbull \(1976\)](#),  
8 in a well-known paper. See also [Jammalamadaka and Mangalam \(2003\)](#).  
9 Because the censoring mechanism is not random in the standard sense, we  
10 need to take into account the coarsening mechanism in a novel manner. In-  
11 tuitively, if bracketing is used, the censoring mechanism is random from the  
12 individual's point of view. In our case, the individual selects his preferred  
13 interval from a subset of intervals that are unknown to the investigator. Our  
14 solution to the estimation problem therefore involves two different probabil-  
15 ities, the first being related to the choice of interval, the second to the condi-  
16 tional probability that the unknown value belongs to a given interval. This  
17 solution is a natural consequence of the way information is elicited when  
18 using the self-selected interval approach. The comparison with the usual ap-  
19 proach to maximum likelihood estimation with interval data is therefore not  
20 straightforward. Nevertheless, we provide an illustrative example invoking  
21 the Turnbull estimator as if the data had been generated by the standard  
22 bracketing approach, in which the brackets are chosen by the investigator,  
23 not the individual. It should be noted that our comparison does not cater  
24 for the fact that a suggested sequence of brackets may be unsuitable from  
25 the individual's point of view. Proper comparison of our approach with the  
26 bracketing approach will be a focus of our continued research.

27 The rest of the paper is structured as follows. In [Section 2](#) we introduce  
28 a very simple economic model that pins down what we want to measure  
29 and detail our basic assumptions. We re-interpret a model suggested in

Hanemann, Kriström and Li (1996) to handle preference uncertainty in valuation studies. The empirical data is introduced in Section 3 and serve as a bridge to the statistical model introduced in Section 4. In Section 5 we apply the proposed ML-estimator and compare it with Turnbull's. The final Section 6 has concluding remarks. The Appendix sketches proofs of existence and consistency of the ML-estimators and consistent estimation of accuracy using resampling methods.

## 2. BEHAVIORAL ASSUMPTIONS AND THE ECONOMIC MODEL

We offer the following simple model to pin down what we want to measure, i.e. a measure of the public's willingness-to-pay (WTP) for a public good. Because the respondent's WTP is an interval, the welfare economic interpretation is subtle. Our line of attack is based on Hanemann, Kriström and Li (1996). We assume that each individual has an underlying concave smooth utility function  $U(c, q)$ , increasing in both its arguments, where  $c$  is money income and  $q$  is an index of environmental quality. The Hicksian compensating variation (WTP)  $x$  for an environmental improvement from  $q_0$  to  $q_1 (q_1 > q_0)$  is then defined by the relation

$$(2.1) \quad U(c_0, q_0) = U(c_0 - x, q_1),$$

where  $c_0$  denotes income in the status quo.

Let the set of individuals' incomes in a population be described by a probability distribution. Then for the  $i^{th}$  individual the unobserved value of compensating variation  $x = x_i$  in (2.1) is a value of a random variable (r.v.)  $X_i$  with distribution function (d.f.)  $F[x] = P[X_i \leq x], x \geq 0$ . We assume that the  $i^{th}$  individual does have a true point of valuation for the change of  $q$  but cannot state it with certainty. Thus, let the uncertainty of particular individual's valuation  $y$  be described by a random variable  $Y$ . Its conditional d.f. given the compensating variation  $x$  is denoted  $G[y | x] = P[Y \leq y | x]$ .

Let  $\mathbf{y}_i = (y_{Li}, y_{Ri})$  be an interval stated by the  $i^{th}$  individual, then  $\mathbf{y}_i$  is a value of a r.v.  $\mathbf{Y}_i = (Y_{Li}, Y_{Ri})$ . The conditional d.f. of  $\mathbf{Y}_i$  given the compensating variation is denoted  $G[\mathbf{y} | x] = P[Y_{Li} \leq y_L, Y_{Ri} \leq y_R | x]$ ,  $\mathbf{y} = (y_L, y_R)$ .

Our empirical data suggest that individuals prefer to state rounded values or intervals, selected from a finite set. To handle this possibility, we proceed as follows. If the  $i^{th}$  individual has stated a point value  $y_i$  then  $y_i$  belongs to a finite set  $\mathcal{U}_p$ ,  $y_i \in \mathcal{U}_p = \{u_1, \dots, u_{m_p}\}$ . All stated intervals are elements of a set  $\mathcal{U}_I = \{\mathbf{u}_1, \dots, \mathbf{u}_{m_I}\}$  with a finite number of intervals having rounded left and right ends. Therefore we assume that the conditional d.f.s  $G[y | x]$  or  $G[\mathbf{y} | x]$  are discrete. The element  $u_h \in \mathcal{U}_p$  and the interval  $\mathbf{u}_h \in \mathcal{U}_I$  is selected with conditional probability  $w_{ph}[x] = P[Y_i = u_h | x]$  and  $w_{Ih}[x] = P[\mathbf{Y}_i = \mathbf{u}_h | x]$ , respectively.

The probability to state the point value  $y_i = u_h$  or interval  $\mathbf{y}_i = \mathbf{u}_h$  can be written as

$$w_{ph} = P[Y_i = u_h] = \int_0^\infty w_{ph}[x] dF[x],$$

or

$$(2.2) \quad w_{Ih} = P[\mathbf{Y}_i = \mathbf{u}_h] = \int_0^\infty w_{Ih}[x] dF[x],$$

respectively. The functions  $w_{ph}[x]$ ,  $w_{Ih}[x]$ , and  $F[x]$  are unknown.

This model can be interpreted in several ways. Under preference uncertainty, the uncertainty arises because the individuals do not know their utility function exactly. As noted, this is a re-interpretation of the model in [Hanemann, Kriström and Li \(1996\)](#), which focuses on integrating valuation uncertainty into a microeconomic model. In their model, the individuals are uncertain about their WTP, but know that it is within a given interval (assumed symmetric and with the same length for all households). But it is possible to give this a more general interpretation; the individuals are uncertain about their valuation because of generic uncertainty about many

factors (suppressed here). These uncertainties are summarized by  $F[x]$  and  $G[y | x]$  or  $G[\mathbf{y} | x]$ . Alternatively, we can use the same set-up as in the industrial organization literature, in which the individual is uncertain about the quality of the good he buys. The uncertainty is not resolved until after the individual has experienced the good. Hence, for each given value of the quality-level, one obtains a particular value of WTP. This gives an interval of possible values. Thirdly, we can expand upon the traditional RUM-approach (see e.g. [McFadden et al. \(2005\)](#)) and add a state-dependent error term that summarizes response uncertainty. The intervals are then simply describing the support of these error terms (the standard error term is usually interpreted as the researcher's ignorance about the utility function). We do not develop these models here, because our focus is on the statistical approach, to which we now turn.

### 2.1. Assumptions

In order to estimate the distribution of WTP, we need to make a number of assumptions related to how individuals respond to the valuation question. We first introduce the following notion of admissibility:

**DEFINITION 1** *Let the compensating variation  $x$  be a true point of valuation and let  $\mathcal{C}(x) = \{\mathbf{u} : \mathbf{u} \in \mathcal{U}_I, x \in \mathbf{u}\}$ . Then any interval  $\mathbf{u} \in \mathcal{C}(x)$  is said to be admissible.*

We collect assumptions in the following.

**ASSUMPTION 1** *Each individual has one true point of compensating variation, but might not be aware of the exact location of this point. Respondents may freely round and state these points or admissible (potentially rounded) intervals.*

1           ASSUMPTION 2   *The true points of compensating variation are independent*  
 2           *of question mode (open-ended, a self-selected interval or a choice between*  
 3           *the two) and the question mode does not change the true points.*

4           ASSUMPTION 3   *The true points, the stated points and the intervals of*  
 5           *compensating variation corresponding to different respondents in a sample*  
 6           *are values of independent identically distributed (i.i.d.) random variables.*

8           Assumption 1 is supported by a shift in the contingent valuation literature  
 9           towards catering for preference uncertainty, as noted above. The question  
 10           of incentive-compatibility is beyond the scope of this paper, so we simply  
 11           assume that the individual tells the truth. Assumption 2 limits the set of  
 12           elicitation methods to open-ended and will be used below in a comparison of  
 13           the open-ended and the self-selected intervals. Assumption 3 is acceptable if  
 14           the number of sampled respondents is negligible relative to the population  
 15           of individuals.

16           Let  $\mathcal{P}_{\mathbf{u}}$  be a group of persons in a population of interest  $\mathcal{P}$  and suppose  
 17           that all persons in  $\mathcal{P}_{\mathbf{u}}$ , if being asked, would state the same interval  $\mathbf{u}$ . Let  
 18            $\mathcal{U} = \{\mathbf{u}\}$  be the set of all stated intervals. The size  $M$  of the set  $\mathcal{U}$  is not  
 19           known. Suppose that a random sample of  $n$  respondent was taken from  $\mathcal{P}$   
 20           and  $m$  different intervals  $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  were stated. In what sense can  
 21           we then say that  $m$  is sufficiently large for inference? If Assumptions 1 - 3  
 22           hold we can restate this question as follows. How large fraction  $q_s$  can be  
 23           excluded from  $\mathcal{P}$  such that  $m$  does not increase as  $n \rightarrow \infty$ ?

24           If Assumptions 1 - 3 hold then we can consider the sampling process  
 25           of stated intervals as a multinomial process with independent events. This  
 26           process is reduced to a binomial process if we collect "often stated" intervals  
 27           in the first group and the rest of the stated intervals and all not yet observed  
 28           intervals in the second group. We estimate the probability  $q_o$  to select an  
 29           interval from the first group by the empirical frequency  $\hat{q}_o$  and  $q_s = 1 - q_o$

by  $\hat{q}_s = 1 - \hat{q}_o$ . After that we find  $\gamma$ -confident interval  $(0, \hat{q}_{s\gamma}]$  containing  $q_s$ .  
 If  $\hat{q}_{s\gamma}$  is “small” then our inferences with fixed  $m$  will be valid for a slightly  
 reduced  $\mathcal{P}$ .

To fix ideas and further motivate the self-selected intervals, we reveal  
 salient features of our application in the next section.

### 3. EMPIRICAL DATA

A contingent valuation study with interval questions was carried out to  
 shed light on the costs and benefits of changing in stream flow at the Stornor-  
 rfors hydropower plant on the Vindel River, in northern Sweden. The sce-  
 nario entails reducing production of electricity, which would increase the  
 number of wild salmon in the river, as more water would be allocated to  
 salmon passage areas. The survey was carried out in the autumn of 2004.<sup>9</sup>  
 Respondents were asked about their WTP for increasing the number of  
 salmon that reach their spawning grounds in the river each year. Here we  
 consider a part of the sampled data from a general register of the Swedish  
 population (SPAR).<sup>10</sup> Our analysis is based on three subsets of the sampled  
 data. In the first sample  $S_1$ , we used a standard open-ended question. In  
 the second sample  $S_2$  we asked only about intervals and in the third sample  
 $S_3$  individuals were free to select either a point or an interval of choice. See  
 Table I for a summary of the data.

There is considerable heaping on a certain set of intervals. Thus, 142 out  
 of 241 respondents, in the samples  $S_2$  and  $S_3$ , stated the following four  
 WTP-intervals:  $(20, 50]$ ,  $(50, 100]$ ,  $(100, 150]$ ,  $(100, 200]$ . The numbers of  
 these stated intervals are 39, 11, 69 and 23, so that four “popular” intervals  
 make up an important part of the data. In Figure 1 we display the stated  
 intervals, ordering them by their left endpoints.

<sup>9</sup>For details about the study, see [Håkansson \(2008\)](#).

<sup>10</sup>All individuals less than 18 years of age were excluded from the register prior to  
 sampling.

TABLE I

Sample	# not answered	# of stated 0	# of stated intervals	# of stated points	total #
$S_1$	97	76	0	72	245
$S_2$	97	88	58	0	243
$S_3$	527	334	183	148	1192

Consider now the difference between the points and the intervals. We compare the empirical survival functions (s.f.s)  $sf_1[x]$  based on WTP-points in sample  $S_1$  and  $sf_2[x]$  on the values of right ends of intervals in sample  $S_2$ . The s.f.s are defined in the following manner:

$$(3.1) \quad sf_1[x] = \sum_{h=1}^{m_P} \frac{t_{ph}}{t_p} I[u_{ph} > x], \quad sf_2[x] = \sum_{h=1}^{m_I} \frac{t_{Ih}}{t_I} I[u_{Rh} > x],$$

where  $t_{ph}$  and  $t_{Ih}$  denote the number of statements of the point  $u_h$  and the right points  $u_{Rh}$  of intervals  $\mathbf{u}_h$  in  $S_1$  and  $S_2$ , and  $t_p = \sum_{h=1}^{m_P} t_{ph}$ ,  $t_I = \sum_{h=1}^{m_I} t_{Ih}$ .  $I[\cdot]$  is the indicator function. These empirical s.f.s are displayed in Figure 2.

The two survival functions for the sample  $S_1$  (thin stepwise line) and for the sample  $S_2$  (thick stepwise line) have jumps proportional to the number of statements points and intervals. The figure suggests that the stated points are rounded up to the right-hand ends of the intervals.

To gain further insight, we plotted the empirical s.f.s of the right intervals' ends and intervals' lengths in  $S_2$  and  $S_3$ , see Figure 3. It suggests there is no "treatment effect", i.e. whether or not the individual must state an interval in  $S_2$  or may choose an interval in  $S_3$  makes little difference. A similar comparison between the distributions of stated points in  $S_1$  and  $S_3$  shows a slight disparity. We find by calculating the area under the s.f.s that the estimated mean values are 139.236 and 177.135, respectively, with  $p$ -value for their difference (37.899) being 0.101.<sup>11</sup>

<sup>11</sup>We use resampling methods for estimation accuracy of statistical inferences in this

Unfortunately, the number of observations in samples  $S_1$  and  $S_3$  are rather small and it is not possible with confidence to reject the hypothesis that the distribution of the points in the sample  $S_1$  is the same as in  $S_3$ . If we reject the hypothesis then Assumption 2 does not hold. If we accept the hypothesis and stated points and right-hand ends of intervals are close to WTP, then this either contradicts Assumption 2, or  $S_1$  respondents rounded their true WTP-points upwards. The last possibility is consistent with our assumptions. Therefore, by Assumption 2, we are forced to conclude that the interval question is, given our assumptions, to prefer.

To sum up: the two intervals-samples are very similar, even though the numbers are based on two different "treatments". The WTP-point responses suggest a rounding upward towards the right-hand endpoints of the stated WTP-intervals.

Let us now return to the question about the size of  $m$ . We split the 241 intervals into two groups, the first having the intervals that are stated more than once, the second group then contains the single-stated intervals. In our data, the sizes of the first and the second groups are 220 and 21, respectively. Using the normal approximation for the distribution of  $\hat{q}_o = 220/241$  we obtain a  $\gamma$ -confident interval  $(0, \hat{q}_{s\gamma}] \simeq (0, 0.14]$  containing  $q_s$  with probability  $\gamma \geq 0.975$ . Hence, we obtain statistical inferences for distributions of large majority of WTP-points based on the observed value  $m$  of all different stated intervals. For samples taken from not less than  $86\% = (1 - 0.14)100\%$  part of the population  $\mathcal{P}$ ,  $m$  will be constant for all  $n > 241$ . Now we turn to our statistical model for interval data.

#### 4. DEFINITION OF THE STATISTICAL MODEL WITH ROUNDED INTERVALS

Suppose that the collected statistical data consist of  $n$  stated rounded intervals  $\mathbf{y}_i = (y_{Li}, y_{Ri}]$  containing unobserved true WTP-points of com-

---

paper, see the Appendix and [Belyaev \(2003, 2007\)](#).



1       pensating variation  $x_i \in \mathbf{y}_i$ ,  $i = 1, \dots, n$ . Let  $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_h, \dots, \mathbf{u}_m\}$  be       1  
2       the list with all different intervals  $\mathbf{u}_h = (u_{Lh}, u_{Rh}]$ ,  $u_{Lh} < u_{Rh}$ ,  $\mathbf{u}_{h_1} \neq \mathbf{u}_{h_2}$ ,       2  
3        $h_1 \neq h_2$ , and for each  $i \in \{1, \dots, n\}$  there is at least one  $\mathbf{y}_i = \mathbf{u}_{h_i}$ , i.e. the       3  
4       stated WTP-intervals  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\} = \{\mathbf{u}_{h_1}, \dots, \mathbf{u}_{h_n}\}$ . As above, let  $t_h$  be the       4  
5       number of cases when  $\mathbf{y}_i = \mathbf{u}_h$ . The collected statistical data can be written       5  
6       as the list  $\mathbf{dat}_m = \{\{\mathbf{u}_1, t_1\}, \dots, \{\mathbf{u}_h, t_h\}, \dots, \{\mathbf{u}_m, t_m\}\}$ ,  $\sum_{h=1}^m t_h = n$ .       6

7       Suppose that Assumptions 1 - 3 stated in Section 2 are valid. By As-       7  
8       sumption 3 we consider  $\{x_1, \mathbf{y}_1\}, \dots, \{x_n, \mathbf{y}_n\}$  as values of i.i.d. pairs of r.v.s       8  
9        $\{X_i, \mathbf{Y}_i\}$ ,  $\mathbf{Y}_i = (Y_{Li}, Y_{Ri}]$ , and let  $\{\mathbf{Y}_i \leq \mathbf{y}\} = \{Y_{Li} \leq y_L, Y_{Ri} \leq y_R\}$ .       9  
10       Their d.f.s  $F[x] = P[X_i \leq x]$  and  $G[\mathbf{y} | x] = P[\mathbf{Y}_i \leq \mathbf{y} | X_i = x_i]$        10  
11       are unknown. By Assumption 1  $X_i \in \mathbf{Y}_i$ . The stated rounded intervals       11  
12        $\mathbf{u}_h = (u_{Lh}, u_{Rh}]$ ,  $h = 1, \dots, n$ , can overlap and their union is contained in the       12  
13       support of the distribution of the r.v.s  $X_i$ ,  $i = 1, \dots, n$ . The  $i^{\text{th}}$  respondent       13  
14       states that the true point of compensating variation  $x_i$  belongs to an inter-       14  
15       val  $\mathbf{u}_h$ , e.g. " $x_i \in \mathbf{u}_h$ ",  $u_{Lh} < x_i \leq u_{Rh}$ . The rounded interval  $\mathbf{u}_h$  has covered       15  
16       the true point  $x_i$ .       16

17       Let  $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  be the division generated by the set of intervals       17  
18        $\mathcal{U}_m$ , i.e.  $\mathcal{V}_k$  is the collection of disjoint intervals  $\mathbf{v}_j = (v_{Lj}, v_{Rj}]$  and each       18  
19        $\mathbf{u}_h = \cup_{j \in \mathcal{C}_h} \mathbf{v}_j$ , where  $\mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}$  is the set of all indices of division       19  
20       intervals which are subsets of  $\mathbf{u}_h$ ,  $h = 1, \dots, m$ . For each  $j = 1, \dots, k$  we define       20  
21       the set  $\mathcal{D}_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}$ , i.e.  $h$  belongs to  $\mathcal{D}_j$  if and only if  $\mathbf{v}_j \subseteq \mathbf{u}_h$ . By       21  
22        $d_j$  we denote the number of  $h \in \mathcal{D}_j$ , i.e.  $d_j$  is the size of  $\mathcal{D}_j$ . The division       22  
23        $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  may be considered as a kind of bracketing generated by       23  
24       respondents due to roundings.       24

25       Let  $X_i = x_i \in \mathbf{v}_j$  and suppose that the  $i^{\text{th}}$ -respondent has stated an       25  
26       admissible interval  $\mathbf{u}_{h_i}$ ,  $h_i \in \mathcal{D}_j$ . The event  $\{X_i \in \mathbf{v}_j\}$  is not observed. We       26  
27       introduce the conditional probability  $w_{h_i j}$  to state  $\mathbf{u}_{h_i}$ ,  $h_i \in \mathcal{D}_j$        27  
28       28

$$(4.1) \quad w_{h_i j} = P["X_i \in \mathbf{u}_{h_i}" | X_i = x_i \in \mathbf{v}_j], \quad \sum_{h_i \in \mathcal{D}_j} w_{h_i j} = 1, \quad w_{h_i j} > 0. \quad 29$$

We use the following notations  $F[\mathbf{u}_h] = F[u_{Rh}] - F[u_{Lh}]$ ,  $F[\mathbf{v}_j] = F[v_{Rj}] - F[v_{Lj}]$  and note that  $F[\mathbf{u}_h] = \sum_{j \in \mathcal{C}_h} F[\mathbf{v}_j]$ .  $\mathbb{W} = \{w_{hj}, h \in \mathcal{D}_j, j = 1, \dots, k\}$  and  $\mathbf{F}_k = \{F[\mathbf{v}_j], j = 1, \dots, k\}$  are the parameters of the statistical model with the list of collected data  $\mathbf{dat}_m$ . From (2.2), Assumption 1 and the formula of total probability it follows that the  $i^{\text{th}}$  respondent states an interval  $\mathbf{u}_{h_i}$  with probability

$$(4.2) \quad w_{h_i} = P[{}^{\prime\prime}X_i \in \mathbf{u}_{h_i}{}^{\prime\prime}] = \sum_{j \in \mathcal{C}_{h_i}} w_{h_i j} F[\mathbf{v}_j], \quad i = 1, \dots, n.$$

Therefore, the probability to obtain a particular data  $\mathbf{dat}_m$  is

$$\prod_{i=1}^n P[{}^{\prime\prime}X_i \in \mathbf{u}_{h_i}{}^{\prime\prime}] = \prod_{h=1}^m w_h^{t_h} = \prod_{h=1}^m \left( \sum_{j \in \mathcal{C}_h} w_{hj} F[\mathbf{v}_j] \right)^{t_h}.$$

The right hand side is the likelihood if we consider it as a function of the parameters in  $\mathbb{W}$  and  $\mathbf{F}_k$ . We obtain

$$(4.3) \quad \text{likelihood}[\mathbb{W}, \mathbf{F}_k \mid \mathbf{dat}_m] = \prod_{h=1}^m \left( \sum_{j \in \mathcal{C}_h} w_{hj} F[\mathbf{v}_j] \right)^{t_h}.$$

It is more convenient consider the average log likelihood (llik). From (4.2) and (4.3) we have

$$(4.4) \quad \text{llik}[\mathbb{W}, \mathbf{F}_k \mid \mathbf{dat}_m] = \frac{1}{n} \sum_{h=1}^m t_h \text{Log}[w_h] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} w_{hj} F[\mathbf{v}_j] \right],$$

$n = \sum_{h=1}^m t_h$ . It is possible to rewrite (4.4) as follows

$$(4.5) \quad \text{llik}[\mathbb{W}, \mathbf{F}_k \mid \mathbf{dat}_m] = \text{llik} A[\mathbb{W}, \mathbf{F}_k \mid \mathbf{dat}_m] + \text{llik} B[\mathbf{F}_k \mid \mathbf{dat}_m],$$

where

$$(4.6) \quad \text{llik} A[\mathbb{W}, \mathbf{F}_k \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} w_{hj} F[\mathbf{v}_j] / F[\mathbf{u}_h] \right],$$

$$(4.7) \quad \text{llik} B[\mathbf{F}_k \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log}[F[\mathbf{u}_h]] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} F[\mathbf{v}_j] \right].$$

The loglikelihood (4.4) contains many unknown parameters:  $\mathbb{W} = \{w_{hj}, h \in \mathcal{D}_j, j = 1, \dots, k\}$  and  $\mathbf{F}_k = \{F[\mathbf{v}_1], \dots, F[\mathbf{v}_k]\}$ . The unknown probabilities in  $\mathbb{W}$  depend on the respondent's choice of interval from the set of admissible intervals  $\mathbf{u}_h, h \in \mathcal{D}_j$ , if  $x_i \in \mathbf{v}_j$ . In our approach it is possible to consider different conditional probabilities  $w_{hj} \in \mathbb{W}$ , depending on the behavior of respondents. For example, if the respondents are indifferent, i.e. the respondents state any  $\mathbf{u}_h, h \in \mathcal{D}_j$ , with the same probability, then  $w_{hj} = 1/d_j, j = 1, \dots, k$ . Note that we may also vary  $w_{hj}$  depending on the position of  $\mathbf{v}_j$  inside  $\mathbf{u}_h \supset \mathbf{v}_j$ . It is essential know whether respondents prefer to state  $\mathbf{u}_h \supset \mathbf{v}_j$  if  $\mathbf{v}_j \ni x_i$  is closer to the left(right)-hand end  $u_{Lh}$  ( $u_{Rh}$ ) or not.

If there is no truncation, then (4.7) formally corresponds to the likelihood suggested by [Turnbull \(1976\)](#) under the (very restrictive) assumption that the selection of intervals  $\mathbf{u}_h, h = 1, \dots, m$ , is independent of the respondents.

We can reduce the number of unknown parameters. Let us introduce a quasi-linear ordering in the set  $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  of all different stated intervals. We say that  $\mathbf{u}_{h_1}$  is more likely to be stated than  $\mathbf{u}_{h_2}$  if  $w_{h_1} > w_{h_2}$ . If  $w_{h_1} = w_{h_2}$  then we say that both  $w_{h_1}$  and  $w_{h_2}$  are equally likely to be stated.

Let us define the selection probabilities  $w_{hj}, h \in \mathcal{D}_j$ , by the relations

$$(4.8) \quad \tilde{w}_{hj} = \frac{w_h I[h \in \mathcal{D}_j]}{\sum_{h' \in \mathcal{D}_j} w_{h'}}, \quad j = 1, \dots, k.$$

The probabilities in (4.8) satisfy the quasi-linear ordering. The set of unknown parameters  $\mathbb{W}$  is thus reduced to the list of probabilities  $\mathbf{w}_m = \{w_1, \dots, w_m\}$  to state intervals  $\mathbf{u}_1, \dots, \mathbf{u}_m$ .

From (4.8) we thus obtain the average loglikelihood (4.4) as a function of unknown parameters  $\mathbf{w}_m = \{w_1, \dots, w_m\}$  and  $\mathbf{F}_k = \{F[\mathbf{v}_1], \dots, F[\mathbf{v}_k]\}$ . We

rewrite (4.4) as follows

$$(4.9) \quad \text{llik}[\mathbf{w}_m, \mathbf{F}_k \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} \tilde{w}_{hj} F[\mathbf{v}_j] \right]$$

where all  $\tilde{w}_{hj}$  are defined by relations (4.8). We consider  $\tilde{w}_{hj}$  as nuisance parameters because our aim is to estimate  $\mathbf{F}_k$ .

llik, in (4.9) implies that when respondents are selecting  $\mathbf{u}_h \in \mathcal{U}_m$ , they state the most "attractive" rounded interval  $u_h$  containing their WTP.

We consider two extensions of conditional probabilities  $w_{hj}, j \in \mathcal{C}_h$ , to select  $\mathbf{u}_h \in \mathcal{U}_m$ . Suppose that there are two independent causes affecting the probability  $w_{hj}$ : the anchoring of the rounded interval  $\mathbf{u}_h$  and the position of the interval  $\mathbf{v}_j \subseteq \mathbf{u}_h$  that contains WTP-point. Let  $\mathbf{u}_h = \mathbf{v}_{j_1} \cup \dots \cup \mathbf{v}_{j_{d_h}}, v_{Rj_s} = v_{Lj_{s+1}}$ . We say that  $\mathbf{v}_j$  has  $h$ -local rank  $r_{hj} = s$  if  $\mathbf{v}_j \subset \mathbf{u}_h$  and  $j = j_s$ . In the first extension we suppose that  $w_{hj}$  is a linear function of the  $h$ -local rank  $r_{hj}$  if the interval  $\mathbf{v}_{j_r}$  contains the WTP-point. Let  $\tilde{r}_{hj} = r_{hj}/r_h, r_h = \sum_{j \in \mathcal{C}_h} r_{hj}$ . Then we have

$$(4.10) \quad w_{hj}[c] = \frac{(1 + c\tilde{r}_{hj})w_h I[h \in D_j]}{\sum_{h' \in D_j} (1 + c\tilde{r}_{h'j})w_{h'}}, \quad h \in D_j, j = 1, \dots, k,$$

where  $c$  is a real number. If  $c > 0$  ( $c < 0$ ) then the probability to select  $\mathbf{u}_h$  increases (decreases) if the  $h$ -local rank of  $\mathbf{v}_{j_s}$  grows (descends). If  $c = 0$  then (4.10) coincides with (4.8). The related llik can be written as follows

$$(4.11) \quad \text{llik}[c, \mathbf{w}_m, \mathbf{F}_k \mid \mathbf{dat}_m] = \sum_{h=1}^m w_h \text{Log} \left[ \sum w_{hj}[c] F[\mathbf{v}_j] \right].$$

Another extension is based on the assumption that the second cause contributes to the factor  $g_{hj}$  proportional to the probabilities (4.8),  $\mathbf{v}_j \subset \mathbf{u}_h$ . Here, we use Beta distributions  $g_{hj} = B[\mathbf{v}_{hj}, \alpha, \beta] = \frac{1}{B[\alpha, \beta]} \int_{v_{Lhj}}^{v_{Rhj}} z^{\alpha-1} (1-z)^{\beta-1} dz, v_{Lhj} = (v_{Lj} - u_{Lh}) / (u_{Rh} - u_{Lh}), v_{Rhj} = (v_{Rj} - u_{Lh}) / (u_{Rh} - u_{Lh})$ . We have

$$(4.12) \quad w_{hj}[\alpha, \beta] = g_{hj} \frac{w_h I[h \in D_j]}{\sum_{h' \in D_j} w_{h'} g_{h'j}}.$$

Then the corresponding llik can be written as follows

$$(4.13) \quad \text{llik}[\alpha, \beta, \mathbf{w}_m, \mathbf{F}_k \mid \mathbf{dat}_m] = \sum_{h=1}^m w_h \text{Log} \left[ \sum_{j \in \mathcal{C}_h} w_{hj} [\alpha, \beta] F[\mathbf{v}_j] \right].$$

From Assumption 3 and (4.4) the ML-estimators of probabilities  $w_h$ ,  $h = 1, \dots, m$ , are  $\hat{w}_h = \frac{t_h}{n}$ ,  $\hat{w}_h - w_h \xrightarrow{P} 0$ , as  $n \rightarrow \infty$ . Here,  $\xrightarrow{P}$  denotes convergence in probability.

The estimators  $\hat{w}_h$ ,  $h = 1, \dots, m$ , can be used to obtain consistent estimators  $\hat{w}_{hj}$  for probabilities  $\tilde{w}_{hj}$ ,

$$(4.14) \quad \hat{w}_{hj} = \frac{t_h I[h \in \mathcal{D}_j]}{\sum_{h' \in \mathcal{D}_j} t_{h'}}, \quad j = 1, \dots, k.$$

Relations (4.14) can be also be used as consistent estimators of factors in the above extensions.

Our main aim is to estimate probabilities  $F[\mathbf{v}_1], \dots, F[\mathbf{v}_k]$  which we consider as  $k$  unknown parameters with the constraint  $\sum_{j=1}^k F[\mathbf{v}_j] \leq 1$ . All unknown parameters  $0 \leq \tilde{w}_{hj} \leq 1$ ,  $0 \leq F[\mathbf{v}_j] \leq 1$ ,  $j = 1, \dots, k$ ,  $h \in \mathcal{D}_j$  are contained in a compact set. Maximum likelihood can thus be applied to finding estimates of  $F[\mathbf{v}_j]$ ,  $j = 1, \dots, k$ , if we in (4.9) use the consistent estimates  $\hat{w}_{hj}$  of  $\tilde{w}_{hj}$ .

We collect an interesting property of our ML-estimators  $\hat{\mathbf{w}}_k$  in a proposition.

**PROPOSITION 1** *The ML-estimators  $\hat{\mathbf{w}}_m = \{\hat{w}_1, \dots, \hat{w}_m\}^T$  minimize the empirical entropy of the distribution  $\mathbf{w}_m = \{w_1, \dots, w_m\}^T$*

$$(4.15) \quad \max_{\mathbf{w}_m} \text{llik}[\mathbf{w}_m, \mathbf{F}_k \mid \mathbf{dat}_m] = - \sum_{h=1}^m \hat{w}_h \text{Log} \left[ \frac{1}{\hat{w}_h} \right].$$

Relation (4.15) follows from maximization of  $\sum_{h=1}^m w_h \text{Log}[w_h]$  with the linear constraint  $\sum_{h=1}^m w_h = 1$ .

We find estimates  $\hat{F}[\mathbf{v}_j]$  as the solutions to the following optimization problem

$$(4.16) \quad \max_{F[\mathbf{v}_j], j=1, \dots, k} \sum_{h=1}^m t_h \text{Log} \left[ \sum_{j \in \mathcal{C}_h} \hat{w}_{hj} F[\mathbf{v}_j] \right] = \sum_{h=1}^m t_h \text{Log} \left[ \sum_{j \in \mathcal{C}_h} \hat{w}_{hj} \hat{F}[\mathbf{v}_j] \right].$$

It is not possible to identify  $F[x]$  for all  $x$  but we can consistently estimate its increments on any division interval  $v_j$  as the number  $n$  of observations grows. The solution to the optimization problem in (4.16) is a non-trivial numerical problem because the number  $k$  of parameters  $F[\mathbf{v}_j]$  can be rather large. We do not try to solve it here, because we have a small data set. Instead, we suppose that the d.f. of interest  $F[x]$  can be approximated by a parametric function  $F_{\boldsymbol{\theta}}[x]$ ,  $\boldsymbol{\theta} \in \Theta$ . Here, we will use mixtures of the Weibull families.

## 5. APPLICATION OF THE STATISTICAL MODEL

The pilot analysis in Section 3 suggests that we can combine the interval data collected in the samples  $S_2$  and  $S_3$  into one sample. We denote this fused sample  $S_f$ . The corresponding  $\mathbf{dat}_m = \{\{\{u_{L1}, u_{R1}\}, t_1\}, \dots, \{\{u_{Lm}, u_{Rm}\}, t_m\}\}$  has size  $m = 46$ . The WTP-intervals are ordered by the values of their left-hands ends and right-hands ends as shown in Figure 1. The related division  $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  has  $k = 23$  intervals. The sample  $S_f$  contains  $n = 241$  stated intervals. The most popular interval  $\mathbf{u}_{27} = (50, 100]$  is, as noted, stated by  $t_{27} = 69$  respondents.

We denote the Weibull d.f.  $W(a, b) = F_{ab}[x]$  as  $1 - e^{-(x/a)^b}$ , where  $a$  is the scale parameter, and  $b$  is the shape parameter. We thus approximate the d.f. of unobserved true WTP-points values  $x_1, \dots, x_m$  by a distribution from the Weibull family  $\mathcal{F}_W$  of distributions. In addition, we use a mixture of the Weibull and the Exponential distributions. We denote the corresponding family by  $\mathcal{F}_{WE}$ . For comparative purposes, we include the Turnbull approach which is reduced here to parametric WTP-d.f. with no observed

WTP-points. The properties of the estimators in our model are collected in Theorem 5.1.

**THEOREM 5.1** *If Assumptions 1 - 3 hold and the d.f. of true WTP-points is  $W(a, b)$  then a consistent ML-estimator  $\hat{\boldsymbol{\theta}}_n = \{\hat{a}_n, \hat{b}_n\}$  exists and its accuracy can be consistently estimated by resampling as  $n \rightarrow \infty$ .*

A proof is outlined in the Appendix. A similar result also holds if the d.f. of true WTP-points is a mixture of the Weibull d.f. and the Exponential d.f.

Let us now turn to estimation results. We use (4.14) to estimate  $\hat{w}_{hj}$  the conditional probability to state an interval  $(u_{Lh}, u_{Rh}]$ . The loglikelihood is

$$(5.1) \quad \text{l lik } W[a, b \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} \hat{w}_{hj} \left( e^{-(v_{Lh}/a)^b} - e^{-(v_{Rh}/a)^b} \right) \right].$$

The contour plot of the loglikelihood (5.1) is shown in the left-hand part of Figure 4. By comparison, a parametric version of the Turnbull estimator entails maximizing

$$(5.2) \quad \text{l lik } WB[a, b \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \left( e^{-(u_{Lh}/a)^b} - e^{-(u_{Rh}/a)^b} \right) \right].$$

The difference between our proposed statistical model and the Turnbull approach can be seen from (5.1) and (5.2). The main difference is that (5.1) includes a sum over the divisions, while (5.2) has a much simpler probability statement. The expressions encapsulate the key difference between the way the data are generated. In (5.1) we present a way to cater for the fact that the individual can freely choose an interval, while (5.2) portrays the likelihood when the individual is presented with certain brackets by an investigator. The cost of this freedom from a computational point of view is displayed in (5.1). To repeat, we will use (5.1) and (5.2) on the same data. Thus, we plug in the data in (5.2) as if the individuals had been presented with the intervals actually stated by them.

The contour plot for the loglikelihood approximation (5.2) is shown in the right-hand part of Figure 4. Let  $\hat{a}, \hat{b}$  and  $\check{a}, \check{b}$  be the ML-estimators of  $a, b$  corresponding (5.1) and (5.2), respectively. The corresponding survival functions are shown in Figure 5. As can be seen,  $\hat{s}_W[x] = \text{Exp}[-(x/\hat{a})^{\hat{b}}]$  and  $\check{s}_W[x] = \text{Exp}[-(x/\check{a})^{\check{b}}]$  are nearly the same. We find estimates for the mean of the true WTP-points by integrating  $\hat{s}_W[x]$  and  $\check{s}_W[x]$  over  $[0, \infty)$ . Distribution consistently imitating deviations of ML-estimators from mean WTP have been obtained by resamplings with 2000 resampled copies. They are shown on the right side of Figure 5 in the normal Quantile-Quantile ( $Q-Q$ ) plot.  $X$ -axis in the Normal  $Q-Q$  plot contains values  $q$ -quantiles  $x(q)$  of the standard Normal distribution, i.e.  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x(q)} e^{-z^2/2} dz = q$ ,  $0 < q < 1$ . The distributions imitating deviations are shown as lines containing points  $\{x(q), y(q)\}$ . Here  $q$ -quantiles  $y(q)$  of these distributions are matched in pairs with  $q$ -quantiles  $x(q)$ ,  $0 < q < 1$ .

To approximate the s.f. of the unobserved true WTP-points by mixtures of the Weibull and the Exponential distributions, we define

$$(5.3) \quad s_{WE}[x, \boldsymbol{\theta}_4] = p \text{Exp} \left[ - \left( \frac{x}{a} \right)^b \right] + (1-p) \text{Exp} \left[ - \frac{x}{m_1} \right], \quad \boldsymbol{\theta}_4 = \{p, a, b, m_1\}.$$

The related loglikelihood function (reduced by plugging in  $\hat{w}_{hj}$  instead of  $w_{hj}$ ) is

$$(5.4) \quad \text{l lik } WE[\boldsymbol{\theta}_4 \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} \left[ \sum_{j \in \mathcal{C}_h} \hat{w}_{hj} (s_{WE}[v_{Lj}, \boldsymbol{\theta}_4] - s_{WE}[v_{Rj}, \boldsymbol{\theta}_4]) \right].$$

Similarly the reduced parametric Turnbull model entails maximizing

$$(5.5) \quad \text{l lik } WEB[\boldsymbol{\theta}_4 \mid \mathbf{dat}_m] = \sum_{h=1}^m \frac{t_h}{n} \text{Log} [s_{WE}[u_{Lh}, \boldsymbol{\theta}_4] - s_{WE}[u_{Rh}, \boldsymbol{\theta}_4]].$$

We collect parameter estimates and model statistics from the above models in Table II.



TABLE II  
 MAXIMUM LOGLIKELIHOOD ESTIMATES FOR THE WEIBULL AND THE MIXED WEIBULL  
 FAMILIES

Model	Our model		Turnbull	
	$\mathcal{F}_W$	$\mathcal{F}_{WE}$	$\mathcal{F}_W$	$\mathcal{F}_{WE}$
Family of d.f.s	(5.1)	(5.4)	(5.2)	(5.5)
Formula for llik	(5.1)	(5.4)	(5.2)	(5.5)
$\hat{p}$	-	0.851	-	0.892
$\hat{a}$	95.87	74.10	94.87	75.98
$\hat{b}$	1.116	1.764	1.178	1.875
$\hat{m}_1$	-	259.36	-	281.98
Maximum llik	-3.0377	-2.9301	-1.5480	-1.4106
Mean <i>WTP</i>	92.07	94.74	89.67	90.63
Left 0.025-limit	79.31	80.65	77.67	80.69
Right 0.975-limit	107.52	112.35	104.07	112.50

The maximum of (5.4) is -2.9301 which is much larger than -3.0377 for (5.1). The maximum of (5.4) is attained at  $\hat{p} = 0.851, \hat{a} = 74.096, \hat{b} = 1.764, \hat{m}_1 = 259.360$ . The estimated mean is  $\hat{m}_{WE} = 94.738$ . Note that if we approximate the s.f. of the true WTP-points distribution with mixtures of two Weibull distributions then the maximum of the related loglikelihood would be nearly the same as the maximum of (5.4). The two extensions of  $w_{hj}, j \in \mathcal{C}_h, h = 1, \dots, m$ , in (4.10) and (4.12) gives moderate but visible improvements the maxima of llik (4.11) and (4.13), respectively. The maximum of llik (4.11) and (4.12) as well as the corresponding values of parameters  $\{a, b, m_1, p, c\}$ , and  $\{a, b, m_1, p, \alpha, \beta\}$ , can be found by using the package of programs for analysis of self-selected interval data, Zhou, Belyaev, and Kriström (2010). The absolute values of the llik maximum are more close the empirical entropy stated in Proposition 1 which is 2.8216 for our empirical data  $\mathbf{dat}_m$ .

The empirical mean value for right-ends of rounded intervals is  $\hat{m}_{1R} = 146.826$ . The coefficient  $c_f = \hat{m}_{WE}/\hat{m}_{1R} = 0.645$  can be used for rescaling.

The rescaled s.f.  $s_{WE}[c_f x, \hat{\theta}_4]$  can be used for approximating the s.f. of right-ends of rounded intervals in accordance with the analysis in Section 3. The rescaled s.f. is shown in Figure 6. This approximation is better than the approximation based on single Weibull s.f.s.

## 6. CONCLUDING REMARKS

We have analyzed an elicitation approach in social surveys in which a respondent can select an interval of choice. Because a point is a special case of an interval, and any suggested bracket is a special case of a self-selected interval, we believe the approach has merit.

Our statistical model is based on the idea that there are two probabilities involved, one being related to the choice of interval, the other to the conditional probability that the unknown value of interest belongs to a given interval. Consequently, our model can be considered as an extension of [Turnbull \(1976\)](#). Given the paucity of our data, we did not attempt to solve for the non-parametric maximum likelihood estimator. There are other interesting statistical problems to be resolved, including, but not limited to the asymptotic properties of the non-parametric estimator. Furthermore, it would be possible to include explanatory variables in our parametric model and this is a natural next step. The mixed Weibull results did suggest a certain clustering of the data, a property that might go away when we add explanatory variables.

The key part of the suggested model contains the division intervals generated by the stated intervals. The division intervals can be understood as an analog to bracketing and the conditional probabilities to state intervals given the division intervals can be used to study respondent behavior. This suggests that an interesting point of departure for future research lies in the connection to psychology. We believe that this model could be a fruitful basis for further joint work between economists and psychologists.

## 7. APPENDIX

The collected statistical data is the sequence  $\mathfrak{u}_n = \{\mathbf{u}_{h_1}, \dots, \mathbf{u}_{h_n}\}$  where  $\mathbf{u}_{h_i}$  is the interval stated by the  $i^{\text{th}}$  respondent. Suppose that Assumptions 1 - 5 hold and we can consistently estimate  $w_h$  and  $\tilde{w}_{h_j}$  as  $n \rightarrow \infty$ . To simplify this presentation we assume that all  $w_h$  and  $\tilde{w}_{h_j}$  are known exactly. Then only the true distribution  $F_0[\cdot]$  of WTP-points of compensating variations has to be estimated. We consider the parametric case with  $F_0[\cdot] = F_{\boldsymbol{\theta}_0}[\cdot]$  belonging to a parametric family  $\mathcal{F}_\Theta = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^s\}$ ,  $\mathbb{R}^s$  is the Euclidian  $s$ -dimensional space with the usual metric  $\|\cdot\|_2$ ,  $\boldsymbol{\theta}_0$  is the true parameter. The loglikelihood function (4.4) can be rewritten as follows

$$(7.1) \quad \text{llik}[\boldsymbol{\theta} \mid \mathfrak{u}_n] = \sum_{i=1}^n l_i[\boldsymbol{\theta}],$$

$$(7.2) \quad l_i[\boldsymbol{\theta}] = I[X_i \in \mathbf{u}_{h_i}] \text{Log} \left[ \sum_{j \in \mathcal{C}_{h_i}} \tilde{w}_{h_j} F_{\boldsymbol{\theta}}[\mathbf{v}_j] \right].$$

Recall that  $\mathbf{u}_1, \dots, \mathbf{u}_m$  are different stated intervals and  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are the corresponding division intervals.

We apply the Maximum Likelihood principle in two steps. In the first step we find the consistent ML-estimator. The related theory is rather well established under some assumptions on  $l_i[\boldsymbol{\theta}]$ , see e.g. [Ferguson \(1996\)](#), [Lehmann and Casella \(1998\)](#). In order to obtain existence of consistent ML-estimators one must check certain properties of the loglikelihood ratio

$$(7.3) \quad R_i[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] = l_i[\boldsymbol{\theta}_1] - l_i[\boldsymbol{\theta}_2],$$

and its infimum in the ball  $\mathcal{B}_\rho(\boldsymbol{\theta}_2) = \{\boldsymbol{\theta}' : \|\boldsymbol{\theta}' - \boldsymbol{\theta}_2\|_2 \leq \rho\}$

$$(7.4) \quad R_i[\boldsymbol{\theta}_1, \mathcal{B}_\rho(\boldsymbol{\theta}_2)] = \inf_{\boldsymbol{\theta}' \in \mathcal{B}_\rho(\boldsymbol{\theta}_2)} R_i[\boldsymbol{\theta}_1, \boldsymbol{\theta}'].$$

The  $l_i(\boldsymbol{\theta})$ ,  $i = 1, \dots, n$  are continuous, locally bounded, and the expectations exist and are positive

$$(7.5) \quad E_{\boldsymbol{\theta}_1}[R_i[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]] > 0, \quad E_{\boldsymbol{\theta}_1}[R_i[\boldsymbol{\theta}_1, \mathcal{B}_\rho(\boldsymbol{\theta}_2)]] > \gamma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0$$

for any  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$  and sufficiently small  $\rho$  and  $\gamma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0$ . Besides that one need to check for a  $g > 2$  that the following expectations are finite

$$(7.6) \quad E_{\boldsymbol{\theta}_1}[|R_i[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]|^g] < \infty, \quad E_{\boldsymbol{\theta}_2}[|R_i[\boldsymbol{\theta}_1, \mathcal{B}_\rho(\boldsymbol{\theta}_2)]|^g] < \infty.$$

It is sufficient to find a compact set  $\mathcal{K}_0 \subset \Theta$  such that  $\boldsymbol{\theta}_0 \in \mathcal{K}_0 \setminus \partial\mathcal{K}_0$ , i.e.  $\boldsymbol{\theta}_0$  is not on the boundary  $\partial\mathcal{K}_0$  of  $\mathcal{K}_0$ . If  $l_i(\boldsymbol{\theta}), i = 1, \dots, n$  are continuous and (7.5), (7.6) hold then the ML-estimators  $\hat{\boldsymbol{\theta}}_n$  exist and are consistent.

Suppose that  $F_{\boldsymbol{\theta}_0}[\cdot]$  belongs to the Weibull family of distributions  $\mathcal{F}_W = \{1 - e^{-(x/a)^b} : \boldsymbol{\theta} = \{a, b\}, a > 0, b > 0\}$ . Let  $\boldsymbol{\theta}_r = \{a_r, b_r\}$  and  $F_{\boldsymbol{\theta}_r}[\mathbf{v}_j] = e^{-(v_{L_j}/a_r)^{b_r}} - e^{-(v_{R_j}/a_r)^{b_r}}$ , then

$$(7.7) \quad l_i[\boldsymbol{\theta}_r] = I[X_i \in \mathbf{u}_{h_i}] \text{Log} \left[ \sum_{j \in \mathcal{C}_{h_i}} \tilde{w}_{hj} F_{\boldsymbol{\theta}_r}[\mathbf{v}_j] \right],$$

and

$$(7.8) \quad R_i[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] = I[X_i = \mathbf{u}_{h_i}] \left( \text{Log} \left[ \sum_{j \in \mathcal{C}_{h_i}} \tilde{w}_{hj} F_{\boldsymbol{\theta}_1}[\mathbf{v}_j] \right] - \text{Log} \left[ \sum_{j \in \mathcal{C}_{h_i}} \tilde{w}_{hj} F_{\boldsymbol{\theta}_2}[\mathbf{v}_j] \right] \right).$$

We can check that inside any rectangular  $\mathcal{K} = [0, a_+] \times [0, b_+]$ ,  $a_+ > 0, b_+ > 0$ ,  $l_i(\boldsymbol{\theta})$  in (7.7) are continuous and inequalities (7.5) and (7.6) hold. If  $\boldsymbol{\theta}_0 = \{a_0, b_0\}$  are parameters of the true Weibull distribution of WTP-points then for each  $\mathbf{u}_h$  stated with positive probabilities  $w_h > 0$  the following inequalities have to be hold

$$(7.9) \quad 0 < w_h = \sum_{j \in \mathcal{C}_h} \tilde{w}_{hj} F_{\boldsymbol{\theta}_0}[\mathbf{v}_j] \leq F_{\boldsymbol{\theta}_0}[\mathbf{u}_h],$$

for all  $h = 1, \dots, m$ . The analysis of asymptotic behavior of  $1 - e^{-(x/a)^b}$  as  $a \rightarrow \infty$  ( $a \rightarrow 0$ ),  $b > 0$  and  $b \rightarrow \infty$  ( $b \rightarrow 0$ ),  $a > 0$  shows that the corresponding Weibull distribution either concentrates at a single point or goes to 0 or to  $+\infty$ . Hence, there are sufficiently large  $a_+ > 0, b_+ > 0$ , such that outside the rectangle  $\mathcal{K}_0 = [0, a_+] \times [0, b_+]$  at least one of inequalities (7.9) does not hold and, therefore,  $\boldsymbol{\theta}_0 \in \mathcal{K}_0$ . Then from inequalities (7.5)

and (7.6) it follows that there exists a unique consistent ML-estimator  $\hat{\boldsymbol{\theta}}_n \in \mathcal{K}$ . Similar arguments can be used in the case of mixtures the Weibull and the Exponential distributions. It is possible to check that for family  $\mathcal{F}_{WE}$  of mixtures the Weibull and the Exponential distribution inequalities (7.5) and (7.6) also hold and there exists the consistent ML-estimator  $\hat{\boldsymbol{\theta}}_n$  of the true parameter  $\boldsymbol{\theta}_0$  of distribution in  $\mathcal{F}_{WE}$ .

In the second step we will consistently estimate accuracy of the ML-estimator  $\hat{\boldsymbol{\theta}}_n$ . If we suppose that a consistent ML-estimator exists and regularity assumptions hold, see Lehmann and Casella (1998), e.g. the first and the second order partial derivatives of  $l_i(\boldsymbol{\theta})$ , exist, and the Fisher matrix has full rank inside the compact set  $\mathcal{K}_0$ , then the resampling methods consistently evaluate accuracies of the ML-estimators. These regularity assumptions hold for  $\mathcal{F}_W$  and  $\mathcal{F}_{WE}$ . We will use resampling copies of the data  $\mathbf{u}_n^* = \{\mathbf{u}_{h_1^*}, \dots, \mathbf{u}_{h_n^*}\}$  where  $\{h_1^*, \dots, h_n^*\}$  are numbers independently and randomly sampled from the list  $\{1, 2, \dots, n\}$ . The corresponding loglikelihood function is

$$(7.10) \quad \text{llik}[\boldsymbol{\theta} \mid \mathbf{u}_n^*] = \sum_{i=1}^n l_i^*[\boldsymbol{\theta}],$$

$$l_i^*[\boldsymbol{\theta}] = I[X_i \in u_{h_i^*}] \text{Log} \left[ \sum_{j \in \mathcal{C}_{h_i^*}} \tilde{w}_{h_i^*, j} F_{\boldsymbol{\theta}}[\mathbf{v}_j] \right].$$

For each copy we find the corresponding ML\*-estimator. We need to generate a rather large number  $R$  of such copies of data. Let  $\hat{\boldsymbol{\theta}}_n^{*c}$  be ML-estimator based on the  $c^{\text{th}}$  resampled copy of data  $\mathbf{u}_n^{*c}$ . Then the empirical distribution of differences  $\hat{\boldsymbol{\theta}}_n^{*c} - \hat{\boldsymbol{\theta}}_n$ ,  $c = 1, \dots, R$ , will imitate the distribution of deviation  $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ . The proof of this fact based on the Central Limit Resampling Theorem, Belyaev (2003, 2007), Belyaev and Sjöstedt-de Luna (2000). A detailed theory for consistent estimation deviations of the ML-estimators of parameters based on resamplings is given in Nilsson (1998).

## REFERENCES

- 1  
2  
3 ARIELY, D., LOEWENSTEIN G. AND PRELEC, D. (2003): Coherent Arbitrariness: Stable  
4 Demand Curves Without Stable Preferences *Quarterly Journal of Economics*, **118**,  
5 73–105.  
6  
7 AKTER, S., BENNETT, J. AND AKHTER, S. (2008): Preference uncertainty in contingent  
8 valuation *Ecological Economics*, **67**, 345–351.  
9  
10 BELYAEV, Y. (2003): Necessary and sufficient conditions for consistency of resampling.  
11 Research report, Center of Biostochastics. Swedish University of Agricultural Sciences.  
12 <http://biostochastics.slu.se/publikationer>.  
13  
14 BELYAEV, Y. (2007): Resampling for Lifetime Distributions. 1653–1657 In *Encyclopedia*  
15 *of Statistics in Quality and Reliability*, John Wiley, Chichester, UK.  
16  
17 BELYAEV, Y., HÅKANSSON, C., AND KRISTRÖM, B. (2008): Rounding it up- Interval  
18 and open ended questions. Department of Forest Economics, Swedish University of  
19 Agricultural Sciences, Umeå, Working paper No 373.  
20  
21 BELYAEV, Y. AND SJÖSTEDT-DE LUNA, S. (2000): Weakly approaching sequences of  
22 random distributions. *J. Applied Probability*, **37**, 807–822.  
23  
24 BROBERG, T. AND BRÄNNLUND, R. (2008): An alternative interpretation of multiple  
25 bounded WTP data-Certainty dependent payment card intervals. *Resources Energy*  
26 *Econ*, doi:10.1016/j.reseneeco.2008.09.001.  
27  
28 FERGUSON, T.S. (1996): *A Course in Large Sample Theory*. Chapman & Hall, London.  
29  
30 FINK, A. (1985): *How to ask survey questions*. Thousand Oaks, CA: SAGE Publications.  
31 Wiley, New York.  
32  
33 GARTHWAITE, P.H., KADANE, J.B. AND A O'HAGAN (2004): "Elicitation" Tech-  
34 nical Report 04/01, 2004, Open University. [http://statistics.open.ac.uk/  
35 TechnicalReports/Handbook-Elicitation.pdf](http://statistics.open.ac.uk/TechnicalReports/Handbook-Elicitation.pdf)  
36  
37 HANEMANN, W.M., KRISTRÖM, B. AND LI, C.-Z. (1996): Nonmarket Valuation un-  
38 der Preference Uncertainty: Econometric Models and Estimation" Department of  
39 Agricultural & Resource Economics, UCB. CUDARE Working Paper 794. [http:  
40 //repositories.cdlib.org/are\\_ucb/794](http://repositories.cdlib.org/are_ucb/794)  
41  
42 HANLEY, N., KRISTRÖM B. AND SHOGREN, J.F. (2008): Coherent arbitrariness: On  
43 value uncertainty for environmental goods, *Land Economics*, **85(1)**: 41–45.  
44  
45 HOBBS, F. (2004): Age and Sex Composition, in Siegel, B. and Swansson,  
46 D. (eds), *The methods and Materials of demography*, Elsevier Academic Press, San  
47 Diego, CA, USA.

- 1 HURD, M.D., MCFADDEN, D., CHAND, H., GAN, L., MERRILL, A. AND ROBERTS, M. 1  
 2 (1998): Consumption and savings balances of the elderly: Experimental evidence on 2  
 3 survey response bias, in D. Wise (Ed.) *Frontiers in the Economics of Aging*, 353–387, 3  
 4 Chicago, IL: University of Chicago Press. 4
- 5 HUTTENLOCHER, J., HEDGES, L.V. AND BRADBURN, N.M. (2008): Reports of elapsed 5  
 6 time: Bounding and rounding processes in estimation. *Journal of Experimental Psy-* 6  
 7 *chology: Learning, Memory, and Cognition* **16**, 196–213. 7
- 8 HÅKANSSON, C. (2008): A new valuation question - Analysis of and insights from interval 8  
 9 open ended data in contingent valuation. *Environmental and Resource Economics* 9  
 10 **39**,2,175–188. 10
- 11 JAMMALAMADAKA, S.R. AND MANGALAM, V. (2003): Non-parametric estimation for 11  
 12 middle-censored data. *J. Nonparametr. Statist.* **15**, 253–265. 12
- 13 KLEIN, J.P. AND MOESCHBERGER, ML (1997): *Survival Analysis: Techniques for Cen-* 13  
 14 *sored and Truncated Data*, Springer-Verlag. 14
- 15 KYNN, M. (2007) 'The 'heuristics and biases' bias in expert elicitation *Journal of the* 15  
 16 *Royal Statistical Society: Series A (Statistics in Society)*. 171(1), 239-264. 16
- 17 LEHMANN, E.L. AND CASELLA, G. (1998): *Theory of point estimation*. 2nd. ed., N. Y.: 17  
 18 Springer-Verlag. 18
- 19 MANSKI, C.F. AND F.MOLINARI (2010) Rounding Probabilistic Expectations in Surveys. 19  
 20 *Journal of Business and Economic Statistics*. **28**, 219-231. 20
- 21 MCFADDEN, D. (1998): Measuring Willingness-to-Pay for Transportation Improve- 21  
 22 ments", in Gärling, T., Laitila, T. and Westin, K. (eds.), *Theoretical Foundations* 22  
 23 *of Travel Choice Modeling*, 339-364, Elsevier Science: Amsterdam. 23
- 24 MCFADDEN, D., BEMMAOR, A., CARO, F., DOMINITZ, J., JUN, B-H., LEWBEL, A., 24  
 25 MATZKIN, R., MOLINARI, F., SCHWARZ, N., WILLIS, R. AND WINTER, J. (2005): 25  
 26 Statistical Analysis of Choice Experiments and Surveys, *Marketing Letters*, **16**, 183– 26  
 27 196. 27
- 28 NILSSON, L. (1998): Parametric Estimation Using Computer Intensive Methods. Ph. 28  
 29 Dissertation, ISBN 91-7191-438-8, Dept. of Mathematical Statistics, Umeå University. 29
- ROSCH, E. (1975): Cognitive reference points. *Cognitive Psychology* **7**, 532–547.
- ROSSI, P.H., WRIGHT, J.D. AND ANDERSON, A.B. (EDS.) (1978): *Handbook of Survey Research*, New York: Academic Press, 1983.
- RUBENSTEIN, S. M. (1994): *Surveying Public Opinion*, Belmont, CA: Wadsworth Publishers.
- SCHAEFFER, N.C. AND BRADBURN, N.M. (1989): Respondent behaviour in magnitude

- 1 estimation. *Journal of the American Statistical Association* **84**, 402–413. 1
- 2 SCHWARTZ, N. (1999): Self Reports: How the uestions shape the answers. *American* 2
- 3 *Psychologist* **54**, 93-105. 3
- 4 SLOVIC, P (1972): FROM SHAKESPEARE TO SIMON: SPECULATIONS - AND SOME EVI- 4
- 5 DENCE - ABOUTMAN’S ABILITY TO PROCESS INFORMATION. *Oregon Research Institute* 5
- 6 *Monograph*, 12, 2. 6
- 7 TURNBULL, B.W. (1976): The empirical distribution function with arbitrarily grouped, 7
- 8 censored and truncated data. *J. Roy. Statist. Soc. Ser. B* 38, pp. 290–295. 7
- 9 TVERSKY, A. AND D. KAHNEMANN (1974): Judgment under uncertainty: Heuristics and 8
- 10 biases. *Science*. **185**, 1124-1130. 8
- 11 VAN EXEL, W.B.F., BROUWER, B., VAN DEN BERG, B. AND KOOPMANSCHAP, M.A. 9
- 12 (2006): With a little help from an anchor discussion and evidence of anchoring effects 10
- 13 in contingent valuation, *The Journal of Socio-Economics*, **35**, 836–853. 11
- 14 WINTER, J. (2003): Bracketing effects in categorized survey questions and the measure- 12
- 15 ment of economic quantities. Working Paper, Mannheim Research Institute for the 13
- 16 Economics of Aging (MEA), February 2003. 13
- 17 WINTER, J. (2004): Response bias in survey-based measures of household consumption. 14
- 18 *Economics Bulletin* **3**, 9, 1–12. 15
- 19 ZHOU, W., BELYAEV, Y., AND KRISTRÖM, B. (2010): iwtp: Software for Analysis of 16
- 20 Self-Selected Interval Data. Department of Forest Economics, Swedish University of 17
- 21 Agricultural Sciences, Umeå. (In preparation). 18
- 19 19
- 20 20
- 21 21
- 22 22
- 23 23
- 24 24
- 25 25
- 26 26
- 27 27
- 28 28
- 29 29



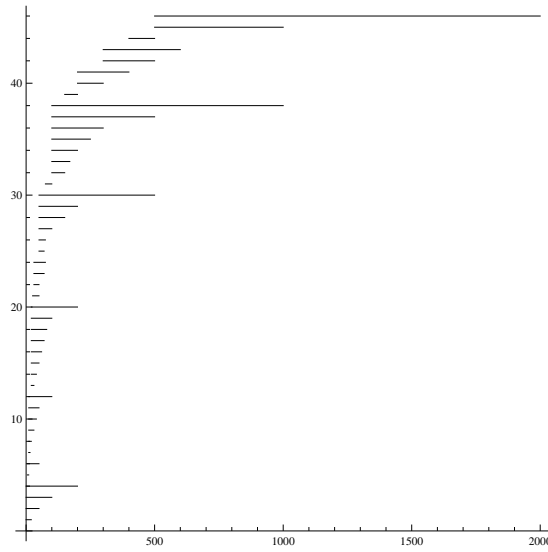


FIGURE 1.— The intervals in the samples  $S_2$  and  $S_3$  are shown as parallel to  $X$ -axis given in SEK (Swedish kronas). They are ordered by their left-ends and if the left ends are the same then they are ordered by their lengths. Their ordering numbers are shown in  $Y$ -axis.

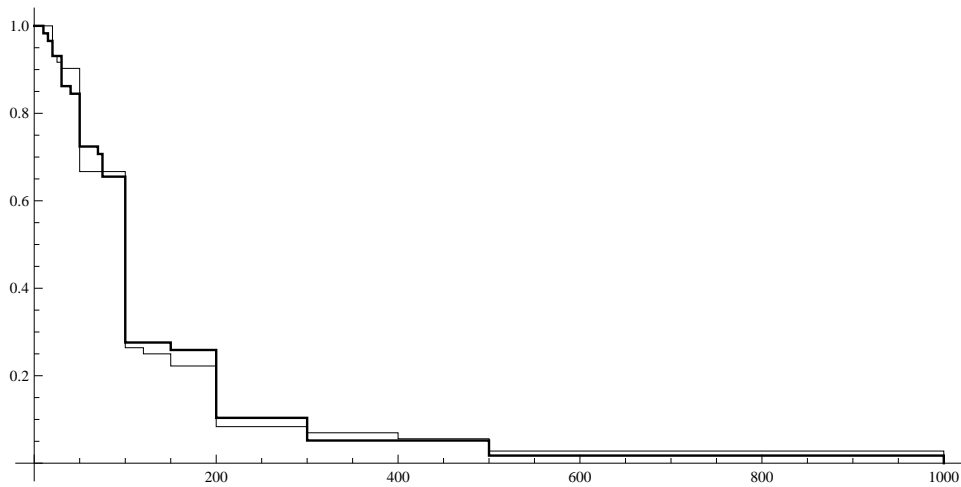


FIGURE 2.— Two empirical s.f.s corresponding to the rounded WTP-points in the sample  $S_1$  (thin stepwise line) and to the right ends of rounded WTP-intervals in the sample  $S_2$  (thick stepwise line).

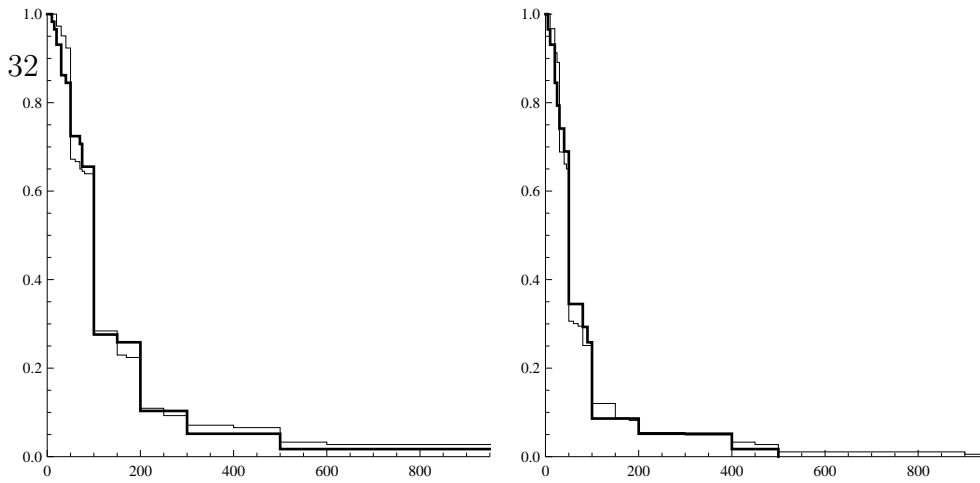


FIGURE 3.— Two empirical s.f.s of the stated right ends intervals in the samples  $S_2$  (thick line) and  $S_3$  (thin line) are shown in the left plot. Two empirical s.f.s of the lengths in the stated intervals in the samples  $S_2$  (thick line) and  $S_3$  (thin line) are shown in the right plot. Both plots are trimmed at 950 SEK.

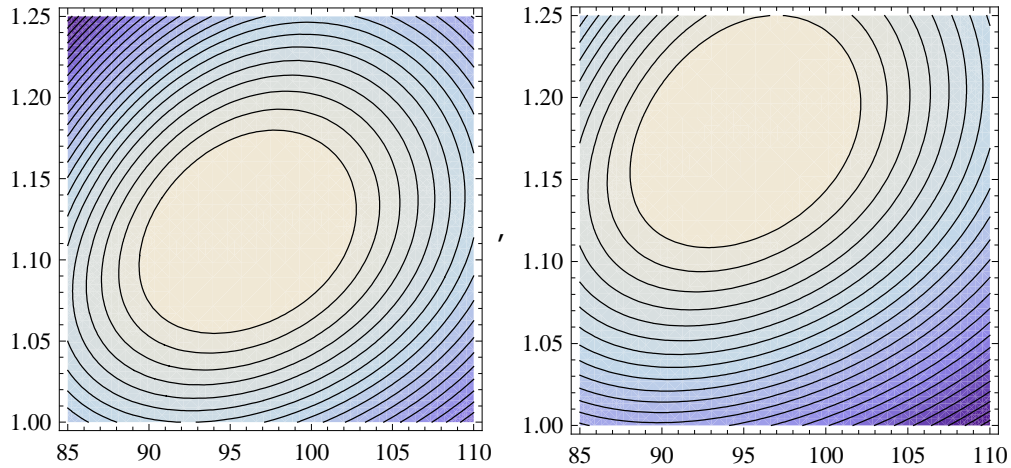


FIGURE 4.— Two contour plots: left for loglikelihood (5.1), right for loglikelihood part (5.2).  $X$ -axis shows values of scale  $a$  and  $Y$ -axis shows values of shape  $b$  parameters of the Weibull distribution  $W(a, b)$ .

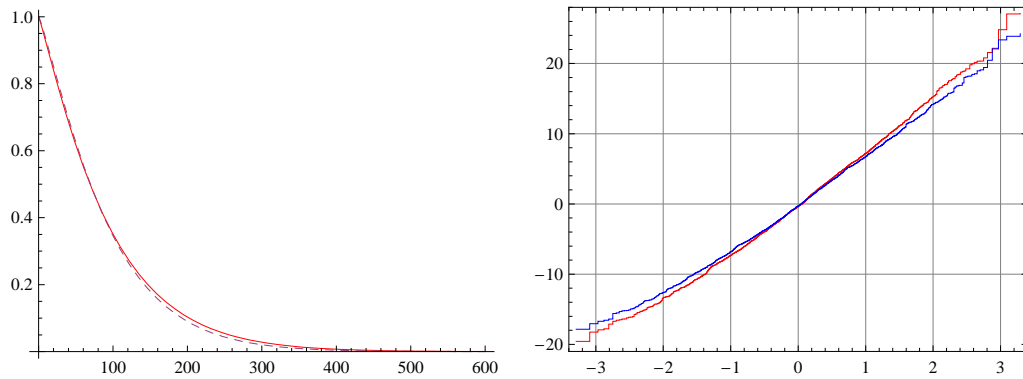


FIGURE 5.— Two Weibull s.f.s, with parameters  $\hat{a}, \hat{b}$  maximizing the loglikelihood (5.1) (smooth line) and the loglikelihood part (5.2) (dashed line), are shown in the left side. Normal  $Q - Q$  plot on the right side shows two distributions imitating deviations of ML-estimators from mean WTP corresponding (5.1) and (5.2).

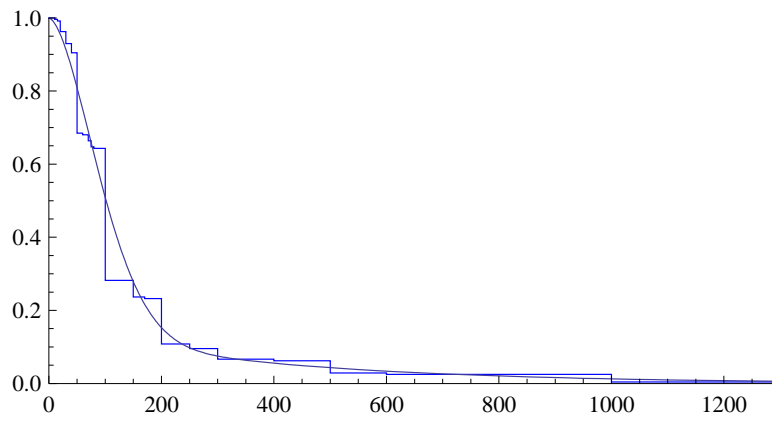


FIGURE 6.— Two survival functions are shown: the empirical s.f. (stepwise line) of the right ends of declared intervals in the fused sample  $S_f$  and the mixed Weibull s.f. (smooth line)  $s_{WE}[c_f x, \hat{\theta}_4]$ , where  $\hat{\theta}_4 = \{\hat{p}, \hat{a}, \hat{b}, \hat{m}_1\}$  is the ML-estimate of the true parameter  $\theta_{04} = \{p_0, a_0, b_0, m_{10}\}$ . The coefficient  $c_f = 0.6452$  is the ratio of the mean of the mixture and the nonparametric estimate of the mean value of the s.f. based on the right ends of intervals in the sample  $S_f$ .