



Two-step approach to Self-Selected Interval Data in Elicitation Surveys

Yuri Belyaev¹ and Bengt Kriström^{1,2}

¹Department of Forest Economics

SLU

S-901 83 Umeå, Sweden

²CERE, Centre for Environmental and Resource Economics

SLU and Umeå University

S-901 83 Umeå, Sweden

The **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 are to tie together research groups at the different departments and universities; provide seminars and workshops within the field of environmental & resource economics and management; and constitute a platform for a creative and strong research environment within the field.



Two-step approach to Self-Selected Interval Data in Elicitation Surveys

Yuri Belyaev and Bengt Kriström

May 30, 2012

1 TWO-STEP APPROACH TO SELF-SELECTED INTERVAL DATA IN 1
2 ELICITATION SURVEYS¹ 2
3

4 YURI BELYAEV^a AND BENGT KRISTRÖM^b 4

5 We propose a novel two-step approach to elicitation in surveys and provide 5
6 supporting statistical theory for the models suggested. The essential idea is to 6
7 combine self-selected intervals in a first step and then employ brackets generated 7
8 from the intervals in a second step. In this way we combine the advantages of self- 8
9 selected intervals, mainly related to the fact that individuals often find it difficult 9
10 to report a precise point-estimate of a quantity of interest, with the documented 10
11 usefulness of brackets. Because the brackets are generated from the first sample 11
12 we sidestep the thorny problem of the optimal design of brackets and additional 12
13 assumptions on dependency between the self-selected intervals and their points of 13
14 interest. Our set-up necessitates development of new statistical models. First, we 14
15 propose a stopping rule for sampling in the first step. Second, Theorem 1 proves 15
16 that the proposed non-parametric ML-estimator of the underlying distribution 16
17 function is consistent. Third, a special recursion for quick estimation of the ML- 17
18

19 ¹Department of Forest Economics, 901 83 Umeå, Sweden. The research presented in 19
20 this paper was carried out as a part of the R&D programme "Hydropower - Environmen- 20
21 tal impacts, mitigation measures and costs in regulated waters". It has been established 21
22 and financed by Elforsk, the Swedish Energy Agency, the National Board of Fisheries 22
23 and the Swedish Environmental Protection Agency. Partial funding from the "PlusMinus 23
24 Program" (<http://www.plusminus.slu.se>) is also gratefully acknowledged. We thank the 24
25 participants at the 15th Ulvön Conference on Environmental Economics, Ulvön, Sweden, 25
26 June 17-19, 2008, 16th Annual Conference of the European Association for Environmen- 26
27 tal and Resource Economics, Göteborg, Sweden, June 25-28, 2008, IISA Conference Jan 27
28 4-8, 2010, Visakhapatnam, (Andhra U., Waltair), India, and seminar participants at the 28
29 Department of Mathematics & Mathematical Statistics, University of Umeå and Depart- 29
ment 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. The usual disclaimer applies.

^aDepartment of Forest Economics SLU, 901 83 Umeå, Sweden.
yuri.belyaev@sekon.slu.se; <http://www.sekon.slu.se>

^bCentre for Environmental & Resource Economics (CERE), 901 83 Umeå, Sweden.
Sweden. bengt.kristrom@sekon.slu.se; <http://www.cere.se>

estimators is suggested. Theorem 2 shows that the accuracy of the estimator can be consistently estimated by resampling. Fourth, we have developed an R-package for efficient application of the method. We illustrate the approach using the problem of eliciting willingness-to-pay for a public good.

KEYWORDS: Interval data, Maximum Likelihood, Turnbull estimator, willingness-to-pay, quantitative elicitation, resampling..

1. INTRODUCTION

When asking quantitative survey-questions, there is an intuitively appealing advantage in letting the respondent choose an interval, rather than forcing him or her to state a point (or to choose one of a number of presented brackets). A point (or a given bracket) is, for all practical purposes, a special case of a self-selected interval. Several empirically observed phenomena provides additional motivation. One is that the respondent might not be completely certain about the answer. When asked about, say, the distance to the next city, we typically use a convenient round number or present an interval ("it is about 15-20 miles from here"). It seems natural that the elicitation mechanism should allow for such behavior. The individual might also refuse to answer the question because of uncertainty, see [Belyaev and Kriström \(2010, 2012\)](#), [Håkansson \(2007, 2008\)](#) and [Manski and Molinari \(2010\)](#) for evidence supporting this claim. Furthermore, because uncertainty may vary in complex ways across individuals, a freely chosen interval intuitively provides a better way of characterizing the underlying uncertainty, compared to brackets or open-ended questions.

Perhaps surprisingly, self-selected intervals are uncommon in the survey literature. One reason could be that the handling of such interval data is not part of the standard statistical literature. We propose to close some of this gap, by proposing a statistical toolbox for such interval-questions, complete with supporting statistical theory and publicly available software (written in R) [Wenchao, Belyaev and Kriström \(2012\)](#).

1 The paper contains several new ideas and results. In survival analysis it 1
 2 is typically assumed that the censoring intervals are independent of the un- 2
 3 observable points of interest and that the intervals cover only some of the 3
 4 points, see [Jammalamadaka and Mangalam \(2003\)](#), [Klein and Moeschberger](#) 4
 5 [\(1997\)](#), [Turnbull \(1976\)](#). Our model does not put any restriction on the 5
 6 dependence between self-selected intervals and their associated (but unob- 6
 7 servable) points; the model can be viewed as a generalization of the non- 7
 8 parametric ML-estimator for general censoring proposed by [Turnbull \(1976\)](#). 8
 9 However, the use of a single self-selected interval turns out to lead to a funda- 9
 10 mental identifiability issue. To overcome this problem, we introduce a new 10
 11 two-step approach. We thus employ a second question in second sample, 11
 12 which we use to fine-grain the information from the first step. The second 12
 13 step depends on the first in a way that has not been explored before in the 13
 14 survey literature: the second step brackets are generated by self-selected 14
 15 intervals collected on the first step. We derive a sampling stopping rule 15
 16 which enables the analyst to formally address the problem of sample size 16
 17 in the first step. We couch our model within a contingent valuation experi- 17
 18 ment, in which the individual is asked about his willingness-to-pay (WTP) 18
 19 for a public good. This is a natural application of our ideas, because such 19
 20 questions typically involve unfamiliar goods (such as improvements in en- 20
 21 vironmental quality) and a choice situation that individuals are typically 21
 22 not familiar with. While this is a natural application our approach has wide 22
 23 applicability. 23

24 The paper is structured as follows. In section 2 we describe basic as- 24
 25 sumptions and present the two-step sampling plan. In Section 3 we analyze 25
 26 intervals, stated by randomly sampled respondents, by using multinomial 26
 27 processes. Section 4 details the statistical model and the corresponding log- 27
 28 likelihood. Theorem 4.1 collects the properties of the proposed estimators 28
 29 and Theorem 4.2 gives properties of their accuracy. Section 5 provides a 29

simulation experiment. Section 6 concludes. Proofs are in an Appendix.

2. BASIC ASSUMPTIONS

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.¹ The choice is essentially between two types of survey questions, the open-ended and the closed-ended.² Rather than eliciting a point estimate or select between given brackets, we first ask the individual to choose any interval of choice containing the point of interest.

Potential applications include, but are not limited to: recall situations ("How many days were you unemployed the first quarter of last year?", "What was your net income the previous year of taxation?"), projections ("What is your best forecast of the next year's interest rate?") or contingent valuation studies ("How much are you maximally willing to pay for the

¹Perhaps 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. \(2005\)](#).

²Each 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 \(1985\)](#).

suggested change?”). As [Manski and Molinari \(2010\)](#) points out intervals are more common in daily communication than we ordinarily think. Thus, weather reports and pilot communications include a form of implicit interval, e.g. when meteorologists report that the wind blows from the ”north means that the wind direction lies in the interval [337.5., 22.5].” This kind of rounding is also prevalent in many types of surveys; it is well-documented that individuals often round their answers to open-ended survey questions, see [Rosch \(1975\)](#), [Schaeffer and Bradburn \(1989\)](#), [Huttenlocher, Hedges and Bradburn \(2008\)](#), [Hurd et al. \(1998\)](#), [Hobbs \(2004\)](#), and [van Exel et al. \(2006\)](#).

The basic problem is that when a respondent is to report a point, he sometimes round it to describe a sentiment that really is an interval. [Manski and Molinari \(2010\)](#) present an approach to deal with such intervals, which is different from the one suggested here. The difference arises partly because we ask the respondent to select a point or an interval, not both. In addition, we take the view that the individual chooses a particular interval from an unobserved set of admissible intervals.

Another advantage with the self-selected intervals is that they arguably provide a richer picture of any underlying response uncertainty, compared to bracketing and some recent approaches to cater for respondent uncertainty in contingent valuation. The currently most popular elicitation approach in contingent valuation is a payment card containing several different costs for a public good, combined with a question about how certain the respondent feels about paying a certain cost (e.g. ”definite yes”, ”probably yes”, ”probably no” and ”definite no”). Recent analysis shows that including such uncertainty-assessments in the survey instrument may affect the estimate of valuations. For a review of this literature in this area see [Broberg and Brännlund \(2008\)](#). The intervals do not burden the respondent with the task of categorizing his uncertainty about a certain quantity and no issues arise

as to how such categories should be represented in an econometric model.³

Our illustration provides an example from contingent valuation, a widely used approach to elicit individual's sentiment about non-priced goods or services (eg environmental quality). Our objective is to measure WTP (for an environmental improvement) and we will base our model on three basic assumptions:

ASSUMPTION 1 *Each respondent might not be aware of the exact location of the true WTP-point. The respondents may freely choose intervals containing their true WTP-points. The ends of stated intervals may be rounded, e.g. to simple sums of coins or paper values of money.*

ASSUMPTION 2 *The true WTP-points are independent of question mode, i.e. the content of a question does not change the true WTP-point in the self-selected WTP-interval.*

ASSUMPTION 3 *The pairs of true WTP-points and the stated self-selected WTP-intervals, corresponding to different sampled individuals, are values of independent identically distributed (i.i.d.) random variables (r.v.s).*

The first part of Assumption 1 has ample support in the contingent valuation literature. In the second part we naturally assume (but do not explicitly state) that the individual chooses an interval such that WTP does not exceed his or her income. We explicitly allow for rounding in the final part of Assumption 1. This is seldom, if ever, made explicit in the contingent valuation literature. Assumption 2 is important. From some perspectives, this assumption is very strong, given the evidence that exists on the differing results between types of valuation questions. Let us come back to it later.

³Each category of uncertainty can be considered in terms of a threshold-parameter in an ordered logit(probit) model. Assumptions are usually imposed on those parameters, an issue avoided here.

Assumption 3 is, typically, innocuous, in that most contingent valuation applications are based on significant national samples.

3. SAMPLING DESIGN

We consider the following *two-step plan* of data collecting. On the *first step* randomly sampled individuals provide self-selected intervals that contain the true point. Let n , randomly sampled respondents, from a population \mathfrak{P} of interest, have stated intervals $\mathbf{y}_1^n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i = (y_{Li}, y_{Ri}]$ containing their WTP-points. Due to rounding, the same intervals can be communicated by different respondents. We consider n as a fixed non-random number.

From our assumptions it follows that we can consider $\mathbf{y}_1^n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ as a realization of a multinomial random process $\{\mathbf{Y}_i\}_{i \geq 1}$ with the discrete time parameter $i = 1, 2, \dots$. The r.v.s $\{\mathbf{Y}_i\}_{i \geq 1}$ are i.i.d. and the set of their values is a set containing all possible self-selected intervals $\mathcal{U}_{all} = \{\mathbf{u}_\alpha : P[\mathbf{Y}_i = \mathbf{u}_\alpha] > 0, \alpha \in A\}$. The set \mathcal{U}_{all} can contain many different intervals \mathbf{u}_α but due to rounding their number is finite.

There is a discrete probability distribution $p_\alpha = P[\mathbf{Y}_i = \mathbf{u}_\alpha]$, $\alpha \in A$. The set \mathcal{U}_{all} and the probability distribution $\{p_\alpha, \alpha \in A\}$ are not known. In our context α is an integer index identifying \mathbf{u}_α , i.e. $\mathbf{u}_{\alpha'} \neq \mathbf{u}_{\alpha''}$ if $\alpha' \neq \alpha''$.

All $m(n) \leq n$ different intervals in \mathbf{y}_1^n can be ordered by their endpoints. We write $\mathbf{y}_{i_1} < \mathbf{y}_{i_2}$ if either $y_{Li_1} < y_{Li_2}$, or $y_{Li_1} = y_{Li_2}$ but $y_{Ri_1} < y_{Ri_2}$. Then $\mathbf{y}_{i'_1} < \mathbf{y}_{i'_2} < \dots < \mathbf{y}_{i'_{m(n)}}$ and we let $\mathbf{u}_{hn} = \mathbf{y}_{i'_h}$ for all \mathbf{y}_i identical with $\mathbf{y}_{i'_h}$.

The collected data can be written as a list:

$$\mathbf{d}\mathbf{1}_n = \{\dots, \{h, \{u_{Lh,n}, u_{Rh,n}\}, t_{h,n}\}, \dots\},$$

where ordering indexes $h = 1, \dots, m(n)$ of different intervals $\mathbf{u}_{h,n} = (u_{Lh,n}, u_{Rh,n}]$ depend on the collected data \mathbf{y}_1^n .

Let $\mathcal{U}_{m(n),n} = \{\mathbf{u}_{1,n}, \dots, \mathbf{u}_{m(n),n}\} \subseteq \mathcal{U}_{all}$ be the set of elicited intervals and

$\mathbf{t}(n) = \{t_{1,n}, \dots, t_{m(n),n}\}$ be the numbers $t_{h,n} = \sum_{i=1}^n I[\mathbf{y}_i = \mathbf{u}_{h,n}]$ of times that $\mathbf{u}_{h,n}$ was found in the data \mathbf{y}_1^n .

We are interested in estimating the fraction of respondents (in \mathfrak{P}) that will state WTP-intervals which are already in the set $\mathcal{U}_{m(n),n}$. Let $p_c(n)$ be probability of the event that the last WTP-interval, $\mathbf{y}_n = \mathbf{u}_{h_n,n}$, in \mathbf{y}_1^n , has been stated by at least one of the previous respondents, $i < n$. Then $t_{h_n,n} \geq 2$. Let H_n be the r.v. to state a WTP-interval $\mathbf{u}_{H_n,n}$ by the n -th respondent $\{H_n = h\} \subseteq \{X_i \in \mathcal{U}_{h,n}\}$. The expectation $E[I[t_{H_n,n} \geq 2]] = p_c(n)$. Therefore $I[t_{h_n,n} \geq 2]$ can be considered as an unbiased estimate of $p_c(n)$, and $\mathbf{t}(n)$ is a sufficient statistic with components $t_{h,n}$, $h = 1, \dots, m(n)$.

Due to uniformly random sampling (simple random sampling without replacement) of respondents from \mathfrak{P} , $\mathbf{t}(n)$ is invariant with respect to any permutation of intervals in \mathbf{y}_i^n . Hence, after averaging $I[t_{H_n,n} \geq 2]$ given $\mathbf{t}(n)$, we obtain the enhanced unbiased estimate of $p_c(n)$,

$$(3.1) \quad \hat{p}_c(n) = r(n)/n,$$

where $r(n) = \sum_{h=1}^{m(n)} t_{h,n} I[t_{h,n} \geq 2]$, and $\hat{}$ over $p_c(n)$ denotes an estimate.

We reduce our problem to a classical urn problem with $r(i)$ white and $i - r(i)$ black balls. Let $\mathcal{U}_{m(i),i}(1) = \{\mathbf{u}_h : \mathbf{u}_h \in \mathcal{U}_{m(i),i}, t_{hi} = 1\}$, $\mathcal{U}_{m(i),i}(2) = \{\mathbf{u}_h : \mathbf{u}_h \in \mathcal{U}_{m(i),i}, t_{hi} \geq 2\}$, $\mathcal{U}_{m(i),i} = \mathcal{U}_{m(i),i}(1) \cup \mathcal{U}_{m(i),i}(2)$. If $\mathbf{y}_{i+1} \in \mathcal{U}_{m(i),i}(2)$ then $\Delta_c(i) = \hat{p}_c(i+1) - \hat{p}_c(i) = (i - r(i))/(i(i+1))$. If $\mathbf{y}_{i+1} \in \mathcal{U}_{m(i),i}(1)$ then $\Delta_c(i) = (2i - r(i))/(i(i+1))$. If $\mathbf{y}_{i+1} \notin \mathcal{U}_{m(i),i}$ then $\Delta_c(i) = -r(i)/(i(i+1))$.

By calculating $\hat{p}_c(i)$ for each $i \leq n$ in (3.1) we can observe the evolution of $\hat{p}_c(i)$, see Fig. 2 in Section 5. The probability $p_c(n)$ is increasing in n . Hence, we can use $\hat{p}_c(n)$ as a lower bound estimate for $p_c(n')$, for any $n' > n$. Then we may interpret $\hat{p}_c(n)100\%$ as a lower bound estimate of the percentage of all individuals in \mathfrak{P} who would claim an interval $\mathbf{u}_h \in \mathcal{U}_{m(n),n}$. We call $p_c(n)$ a *coverage probability*.

The decision to stop collecting data on the first step can thus depend on the value of $\hat{p}_c(n)$. If this value is not sufficiently close to 1, and if it is possible to extend data collection, we have information about the value of so doing. Note that if the collection of data has stopped with $n_1 \geq n$ and $\hat{p}_c(n_1) = r(n_1)/n_1$ then inferences about the WTP-distribution will only correspond to the subset of individuals who have WTP-intervals in $\mathcal{U}_{m(n_1), n_1}$.

Henceforth, the number n_1 of randomly sampled respondents is fixed and we suppress indexes, e.g. we will write $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ instead of $\mathcal{U}_{m(n_1), n_1} = \{\mathbf{u}_{1, n_1}, \dots, \mathbf{u}_{m(n_1), n_1}\}$, and t_h instead of t_{h, n_1} . We write $\mathcal{U}_m, \mathbf{u}_h, \mathbf{v}_j, \mathcal{C}_h$ and \mathcal{D}_j instead of $\mathcal{U}_{m, n_1}, \mathbf{u}_{h n_1}, \mathbf{v}_{j n_1}, \mathcal{C}_{h n_1}$ and $\mathcal{D}_{j n_1}$.

Let $v_0 < v_1 < \dots < v_{k-1} < v_k$ be ordered values of the end points of all intervals $\mathbf{u}_h \in \mathcal{U}_m$. We let $v_{Lj} = v_{j-1}$ and $v_{Rj} = v_j$. Then the intervals in the following collection $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k\}, \mathbf{v}_j = (v_{Lj}, v_{Rj}]$ $j = 1, \dots, k$, are called *division intervals*. Each interval $\mathbf{u}_h \in \mathcal{U}_m$ is the union of disjoint division intervals

$$(3.2) \quad \mathbf{u}_h = \bigcup_{j \in \mathcal{C}_h} \mathbf{v}_j, \quad \mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}.$$

Below we will also use the following sets of indices h ,

$$(3.3) \quad \mathcal{D}_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}, \quad j = 1, \dots, k.$$

Respondents' WTP-points $\{x_i\}$ are values of i.i.d. r.v.s $\{X_i\}$. We need to estimate the d.f. $F[x] = P[X_i \leq x]$, or equivalently, the survival function (s.f.) $S[x] = 1 - F[x]$ of the true WTP-distribution. Let $H_i, i = 1, \dots, n_1$, be i.i.d. r.v.s, $H_i = h$ given that, on the first step, the i -th respondent has stated the interval \mathbf{u}_h containing his/her WTP-point x_i , $\{H_i = h\} \subseteq \{X_i \in \mathbf{u}_h\}$. We let $w_h = P[\{H_i = h\} \cap \{X_i \in \mathbf{u}_h\}]$, $w_h \leq p_h = P[X_i \in \mathbf{u}_h], h = 1, \dots, m$.

The events $\{H_i = h\} \cap \{X_i \in \mathbf{u}_h\}$ are observable, while events $\{X_i \in \mathbf{u}_{h'}\}, h' \neq h$, are not. The probability of the number of times t_1, \dots, t_m that

$\mathbf{u}_1, \dots, \mathbf{u}_m$ appears is $\prod_{h=1}^m w_h^{t_h}$, i.e. we have a multinomial distribution. The corresponding normed log likelihood (lik) is

$$\text{lik}[w_1, \dots, w_m | t_1, \dots, t_m] = \sum_{h=1}^m \frac{t_h}{n_1} \text{Log}[w_h], \quad \sum_{h=1}^m t_h = n_1.$$

The maximum of lik over $w_h \geq 0, \sum_{h=1}^m w_h = 1$, is attained at $\check{w}_h = t_h/n_1, h = 1, \dots, m$. Note that $-\sum_{h=1}^m \check{w}_h \text{Log}[\check{w}_h]$ is the empirical Entropy of the multinomial distribution with probabilities $\{\hat{w}_1, \dots, \hat{w}_m\}$.

Let us now introduce the *second step* of data collection. We prolong random sampling of new (not yet sampled) individuals from the population \mathfrak{P} . In the second step each individual is to announce an interval containing his/her WTP-point. If the interval does not belong to \mathcal{U}_m then we do not include it in the collected data. If the interval \mathbf{u}_h belongs to \mathcal{U}_m then this respondent is asked to select an interval $\mathbf{v}_j \in \mathcal{V}_k, \mathbf{v}_j \subseteq \mathbf{u}_h$, from the division \mathcal{V}_k containing his/her true WTP-point. The respondents may well abstain from answering this second question and we will cater for these events in the collected data.

The collected data will be the list of triples $\mathbf{d}\mathbf{2}_{n.2} = \{\mathbf{z}_1, \dots, \mathbf{z}_{n.2}\}, \mathbf{z}_i = \{i, \mathbf{u}_{h_i}, NA\}$ or $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$, *NA* is "no answer" to the additional question; we call these triples *singles* and *pairs*. We suppose that $n.2$ is sufficiently large and that any $\mathbf{v}_j \in \mathcal{V}_k$ was reported sufficiently many times. The size $n.2$ of the data collected in the second step should be significantly larger than n_1 .

4. STATISTICAL MODEL

Henceforth we consider r.v.s X_i and probabilities of events for individuals who being asked would have intervals included in \mathcal{U}_m .

Let us consider the conditional probabilities

$$w_{hj} = P[H_i = h | X_i \in \mathbf{v}_j].$$

We call a pair of h, j *compatible* if $\mathbf{v}_j \subseteq \mathbf{u}_h$. It is not possible to identify the true WTP-distribution if we only have the data $\mathbf{d}\mathbf{1}_{n_1}$ obtained on the first step. For identification of the true WTP-distribution on the division intervals \mathcal{V}_k it is necessary to consistently estimate the conditional probabilities $\mathbb{W} = (w_{hj})$ by using data collected on the second step. We obtain this identification asymptotically as the size of data collected in the second step is growing without bounds. The values of probabilities w_{hj} for a given $j = 1, \dots, k$ reflect behavior of individual's selecting $\mathbf{u}_h \supseteq \mathbf{v}_j$, if $X_i = x_i \in \mathbf{v}_j$. By the formula of total probability we can write

$$(4.1) \quad w_{trh} = \sum_{j \in \mathcal{C}_h} w_{hj} q_{trj},$$

where $q_{trj} = P[X_i \in \mathbf{v}_j]$, is the true probability of the event $\{\mathbf{X}_i \in \mathbf{v}_j\}$.

We have no information about the process of selecting intervals but our approach avoids the necessity of knowing the true behavioral model. The collected data with pairs $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ provides a simple way to estimate $q_{trj} = P[X_i \in \mathbf{v}_j], j = 1, \dots, k$. Let $c_{pj} = \sum_{i=1}^{n_{.2}} I[\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_j\}]$, $c_{phj} = \sum_{i=1}^{n_{.2}} I[\mathbf{z}_i = \{i, \mathbf{u}_h, \mathbf{v}_j\}]$, $n_{s2} = \sum_{h=1}^m t_{sh}$, $n_{p2} = \sum_{j=1}^k c_{pj}$, $n_{.2} = n_{s2} + n_{p2}$. The sub-indeces s and p correspond to *singles* and *pairs*, $t_{sh} = \sum_{i=1}^{n_{.2}} I[\mathbf{z}_i = \{i, \mathbf{u}_h, NA\}]$.

We consider $n_{p2}/n_{.2}$ as a frequency in the Binomial process with singles and pairs and $n_{p2}/n_{.2} \rightarrow \alpha_p \in (0, 1)$ almost surely (a.s.) as $n_{.2} \rightarrow \infty$. We can also consider the conditional Binomial process with pairs containing \mathbf{v}_j together with any $\mathbf{u}_h \supseteq \mathbf{v}_j, \mathbf{u}_h \in \mathcal{U}_m$. To reduce notational complexity for division intervals with $q_{trj} = 0$, we consider the case when all $q_{trj} > 0, j = 1, \dots, k$. The cases with some of $q_{trj} = 0$ may be considered similarly.

The strongly consistent estimates of q_{trj} in (4.1) are

$$(4.2) \quad \tilde{q}_{pj} = \frac{c_{pj}}{n_{p2}} \rightarrow q_{trj}, \quad \text{a.s.} \quad j = 1, \dots, k, \quad \text{as } n_{.2} \rightarrow \infty.$$

We can also use the pairs $\{\mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ to obtain the following strongly consis-

tent estimates of w_{hj}

$$(4.3) \quad \hat{W}_{mk} = \{\hat{w}_{hj} = \frac{c_{phj}}{c_{pj}}, \quad h \in \mathcal{D}_j, \quad j = 1, \dots, k\}.$$

In (4.3) c_{phj} is the number of saved pairs $\{i, \mathbf{u}_h, \mathbf{v}_j\}, h \in \mathcal{D}_j$, in the data $\mathbf{d2}_{n.2}$. \mathcal{D}_j is defined in (3.3). Note that $\sum_{h=1}^m \hat{w}_{hj} = 1$ for each $j = 1, \dots, k$.

Our approach is nonparametric in the sense that we do not assume that $q_j, j = 1, \dots, k$ are known functions of some parameters. In addition, we do not impose any restrictions on the behavior of respondents, in terms of the relative position of their true WTP-points inside the (potentially rounded) intervals.

Note that $\mathbf{q}_k = \{q_1, \dots, q_k\}$ is a point in the polyhedron \mathcal{S}_{k-1}

$$(4.4) \quad \mathcal{S}_{k-1} = \{\mathbf{q}_k : 0 \leq q_j \leq 1, \sum_{j=1}^k q_j = 1\}.$$

The log likelihood related to the saved data from the i th respondent is

$$(4.5) \quad \begin{aligned} \text{ll}_i[\mathbf{q}_k \mid \hat{W}_{mk}, \mathbf{d2}_{n.2}] &= I[z_i = \{i, \mathbf{u}_{h_i}, NA\}] \text{Log} \left[\sum_{j \in \mathcal{C}_{h_i}} \hat{w}_{h_i j} q_j \right] \\ &+ I[z_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}] \text{Log}[\hat{w}_{h_i j_i} q_{j_i}], \end{aligned}$$

where \hat{w}_{hj} and \mathcal{C}_{h_i} are defined in (4.3) and (3.2), respectively. In short we write $\text{ll}_i[\mathbf{q}_k]$ instead of $\text{ll}_i[\mathbf{q}_k \mid \hat{W}_{mk}, \mathbf{d2}_{n.2}]$.

Now we can write the following, normed by $n_{.2} = n_{s2} + n_{p2}$, log likelihood function corresponding to the data $\mathbf{d2}_{n.2}$ containing both singles and pairs

$$(4.6) \quad \text{llik}[\mathbf{q}_k \mid \hat{W}_{mk}, \mathbf{d2}_{n.2}] = \frac{1}{n_{.2}} \sum_{i=1}^{n_{.2}} \text{ll}_i[\mathbf{q}_k].$$

From (4.5) and (4.6) we obtain the following estimate of the normed by $n_{.2}$ log likelihood function of $\mathbf{q}_k = \{q_1, \dots, q_k\}$ corresponding to the all data,

with pairs and singles, collected on the second step

$$(4.7) \quad \begin{aligned} \text{llik}[\mathbf{q}_k \mid \hat{W}_{mk}, \mathbf{d}_{2n_2}] &= \frac{n_{s2}}{n_2} \sum_{h=1}^m \frac{t_{sh}}{n_{s2}} \text{Log} \left[\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j \right] \\ &+ \frac{n_{p2}}{n_2} \sum_{j=1}^k \frac{c_{pj}}{n_{p2}} \text{Log}[q_j] + \frac{1}{n_2} \sum_{i=1}^{n_2} \text{Log}[\hat{w}_{h_i j_i}] I[\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}]. \end{aligned}$$

Let $\partial \mathcal{S}_{k-1}$ be all boundary points of \mathcal{S}_{k-1} . The log likelihood function (4.7) is concave on $\mathcal{S}_{k-1} \setminus \partial \mathcal{S}_{k-1}$, see Theorem 7.1 in Appendix.

We have obtained (see Appendix) the following recursion, $r = 1, 2, \dots$,

$$(4.8) \quad q_j^{(r+1)} = \frac{n_{p2}}{n_{s2} + n_{p2}} q_j^{(1)} + \frac{n_{s2}}{n_{s2} + n_{p2}} \sum_{h \in \mathcal{D}_j} \frac{\hat{w}_{hj} q_j^{(r)}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}^{(r)}} \hat{w}_{sh},$$

$$q_j^{(1)} = \check{q}_{pj}, \quad \hat{w}_{sh} = \frac{1}{n_{s2}} \sum_{i=1}^{n_2} I[z_i = \{i, \mathbf{u}_h, NA\}].$$

$q_j^{(1)} = q_j^{(1)}(n_{p2}), q_j^{(r)} = q_j^{(r)}(n_{p2}, n_{s2}), \hat{w}_{sh} = \hat{w}_{sh}(n_{s2}), j = 1, \dots, k, h = 1, \dots, m, r = 2, 3, \dots$. If n_{p2} and n_{s2} are sufficiently large then

$$(4.9) \quad \hat{w}_{sh} > 0, \quad \sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_j^{(1)} > 0 \quad \text{and} \quad q_j^{(1)} = \check{q}_{pj} > 0,$$

for each $h = 1, \dots, m, j = 1, \dots, k$. If (4.9) holds then the recursion (4.8) can be applied for obtaining the ML-estimators of probabilities \mathbf{q}_{tr} maximizing llik (4.7).

THEOREM 4.1 *Suppose that Assumptions 1–3 are valid and the sizes n_{p2} and n_{s2} of collected pairs and singles are growing unboundedly as $n_2 \rightarrow \infty$. Then for any sufficiently large n_{p2} and n_{s2} the ML-estimator $\check{\mathbf{q}}_k = \{\check{q}_1, \dots, \check{q}_k\}$ (based on the data \mathbf{d}_{2n_2} with singles and pairs) of the log likelihood (4.7) exists, is strongly consistent, and can be found as the stationary point of recursion (4.8).*

Proof. From (4.2) and (4.8) it follows that $q_j^{(1)} \rightarrow q_{trj}$ a.s. $n_2 \rightarrow \infty, j = 1, \dots, k, \mathbf{q}_{trk} \in \mathcal{K}_{\alpha_p} = \{\mathbf{q}_k : q_j \geq q_{trj} \alpha_p / 2, \sum_{j=1}^k q_j = 1\} \subset \mathcal{S}_{k-1} \setminus \partial \mathcal{S}_{k-1}$.

For all sufficiently large $n_{.2}$ all $\mathbf{q}_k^{(r)} \in \mathcal{K}_{\alpha_p}$, $r \geq 1$. Due to concavity of (4.7) the sequence $\{\mathbf{q}_k^{(r)}\}_{r \geq 1}$ has only one limit stationary point of the recursion (4.8) in the compact set \mathcal{K}_{α_p} where (4.7) has its maximum. We summarize these results. From Theorem 7.1, Corollaries 7.2 - 7.3 follows existence of the ML-estimates $\check{\mathbf{q}}_k = \check{\mathbf{q}}_{kM}$ which converge a.s. to \mathbf{q}_{trk} . Corollary 7.2 justifies usage of the recursion (4.8) to approximate the ML-estimates $\check{\mathbf{q}}_k$. \square

Hence, we can find consistent ML-estimates \check{q}_j , $j = 1, \dots, k$, of the projected WTP-distribution on the division intervals $\mathbf{v}_j \in \mathcal{V}_k$, i.e. $\check{q}_j \rightarrow q_{trj}$ a.s. $j = 1, \dots, k$, if n_{p2} and n_{s2} are growing. Note that this inference can be only applied to the respondents in \mathfrak{P} who choose $\mathbf{u}_h \in \mathcal{U}_m$.

Besides consistency we need to know the accuracy of the obtained ML-estimate $\check{\mathbf{q}}_k$, or $\check{\mathbf{q}}_{k-1} = \{\check{q}_1, \dots, \check{q}_{k-1}\}$, $\check{q}_k = 1 - \sum_{j=1}^{k-1} \check{q}_j$. In our case we are interested in consistently estimating the distribution of rescaled deviations $\sqrt{n_{.2}}(\check{\mathbf{q}}_{k-1} - \mathbf{q}_{tr,k-1})$,

$$(4.10) \quad Q_{q_{tr,k-1}}[x_1, \dots, x_{k-1} \mid \hat{\mathbb{W}}_{mk}, \mathbf{d}\mathbf{2}_{n_{.2}}] = P \left[\bigcap_{j=1}^{k-1} \{ \sqrt{n_{.2}}(\check{q}_j - q_{trj}) \leq x_j \} \right].$$

The log likelihoods $\text{ll}_i[\mathbf{q}_k]$ in (4.5) are values of r.v.s

$$(4.11) \quad \begin{aligned} \text{LL}_i[\mathbf{q}_k] &= I[Z_i = \{i, \mathbf{u}_{H_i}, NA\}] \text{Log} \left[\sum_{j \in \mathcal{E}_{H_i}} \hat{w}_{H_i j} q_j \right] \\ &+ I[Z_i = \{i, \mathbf{u}_{H_i}, \mathbf{v}_{J_i}\}] \text{Log}[\hat{w}_{H_i J_i} q_{J_i}], \quad i = 1, \dots, n_{.2}, \end{aligned}$$

where H_i and J_i are random indexes of \mathbf{u}_{H_i} and \mathbf{v}_{J_i} . The asymptotics of deviations ML-estimates in parametric statistical models with independent observations are rather completely investigated if regularity properties (RP), of the observations' log likelihoods $\text{LL}_i[\mathbf{q}_k]$, hold (Lehmann and Casella (1998)). In our case these RP can be described as follows.

RP1: all first and second order partial derivatives

$$\frac{\partial \text{LL}_i[\check{\mathbf{q}}_k]}{\partial q_{j_1}}, \frac{\partial^2 \text{LL}_i[\check{\mathbf{q}}_k]}{\partial q_{j_1} \partial q_{j_2}}, \quad 1 \leq j_1, j_2 \leq k-1, \quad \check{\mathbf{q}}_k = \{q_1, \dots, q_{k-1}, 1 - \sum_{j=1}^{k-1} q_j\},$$

exist and are continuous functions on the set of all possible values \mathbf{q}_k .

RP2: these derivatives possess finite absolute moments of an order $\gamma > 2$.

RP3: the second order partial derivatives are uniformly continuous on any containing \mathbf{q}_{trk} compact set of possible values \mathbf{q}_k ,

RP4: for all $1 \leq j_1, j_2 \leq k - 1$

$$E_{\mathbf{q}_{trk}} \left[\frac{\partial \text{LL}_i[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} \right] = 0,$$

$$E_{\mathbf{q}_{trk}} \left[\frac{\partial^2 \text{LL}_i[\tilde{\mathbf{q}}_k]}{\partial q_{j_1} \partial q_{j_2}} \right] = -E_{\mathbf{q}_{trk}} \left[\frac{\partial \text{LL}_i[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} \frac{\partial \text{LL}_i[\tilde{\mathbf{q}}_k]}{\partial q_{j_2}} \right].$$

In order to formulate **RP5** we introduce the Fisher information (FI-) matrices

$$\mathbb{F}_{n_2}[\tilde{\mathbf{q}}_k] = \frac{1}{n_2} \sum_{j=1}^{n_2} \left(E_{\mathbf{q}_{tr,k-1}} \left[\frac{\partial^2 \text{LL}_i[\mathbf{q}_k]}{\partial q_{j_1} \partial q_{j_2}} \right] \right)_{1 \leq j_1 \leq j_2 \leq k-1}.$$

RP5: for each possible \mathbf{q}_k the FI-matrix has full rank $k - 1$.

In Appendix we show that **RP1 - RP5** are valid for the statistical model with log likelihood (4.7). These properties imply the standard regularity properties.

The distributions, of rescaled deviations of ML-estimators from true values of parameters, thus converge to normal distributions with the inverse FI-matrix (Lehmann and Casella (1998)). In our case, with large number ($k - 1$) of parameters to be estimated, it is more convenient to use an universal method for assessing accuracies of estimated parameters based on resamplings (Davison and Hinkley (1995); Belyaev and Nilsson (1997); Nilsson (1998); Belyaev (2007)). We proceed by outlining a procedure for resampling in the considered statistical model of the data \mathbf{d}_{2n_2} .

Let $\mathbf{J}_{n_2}^* = \{J_{1n_2}^*, \dots, J_{n_2n_2}^*\}$ be the list of i.i.d. r.v.s uniformly distributed over the range of natural numbers $\{1, 2, \dots, n_2\}$, i.e. $P^*[J_{i'n_2}^* = i] = \frac{1}{n_2}, 1 \leq i' \leq n_2$. We use “*” to denote r.v.s, probabilities of events, mathematical

expectations, etc., related to the r.v.s $J_{in_2}^*$, $1 \leq i \leq n_2$. The values of these r.v.s can be simulated via a random numbers generator.

It is of interest to simulate resampled, with replacement, copies of data $\mathbf{d2}_{n_2}$. Let $\{j_{1n_2}^{*c}, \dots, j_{n_2, n_2}^{*c}\}$, be the c -th resampled copy of the list $\{1, \dots, n_2\}$, obtained by simulation i.i.d. values of (pseudo) r.v.s $J_{in_2}^{*c}$. Then the c -th copy of $\mathbf{d2}_{n_2}$ is the list $\mathbf{d2}_{n_2}^{*c} = \{\mathbf{z}_1^{*c}, \dots, \mathbf{z}_{n_2}^{*c}\}$, where triples $\mathbf{z}_i^{*c} = \mathbf{z}_{\tilde{i}^{*c}}, \tilde{i}^{*c} = j_{in_2}^{*c}, i = 1, \dots, n_2$. We can consider the list $\mathbf{d2}_{n_2}^{*c}$ as a realisation of a multinomial process with the resampled log likelihood

$$(4.12) \quad \text{lik}^{*c}[\mathbf{q}_k \mid \hat{\mathbb{W}}_{mk}, \mathbf{d2}_{n_2}^{*c}] = \sum_{i=1}^{n_2} \tilde{n}_{i, n_2}^{*c} \text{ll}_i[\mathbf{q}_k],$$

where $\text{ll}_i[\mathbf{q}_k]$ are defined in (4.5), and $\tilde{n}_{i, n_2}^{*c} = \sum_{i=1}^{n_2} I[\tilde{i}^{*c} = i]$.

All methods exposed in Appendix can be applied to the analysis of the statistical model with data $\mathbf{d2}_{n_2}^{*c}$ and the log likelihood (4.12). We can do that with the following notations:

$$\tilde{n}_{s2}^{*c} = \sum_{i=1}^{n_2} I[\mathbf{z}_i^{*c} = \{\tilde{i}^{*c}, \mathbf{u}_{h_{\tilde{i}^{*c}}}, NA\}],$$

$$\tilde{n}_{p2}^{*c} = \sum_{i=1}^{n_2} I[\mathbf{z}_i^{*c} = \{\tilde{i}^{*c}, \mathbf{u}_{h_{\tilde{i}^{*c}}}, \mathbf{v}_{j_{\tilde{i}^{*c}}}\}],$$

$$\tilde{n}_{sh}^{*c} = \sum_{i=1}^{n_2} I[\mathbf{z}_i^{*c} = \{\tilde{i}^{*c}, \mathbf{u}_h, NA\}],$$

$$\tilde{n}_{ph}^{*c} = \sum_{i=1}^{n_2} I[\mathbf{z}_i^{*c} = \{\tilde{i}^{*c}, \mathbf{u}_h, \mathbf{v}_{j_{\tilde{i}^{*c}}}\}],$$

$$\tilde{n}_{phj}^{*c} = \sum_{i=1}^{n_2} I[\mathbf{z}_i^{*c} = \{\tilde{i}^{*c}, \mathbf{u}_h, \mathbf{v}_j\}],$$

We can rewrite the log likelihood (4.12) as follows

$$\begin{aligned}
& \text{lik}^{\star c}[\mathbf{q}_k \mid \hat{\mathbb{W}}_{mk}, \mathbf{d}\mathbf{2}_{n.2}^{\star c}] \\
&= \frac{\tilde{n}_{s2}^{\star c}}{n.2} \sum_{h=1}^m \frac{t_{sh}^{\star c}}{\tilde{n}_{s2}^{\star c}} \text{Log} \left[\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j \right] + \frac{\tilde{n}_{p2}^{\star c}}{n.2} \sum_{j=1}^k \frac{t_{pj}^{\star c}}{\tilde{n}_{p2}^{\star c}} \text{Log}[q_j] \\
(4.13) \quad &+ \frac{\tilde{n}_{p2}^{\star c}}{n.2} \sum_{j=1}^k \sum_{j \in \mathcal{D}_j} \frac{t_{phj}^{\star c}}{\tilde{n}_{p2}^{\star c}} \text{Log}[w_{hj}],
\end{aligned}$$

The recursion corresponding the log likelihood (4.13) is

$$(4.14) \quad q_j^{\star c(r+1)} = \frac{\tilde{n}_{p2}^{\star c}}{n.2} q_j^{\star c(1)} + \frac{\tilde{n}_{s2}^{\star c}}{n.2} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}^{\star c}}{\tilde{n}_{s2}^{\star c}} \frac{\hat{w}_{hj} q_j^{\star c(r)}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}^{\star c(r)}}, \quad j = 1, \dots, k,$$

where $q_j^{\star c(1)} = t_{pj}^{\star c} / \tilde{n}_{p2}^{\star c}$. The limit stationary point is also the solution of (4.14) as $n.2 \rightarrow \infty$.

Theorem 7.1 and 7.2, Corollaries 7.1 - 7.3 and regularity properties **RP1** - **RP5** are valid for resampled data $\mathbf{d}\mathbf{2}_{n.2}^{\star c}$ with the log likelihood (4.13). Hence, the following Corollary is valid.

COROLLARY 4.1 *Statistical model with resampled data $\mathbf{d}\mathbf{2}_{n.2}^{\star c}$ and the log likelihood function (4.12) possesses regularity properties **RP1** - **RP5** and has consistent ML-estimator $\check{\mathbf{q}}_k^{\star c} - \mathbf{q}_{trk} \rightarrow 0$, as $n.2 \rightarrow \infty$.*

In order to estimate the distribution of deviations $Q_{\mathbf{q}_{tr,k-1}}[x_1, \dots, x_{k-1} \mid \hat{\mathbb{W}}_{mk}, \mathbf{d}\mathbf{2}_{n.2}]$ defined in (4.10) we need to simulate sufficiently large number R of data copies $\mathbf{d}\mathbf{2}_{n.2}^{\star c}$ and find corresponding ML-estimates $\check{\mathbf{q}}_k^{\star c}$, $c = 1, \dots, R$. Let

$$(4.15) \quad Q_{n.2, R}^{\star}[x_1, \dots, x_{k-1}] = \frac{1}{R} \sum_{c=1}^R \prod_{j=1}^{k-1} I[\sqrt{n.2}(\check{q}_j^{\star c} - \check{q}_j) \leq x_j],$$

where \check{q}_j is the j th component in the vector ML-estimate $\check{\mathbf{q}}_{kM}$ of \mathbf{q}_{trk} .

THEOREM 4.2 *Suppose that Assumptions 1 - 3 are valid. Then the components of log likelihood (4.6) and (4.12) possess the regularity properties,*

and for any real values x_1, \dots, x_{k-1}

$$(4.16) \quad \sup_{x_1, \dots, x_{k-1}} |Q_{n_2, R}^*[x_1, \dots, x_{k-1}] - Q_{tr, k-1}[x_1, \dots, x_{k-1}]| \rightarrow 0,$$

in probability as $n_2 \rightarrow \infty$ and $R \rightarrow \infty$.

Proof. From Theorem 4.1 and Corollary 4.1 we have that in the considered original statistical model with data $\mathbf{d2}_{n_2}$ and in the statistical models with resampled data $\mathbf{d2}_{n_2}^{*c}$ exist consistent ML-estimates $\check{\mathbf{q}}_{kM}$ and $\check{\mathbf{q}}_k^{*c}$ for all sufficiently large n_2 . Properties **PD**, **PE**, **PR** and **PD***, **PE***, **PR*** are also valid (see Appendix). Then the statement in Theorem 4.2 is a special case of Proposition 2 in Appendix. It follows that the distributions of deviations $\sqrt{n_2}(\check{\mathbf{q}}_{kM} - \mathbf{q}_{trk})$ and $\sqrt{n_2}(\check{\mathbf{q}}_k^{*c} - \check{\mathbf{q}}_{kM})$ are approaching in probability as $n_2 \rightarrow \infty$. These distributions of deviations are asymptotically normal and therefore convergence in the uniform metric stated in (4.16) is valid.

Mean WTP, i.e. $m_{tr1} = E[X_i]$, is an essential parameter in cost-benefit analysis. Here, however, the rounding of WTP-intervals excludes the possibility to find an unbiased estimate of m_{tr1} . Therefore, we introduce an approximation to m_{1tr} by the following consistently estimable *medium mean* of the true WTP-distribution

$$(4.17) \quad mm_{1tr} = v_{L1} + \sum_{j=1}^k \left(q_{tr, j+1}^k + \frac{1}{2} q_{trj} \right) (v_{Rj} - v_{Lj}),$$

where $q_{tr, j+1}^k = \sum_{i=j+1}^k q_{tri}$, $q_{tr, k+1} = 0$.

We can also consistently estimate the low m_{1tr}^- and upper m_{1tr}^+ boundaries for m_{1tr} . Let $q_j^k = q_j + \dots + q_k$, $q_k^k = q_k$, $q_{k+1}^k = 0$, $q_1^k = 1$. Hence, we have

$$(4.18) \quad m_{1tr}^- = v_{L1} + \sum_{j=1}^{k-1} q_{j+1, tr}^k (v_{Rj} - v_{Lj}), \quad m_{1tr}^+ = v_{L1} + \sum_{j=1}^k q_{j, tr}^k (v_{Rj} - v_{Lj}).$$

From (4.17), (4.18) we have

$$m_{1tr}^- \leq m_{1tr} \leq m_{1tr}^+, \quad m_{1t}^- \leq mm_{1tr} \leq m_{1tr}^+.$$

Note that m_{1tr}^-, m_{1tr}^+ and mm_{1tr} are linear forms of $q_{jtr}, j = 1, \dots, k$. We introduced mm_{1tr} because it better approximates m_{1tr} than m_{1tr}^- and m_{1tr}^+ . Under some additional assumptions on the true WTP-distribution the estimate mm_{1tr} can be rather close to m_{1tr} . The characteristic mm_{1tr} can be consistently estimated and its accuracy can be also consistently estimated as $n_{.2} \rightarrow \infty$. The same is also true for the low and upper boundaries m_{1tr}^+, m_{1tr}^- . To underline dependency of the considered characteristics of data size $n_{.2}$ we add it into the indices of considered characteristics. We obtain consistent estimators substituting for all q_j the corresponding ML-estimates $\check{q}_{jn_{.2}}, j = 1, \dots, k$.

The characteristics (4.17) and (4.18) are linear forms of $\check{q}_j = \check{q}_{jn_{.2}}, j = 1, \dots, k$. Hence, by substitution ML-estimates $\check{\mathbf{q}}_{kn_{.2}}$ instead of \mathbf{q}_k in (4.18) we obtain the following strongly consistent ML-estimates of $\check{m}_{1n_{.2}}^+, \check{m}_{1n_{.2}}^-$ and $\check{m}\check{m}_{1n_{.2}}$

$$(4.19) \quad \check{m}_{1n_{.2}}^- = v_{L1} + \sum_{j=1}^{k-1} \check{q}_{j+1, n_{.2}}^k (v_{Rj} - v_{Lj}), \quad \check{m}_{1n_{.2}}^+ = v_{R1} + \sum_{j=1}^k \check{q}_{j, n_{.2}}^k (v_{Rj} - v_{Lj}).$$

We have the following strongly consistent ML-estimate of mm_{1tr}

$$(4.20) \quad \check{m}\check{m}_{1n_{.2}} = v_{L1} + \sum_{j=1}^k \left(\check{q}_{j+1, n_{.2}}^k + \frac{1}{2} \check{q}_{jn_{.2}} \right) (v_{Rj} - v_{Lj}).$$

$\check{q}_{j+1, n_{.2}}^k = \sum_{i=j+1}^k \check{q}_{in_{.2}}, j = 1, \dots, k, \check{q}_{k+1, n_{.2}} = 0$. Note that $\check{m}\check{m}_{1n_{.2}}$ is the area below the broken line, with points $\{v_{Rj}, \check{q}_{j+1, n_{.2}}^k\}, j = 1, \dots, k-1$ and $\{v_{Rk}, 0\}$ on the estimated WTP-survival function, connected by intervals.

By replacing in (4.19) and (4.20) the estimates $\check{q}_{jn_{.2}}$ by $\check{q}_{jn_{.2}}^{*c}, j = 1, \dots, k$, we obtain copies of ML-estimates based on the resampled log likelihoods (4.12)

$$\check{m}_{1n_{.2}}^{-*c} = v_{L1} + \sum_{j=1}^{k-1} \check{q}_{j+1, n_{.2}}^{*ck} (v_{Rj} - v_{Lj}),$$

$$\check{m}_{1n.2}^{+*c} = v_{L1} + \sum_{j=1}^k \check{q}_{j,n.2}^{*ck} (v_{Rj} - v_{Lj}),$$

$$(4.21) \quad \check{m}m_{1n.2}^{*c} = v_{L1} + \sum_{j=1}^k \left(\check{q}_{j+1,n.2}^{*ck} + \frac{1}{2} \check{q}_{j,n.2}^{*c} \right) (v_{Rj} - v_{Lj}),$$

$c = 1, \dots, C$.

All these estimates (4.19) - (4.21) are linear forms w.r.t. q_{jtr} , $\check{q}_{j,n.2}$ and $\check{q}_{j,n.2}^{*c}$. Therefore, from Theorem 4.2 we obtain

COROLLARY 4.2 *If Assumptions 1 - 3 are valid, then the regularity properties of components in log likelihood (4.6), (4.12) are also valid and*

$$\frac{1}{C} \sum_{c=1}^C I[\sqrt{n.2}(\check{m}_{1n.2}^{-*c} - \check{m}_{1n.2}^-) \leq x] - P[\sqrt{n.2}(\check{m}_{1n.2}^- - m_{1tr}^-) \leq x] \rightarrow 0,$$

$$\frac{1}{C} \sum_{c=1}^C I[\sqrt{n.2}(\check{m}_{1n.2}^{+*c} - \check{m}_{1n.2}^+) \leq x] - P[\sqrt{n.2}(\check{m}_{1n.2}^+ - m_{1tr}^+) \leq x] \rightarrow 0,$$

$$\frac{1}{C} \sum_{c=1}^C I[\sqrt{n.2}(\check{m}m_{1n.2}^{*c} - \check{m}m_{1n.2}) \leq x] - P[\sqrt{n.2}(\check{m}m_{1n.2}^- - mm_{1tr}) \leq x] \rightarrow 0,$$

in probability as $n_{p2} \rightarrow \infty$ and $C \rightarrow \infty$.

5. NUMERICAL EXPERIMENTS

We use data collected and analyzed by Håkansson (2007) as if they were obtained on the first step in our two-step design of data collecting. During the first step of data collecting randomly sampled $n_1 = 241$ respondents report self-selected WTP-intervals, containing their true WTP-points. The set $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ contains all intervals $\mathbf{u}_h = (u_{Lh}, u_{Rh}]$, $h = 1, \dots, m$. The following four WTP-intervals (50, 100], (20, 50], (100, 200], and (20, 100] have been stated 69, 39, 23, and 11 times respectively, i.e. near 60% of all respondents have picked these four intervals. The variability of

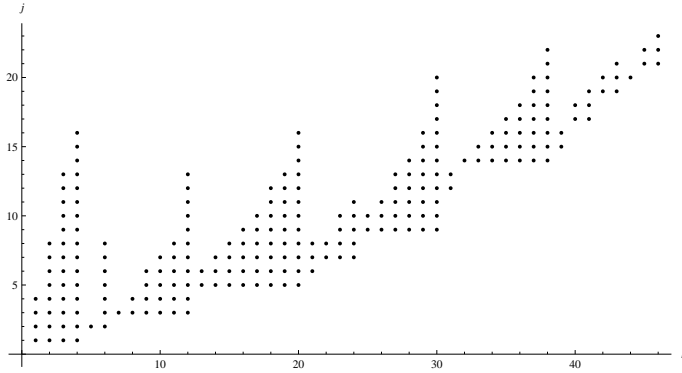


FIGURE 1.— The set of all compatible indexes $h, j, h = 1, \dots, 46, j = 1, \dots, 23$. The sets $\mathcal{C}_h = \{j, \mathbf{v}_j \subseteq \mathbf{u}_h\}, \mathcal{D}_j = \{h, \mathbf{v}_j \subseteq \mathbf{u}_h\}$, are h -cuts and j -cuts of the shown set.

rounded self-selected intervals can be described by disjoint division intervals $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ corresponding to \mathcal{U}_m . In our numerical experiment we have the following, corresponding to $U_m, m = 46$, division intervals: $v_1 = (0, 5], v_2 = (5, 10], v_3 = (10, 15], v_4 = (15, 20], v_5 = (20, 25], v_6 = (25, 30], v_7 = (30, 40], v_8 = (40, 50], v_9 = (50, 60], v_{10} = (60, 70], v_{11} = (70, 75], v_{12} = (75, 80], v_{13} = (80, 100], v_{14} = (100, 150], v_{15} = (150, 170], v_{16} = (170, 200], v_{17} = (200, 250], v_{18} = (250, 300], v_{19} = (300, 400], v_{20} = (400, 500], v_{21} = (500, 600], v_{22} = (600, 1000], v_{23} = (1000, 2000]$.

Each $\mathbf{u}_h \in \mathcal{U}_m$ is the union of related division intervals. Pairs of indexes $\{h, j\}$ are compatible if $\mathbf{v}_j \subseteq \mathbf{u}_h$. We order the indexes as follows: $h_1 < h_2$ if $u_{Lh_1} < u_{Lh_2}$ or if $u_{Lh_1} = u_{Lh_2}$ and $u_{Rh_1} < u_{Rh_2}$, and then let $j_1 < j_2$ if $v_{Lj_1} < v_{Lj_2}$. The set of all compatible indexes $\{h, j\}$ is shown in Figure 1. The sets of indexes $\mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}$ and $\mathcal{D}_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}$, are the vertical cuts of the set with fixed h , and the horizontal cuts with fixed j , respectively. $n_1 = 241$ intervals have been collected and 21 intervals have been stated only once.

From (3.1) we estimate the coverage probability by $\hat{p}_c[241] = 0.9128$. In

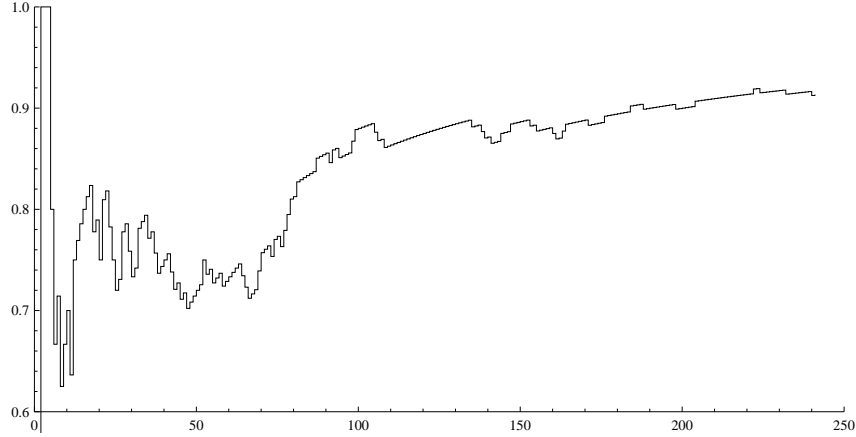


FIGURE 2.— The dynamics of the coverage probability estimates $\hat{p}_c[i]$ as a function of i respondents sequentially selected on the first step.

Figure 2 the part of the plot, with values $\hat{p}_c[i], i = 1, 2, \dots, 241$, is shown. The coverage probability $\hat{p}_c[i]$ characterizes the fraction of respondents in the whole population, who being selected after the i th respondent, that happen to communicate an interval that has been chosen among the first i selected respondents. A more detailed evolution of $\hat{p}_c[i], 180 \leq i \leq 241$ is shown in Figure 3.

If we knew all conditional probabilities $w_{hj} = P[H_i = h \mid X_i \in \mathbf{v}_j]$ then we could consistently estimate many useful characteristics of the true WTP-distribution using only WTP-intervals in the first step. If we know $\mathbb{W}_{mk} = (w_{hj})$ then under the Assumptions 1 - 3 we can write the log likelihood normed by n_1 as follows

$$(5.1) \quad \text{llik}[\mathbf{q}_k \mid \mathbb{W}_{mk}, t_1, \dots, t_m] = \sum_{h=1}^m \frac{t_h}{n_1} \text{Log} \left[\sum_{j \in \mathcal{C}_h} w_{hj} q_j \right], \quad q_j \geq 0, \quad \sum_{j=1}^n q_j = 1.$$

One can consider many examples with different arrays \mathbb{W}_{mk} . Each \mathbb{W}_{mk} describes a specific behavior of respondents in their selection of intervals. We

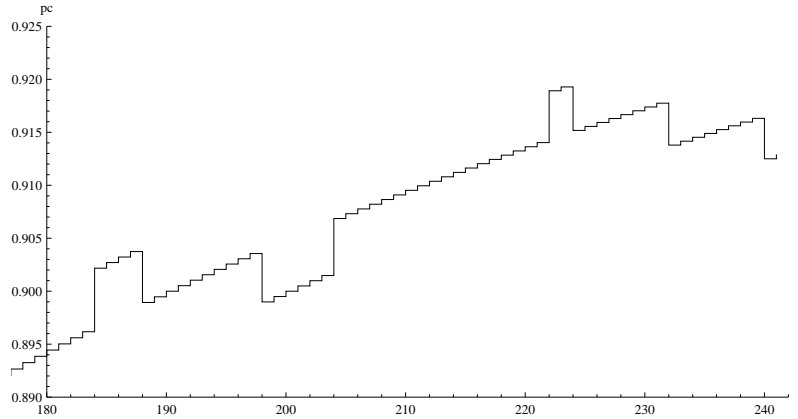


FIGURE 3.— The dynamic in i of the probability $\hat{p}_c[i]$ for i in $\{180, \dots, 241\}$.

have considered the following five hypothetical behavior models of respondents:

BM1 is the behavior model of indifferent respondents with $w_{hj} = 1/d_j$, $h \in \mathcal{D}_j$, $j = 1, \dots, k (= 23)$, d_j is the size of \mathcal{D}_j .

BM2 is the behavior model of respondents who, with \mathbf{v}_j containing their WTP-point, that chooses \mathbf{u}_h in which \mathbf{v}_j is the last division interval.

BM3 is the behavior model of respondents who, with \mathbf{v}_j containing their WTP-point, pick \mathbf{u}_h , $h \in \mathcal{D}_j$ with probability proportional to w_{trh} , $h = 1, \dots, m$.

Here, by using estimated frequencies $\hat{w}_h = \frac{t_h}{n}$ we take

$$(5.2) \quad w_{hj} = \frac{t_h}{\left(\sum_{h' \in \mathcal{D}_j} t_{h'}\right)}.$$

If $\mathbf{u}_h = \cup_{j \in \mathcal{C}_h} \mathbf{v}_j = \mathbf{v}_{j_h+1} + \mathbf{v}_{j_h+2} + \dots + \mathbf{v}_{j_h+M}$ then $\mathbf{v}_j = \mathbf{v}_{j_h+a} \subseteq \mathbf{u}_h$ has internal rank $r_{hj} = a \leq M$. The notion of the interval rank is used in two following behavior models.

BM4 is the behavior model of respondents who, with \mathbf{v}_j containing their WTP-point, report \mathbf{u}_h , $h \in \mathcal{D}_j$, with probabilities proportional to

$w_h(r_{hj}/r_{hM})^{c1-1}(1-(r_{hj}/r_{hM})^{c2-1})$, i.e. with the discrete Beta-type weighting of probabilities $w_h, h = 1, \dots, m$.

BM5 is the behavior model of respondents who, with \mathbf{v}_j containing their WTP-point, select $\mathbf{u}_h, h \in \mathcal{D}_j$, with probability proportional to

$$w_h(1 + c(r_{hj}/r_h)), h \in \mathcal{D}_j, j = 1, \dots, k, r_h = \sum_{j \in \mathcal{C}_h} r_{hj}.$$

Let $\text{mllik}^{(BM)}$ be the maximum of llik in (5.1) with $\mathbb{W}_{mk} = (w_{hj})$ corresponding behavior model BM . We can interpret the maxima of llik functions corresponding to different behavior models as measures of their plausibility to approximate real behavior of respondents. Note that $-\text{mllik}^{(BM)}$ is the entropy of the distribution $\{\hat{w}_1^{(BM)}, \dots, \hat{w}_m^{(BM)}\}$ of frequencies $\hat{w}_h^{(BM)} = \sum_{j \in \mathcal{D}_j} w_{hj}^{(BM)} \check{q}_j^{(BM)}, BM = BM1, \dots, BM5$. Here $\check{q}_j^{(BM)}$ are the ML-estimates of $q_{trj}, j = 1, \dots, k$. The empirical entropy $eew = \sum_{h=1}^m t_h/n_1 \text{Log}[n_1/t_h] = 2.8224$, where $t_h, h = 1, \dots, m = 46$, are the numbers of respondents who chose $\mathbf{u}_h \in \mathcal{U}_m$. In our calculations we have supposed that the true WTP-survival function is a mixture of the Weibull and the Exponential survival functions with unknown parameters p, a, b, m_1 , i.e. $sf_{WTP}[x] = pe^{-(x/a)^b} + (1-p)e^{-x/m_1}$. Results of our calculations are collected in the following Table 1.

Table 1. **Comparisons with the empirical entropy $eew = 2.8224$.**

Behavior Model	BM1	BM2	BM3	BM4	BM4	BM5
c1	-	-	-	1.9256	1.9565	16.872
c2	-	-	-	1.5362	1.9565	-
p	0.8849	0.7766	0.8512	0.8305	0.8426	0.8160
a	77.4627	4.4643	74.0964	76.7255	74.7109	74.9982
b	1.8425	1.9342	1.7638	1.8547	1.7942	1.8374
m1	286.7622	276.6101	259.3601	272.9070	271.5213	254.7344
mean WTP	93.8918	126.854	94.7380	102.8351	98.73626	101.2456
mllik ^(BM)	-3.9388	-3.4576	-2.9301	-2.9079	-2.9109	-2.8591
$-\frac{\text{mllik}^{(BM)}}{\text{eew}}$	1.396	1.225	1.038	1.030	1.031	1.013

We see that the values of $\frac{\text{mllik}^{(BM)}}{\text{eew}}$ depends on the behavior model BM . The best correspondence we have is with $BM = BM5$. This dependency on the assumed behavioral model shows why the direct usage of the survival theory, with the assumption of independency of censoring intervals from positions of covered true WTP-points, will lead to erroneous estimation of the true WTP-distribution. From Table 1 we conclude that to estimate conditional probabilities $\mathbb{W}_{mk} = (w_{hj})$ consistently, we need extended empirical data. We can do that on the second step of data collecting.

In our numerical experiment we suppose that the WTP-distribution is a p_{WE} -mixture WE of the Weibull distribution $W(a, b)$ and the Exponential distribution $E(m_1)$ with parameters $p_{WE} = 0.8160, a = 74.8992, b = 1.8374, m_1 = 254.7344$,

$$(5.3) \quad sf_{WTP}[x] = p_{WE}e^{-(x/a)^b} + (1 - p_{WE})e^{-x/m_1}.$$

We use the behavior model $BM3$ with array $\mathbb{W}_{mk} = (w_{hj})$, for any $j = 1, \dots, k$, and $h \in \mathcal{D}_j, w_{hj} = t_h / \left(\sum_{h' \in \mathcal{D}_j} t_{h'} \right)$. The numbers $t_h, h = 1, \dots, m$, are taken from $\mathbf{d1}_{n_1}$ and w_{hj} correspond to (5.2). We could take any other array with $w_{hj} \geq 0$, and $\sum_{h' \in \mathcal{D}_j} w_{hj} = 1$.

Let p_{NA} be the probability that a respondent states only a self-selected interval. $(1 - p_{NA})$ is the probability of finding a respondent that reports a self-selected interval and a division interval, both containing the true point. If \mathbf{u}_h has already been reported then the only \mathbf{u}_h interval compatible with $\mathbf{v}_j \in \mathcal{C}_h$ containing the true WTP-point X_i can be added. In the our example we let $p_{NA} = 2/3$. Then, in the mean, two of each three respondents will reject to add division intervals to their self-selected intervals. Recall that we use the parameters stated just before (5.3).

Let $q_j = P[X_i \in \mathbf{v}_j] = sf_{WTP}[v_{Lj}] - sf_{WTP}[v_{Rj}]$ where the survival function is given in (5.3). With probability $p_{NA}q_jw_{hj}$ the i -th respondent reports \mathbf{u}_h and rejects to add a \mathbf{v}_j . Then the single triple $\{i, \mathbf{u}_h, NA\}$ is added to simulated data. With probability $(1 - p_{NA})q_jw_{hj}$ the i -th respondent states both \mathbf{u}_h and \mathbf{v}_j . Then the triple $\{i, \mathbf{u}_h, \mathbf{v}_j\}$ is added to simulated data.

We have simulated data $\mathbf{d}\mathbf{2}_{n.2}$, $n.2 = 9000$, containing $n_{s2} = 5856$ singles and $n_{p2} = 3144$ pairs. Henceforth we call $\mathbf{d}\mathbf{2}_{n.2}$ the true data. As well we call the WTP-survival function in (5.3) and corresponding characteristics "true". By using values $\mathbf{u}_{h_i}, \mathbf{v}_{j_i}$ in all n_{p2} pairs we obtain estimates \check{q}_{pj} as in (4.2), $j = 1, \dots, 23$. To improve these estimates we have to use all data $\mathbf{d}\mathbf{2}_{n.2}$ with pairs and singles. The improved ML-estimate $\check{\mathbf{q}}_{kM}$, corresponding to log likelihood (4.7), can be obtained by the recursion (4.8). After 10 iterations with $q_j^{(1)} = \check{q}_{pj}$ we obtain the following list with approximating q_{trj} values $\check{q}_j^{(10)}$, $j = 1, \dots, 23$,

{0.01267, 0.01848, 0.02284, 0.03063, 0.03495, 0.04188, 0.09621, 0.09395,
0.08782, 0.08755, 0.04739, 0.03548, 0.12409, 0.12905, 0.02921, 0.01713,
0.01930, 0.01355, 0.02429, 0.00750, 0.00804, 0.01347, 0.00451}.

In the Figure 4 the histogram of approximated estimates $\check{q}_j^{(10)}$ of \check{q}_{trj} , $j = 1, \dots, 23$, is shown together with the true probability density of the true WTP-distribution (5.3). The histogram is shown over the part of range $0 \leq x \leq 600SEK$.

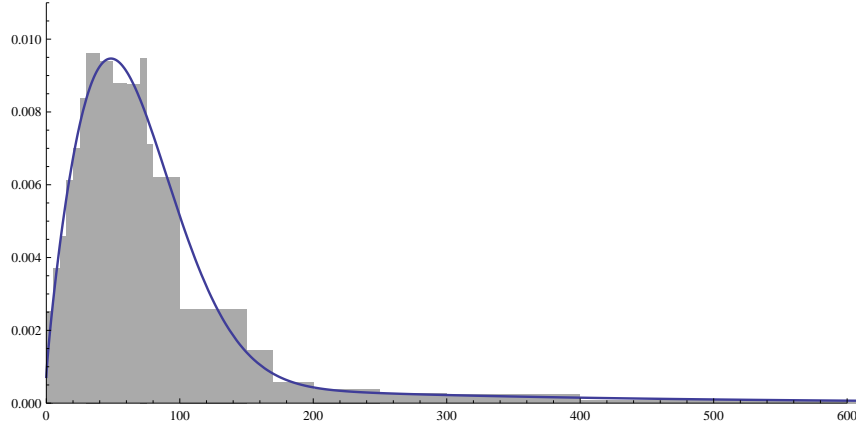


FIGURE 4.— The true probability density (smooth line) and the ML-estimate of the true WTP-histogram (light gray area) with stepwise boundary line.

Let $v_{Rj}^k = \sum_{j'=j}^k v_{Rj'}$ be values of the right divisions' ends, and let $q_{trj}^k = \sum_{j'=1}^k q_{trj'}$, $\tilde{q}_j^{(10)k} = \sum_{j'=1}^k \tilde{q}_{j'}^{(10)k}$, where $\tilde{q}_j^{(10)k}$ are obtained after 10 iterations of recursion (4.8). The true points $\{v_{Rj}^k, q_{trj}^k\}$ on the true survival function (5.3) are approximated by points $\{v_{Rj}^k, \tilde{q}_j^{(10)k}\}$. The obtained approximation is rather accurate. To illustrate the result more clearly, Figure 5 shows the true WTP-survival function (5.3) together with its true points and approximated points $\{v_{Rj}^k, \tilde{q}_j^{(10)k}\}$ over the part $0 \leq x \leq 200SEK$ of the whole range $0 \leq x \leq 2000SEK$. The area below the true WTP-survival function, shown in Figure 5, is the true mean WTP-value $m_{1tr} = 101.161SEK$. However, in an application we do not know the true WTP-survival function. As explained in Section 4 we may consider, as an approximation of m_{1tr} , the WTP-medium mean value mm_{1tr} . The true value is $mm_{1tr} = 103.718$. In our numerical experiment we can consistently estimate it as $n_2 \rightarrow \infty$. By using 10 iteration with recursion (4.8) we obtain $q_j^{(10)}, j = 1, \dots, 23$. We substitute these values in (4.20) and obtain that the approximated ML-estimate $m\check{m}_1^{(10)} = 104.281SEK$. The deviations $m\check{m}_1^{(10)} - mm_{1tr} = 0.563SEK$,

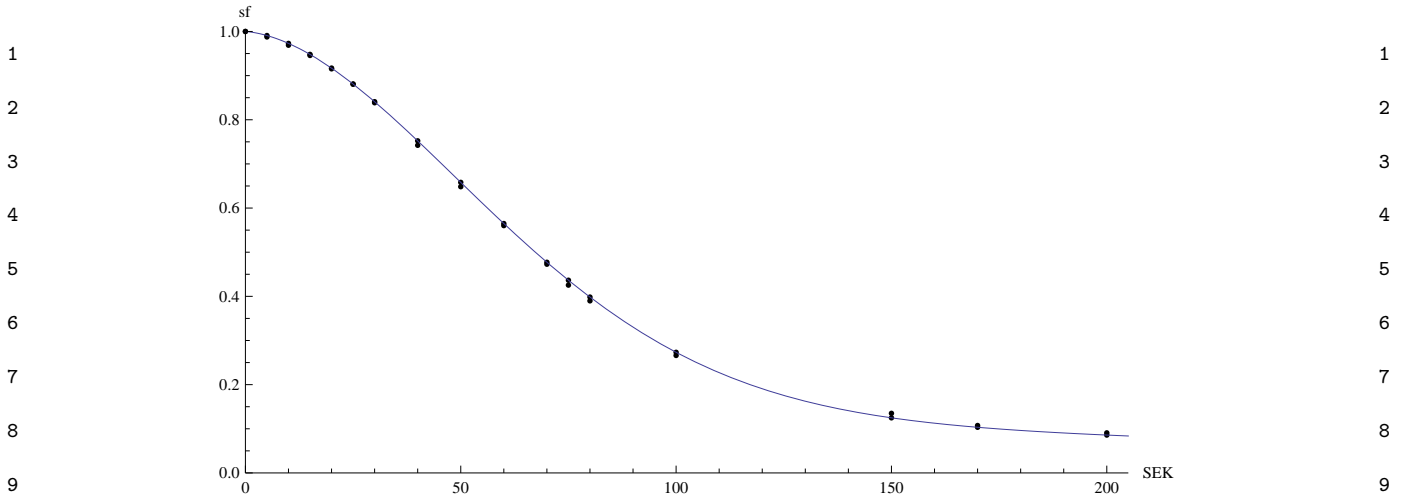


FIGURE 5.— The true WTP-survival function (5.3) is shown over interval $[0, 200]$ together with its true points $\{v_{Rj}^k, q_{trj}^k\}$ and the approximated ML-estimated points $\{v_{Rj}^k, \check{q}_j^{(10)k}\}$, $1 \leq j \leq 16$.

while the bias $mm_{1tr}^{(10)} - m_{1tr} = 3.120SEK$. Of course in reality we do not know the bias and these deviations. The bias would be smaller if, in the second step, the two subintervals $(100, 120]$ and $(120, 150]$ were used as the division intervals, rather than $v_{15} = (100, 150]$. Note that the estimate $\check{m}_1^{(10)}$ is the area below the broken line joining neighbor points $\{v_{Rj}^k, \check{q}_j^{(10)k}\}$, $j = 1, \dots, k$, $k = 23$. The broken line is shown in Figure 6.

Below we consider resampling methods to evaluation accuracies of ML-estimators. In our numerical experiment we know all necessary parameters for simulation any number of data independent copies $\mathbf{d2}_{n,2}^c$, $c = 1, 2, \dots, C$ which have the distribution identical with the considered above data $\mathbf{d2}_{n,2}^c$, $n,2 = 9000$. In our numerical experiment we simulate $C = 2000$ such i.i.d. copies. Each simulated copy contains 9000 triplets (singles or pairs). Here the total collected data may be thought of as registered answers of 18 millions of respondents.

For each copy of data $\mathbf{d2}_{n,2}^c$ we apply 5 iterations of the recursion (4.8) and

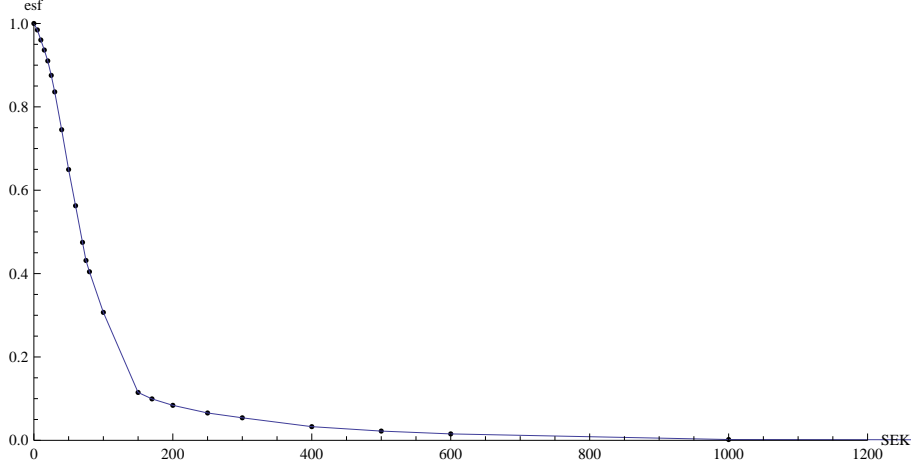


FIGURE 6.— The broken line joins by intervals the ML-estimates of division points of WTP-survival function.

obtain sufficiently accurate approximation $\check{\mathbf{q}}_k^{c(5)}$ of the ML-estimate $\check{\mathbf{q}}_{kM}^c$. Here, we use all copies of data, i.e. both pairs and singles. We can also find the ML-estimates $\check{\mathbf{q}}_{pk}^c$ of \mathbf{q}_{trk} by using (4.2) and the part of data $\mathbf{d}\mathbf{2}_{n,2}^c$ with triples containing pairs. By substituting $\check{\mathbf{q}}_{pk}^c$ and $\check{\mathbf{q}}_k^{c(5)}$ in (4.21) we obtain estimates of WTP-medium mean values $\check{m}\check{m}_{1p}^c$ and $\check{m}\check{m}_1^{c(5)}$, $c = 1, \dots, R$. To underline the gain of usage the whole data $\mathbf{d}\mathbf{2}_{n,2}^c$ compared with only the part of all pairs we calculate two lists with $C = 2000$ deviations

$$(5.4) \quad \check{d}\check{e}v_1^{c(5)} = \check{m}\check{m}_1^{c(5)} - mm_{1tr},$$

$$(5.5) \quad \check{d}\check{e}v_{1p}^c = \check{m}\check{m}_{1p}^c - mm_{1tr}, \quad c = 1, \dots, C.$$

The deviations (5.4) and (5.5) can be considered as values $\{\check{d}\check{e}v_1^{c(5)}, \check{d}\check{e}v_{1p}^c\}$ of i.i.d. two-dimensional r.v.s. Then their standard deviation estimates are

$$\check{std}_1^{(5)} = \left(\frac{1}{R-1} \sum_{c=1}^R (\check{d}\check{e}v_1^{c(5)})^2 \right)^{1/2} = 1.780,$$

$$\check{std}_{1p} = \left(\frac{1}{R-1} \sum_{c=1}^k (\check{d}\check{e}v_{1p}^c)^2 \right)^{1/2} = 2.507.$$

It is useful to know the following empirical distribution functions (e.d.f. s) of deviations (5.4) and (5.5)

$$(5.6) \quad e\check{d}f_{1C}^{(5)}[x] = \frac{1}{C} \sum_{c=1}^C I[d\check{e}v_1^{c(5)} \leq x],$$

$$(5.7) \quad e\check{d}f_{1pC}[x] = \frac{1}{C} \sum_{c=1}^C I[d\check{e}v_{1p}^c \leq x].$$

In Appendix we prove that the consistent ML-estimate of \mathbf{q}_{trk} exists and all regularity properties **RP1** - **RP5** are valid. Then we may apply resampling methods which will give us consistent estimate of characteristics of deviations (5.4) - (5.7). We simulate $C = 2000$ independent copies resampled data $\mathbf{d}\mathbf{2}_{n.2}^{*c}$, $c = 1, \dots, C$. Then we calculate corresponding estimates of WTP-medium mean values $\check{m}\check{m}_1^{*(5)}$ and $\check{m}\check{m}_{1p}^{*c}$ similarly as above using $\mathbf{d}\mathbf{2}_{n.2}^{*c}$ instead of the true data $\mathbf{d}\mathbf{2}_{n.2}$ and five iterations by the recursion (4.14). Here we also use the approximated ML-estimate $\check{m}\check{m}_1^{(5)}$ based on true data $\mathbf{d}\mathbf{2}_{n.2}$. The deviations $\check{d}\check{e}v_1^{*(5)}$ and $\check{d}\check{e}v_{1p}^{*c}$ from the ML-estimates of $\check{m}\check{m}_{1tr}$ based on true whole data $\mathbf{d}\mathbf{2}_{n.2}$ and on its part with pairs are

$$(5.8) \quad \check{d}\check{e}v_1^{*(5)} = \check{m}\check{m}_1^{*(5)} - \check{m}\check{m}_1^{(5)},$$

$$(5.9) \quad \check{d}\check{e}v_{1p}^{*c} = \check{m}\check{m}_{1p}^{*c} - \check{m}\check{m}_{1p}.$$

These deviations are used in the following e.d.f.s

$$(5.10) \quad e\check{d}f_{1C}^{*(5)}[x] = \frac{1}{C} \sum_{c=1}^C I[\check{d}\check{e}v_1^{*(5)} \leq x],$$

$$(5.11) \quad e\check{d}f_{1pC}^*[x] = \frac{1}{C} \sum_{c=1}^C I[\check{d}\check{e}v_{1p}^{*c} \leq x].$$

These e.d.f.s (5.10) and (5.11) correspond to the usage of all triples in $\mathbf{d}\mathbf{2}_{n.2}^{*c}$ and only triples with pairs in $\mathbf{d}\mathbf{2}_{n.2}^{*c}$, respectively.

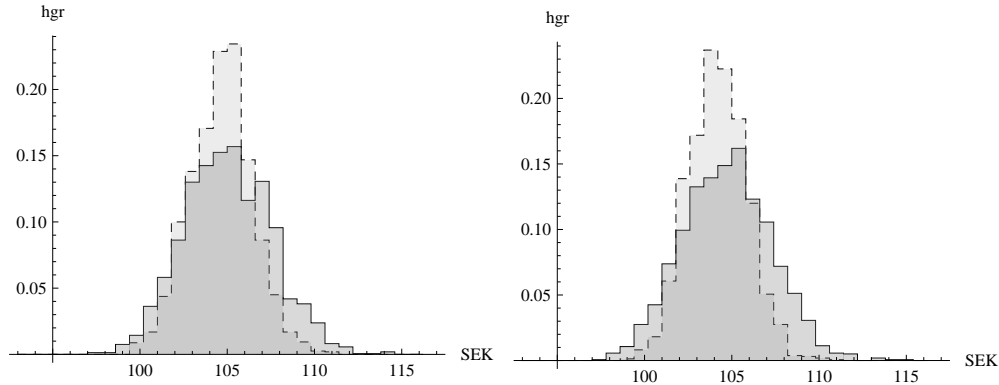


FIGURE 7.— Two pairs of histograms, corresponding to $C = 2000$ deviations (5.4), (5.5) and (5.8), (5.9), are given in the left and right plots, respectively. Dashed contour lines are used for deviations (5.4) and (5.9) based on triples with both pairs and singles, otherwise for triples with only pairs solid contour lines are used.

The lists with deviations (5.4), (5.5) and (5.8), (5.9) can be presented as histograms, composed from rectangles with low edges intervals $I_r = (2r, 2(r + 1)]$, $r = 0, 1, \dots$ on the x -axis. The length of I_r is $2SEK$. The rectangles side intervals are proportional to the numbers of $\check{m}m_1^{c(5)} \in I_r$ and $\check{m}m_{1p}^c \in I_r$, $c = 1, \dots, C$. Two corresponding histograms are in the left plot in Figure 7. The histograms corresponding to deviations (5.8) and (5.9) are in the right plot of Figure 7.

We may visualize these e.d.f.s (5.6) and (5.7) in a QQ -plot with Normal quantiles. QQ -plots are common in statistical applications because they give essential information about deviations of e.d.f.s of interest from corresponding limit distributions as size data is growing. The left QQ -plot in Figure 8 contains two empirical distribution functions (5.6) and (5.7) (solid and dashed lines, respectively). We see that the empirical distributions (5.6) and (5.7), shown in the left QQ -plot in Figure 8 are close to Normal. The

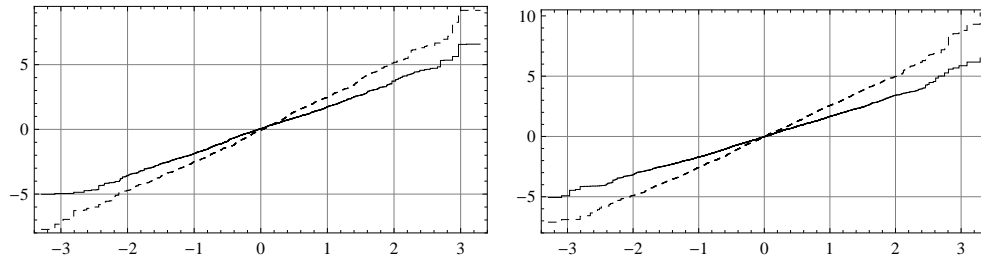


FIGURE 8.— Two QQ -plots, with Normal quantiles, contain lines with empirical distribution functions of deviations of (5.6) and (5.7) in the left QQ -plot and (5.10) and (5.11) in the right QQ -plot. Solid lines correspond to usage copies of data with both pairs and singles. Dashed lines correspond to usage only copies of triples with pairs.

deviations, in the case when both triples with singles and pairs are used, are essentially less than if only triplets with pairs are used.

In the concluding Figure 9 we display two e.d.f.s (5.7) and (5.10). In a real application, the dashed line is not known and an estimate of it (solid line) is the main aim of the suggested approach based on resampling methods. Figure 9 illustrates the consistency of the suggested estimators.

6. SUMMARY AND CONCLUSION

We have proposed a natural extension of the self-selected intervals approach to information elicitation in surveys. Because a point is a special case of an interval, self-selected intervals are very general and an interesting candidate in many types of surveys. Interest is also on the rise, as witnessed e.g. by recent changes towards the use of self-selected intervals in the Michigan Health Survey. While a single self-selected interval provides much useful information about respondent sentiments, our analysis suggest that identification of non-parametric estimators is difficult, if not impossible. We solve this identification problem by adding more information in a second step.

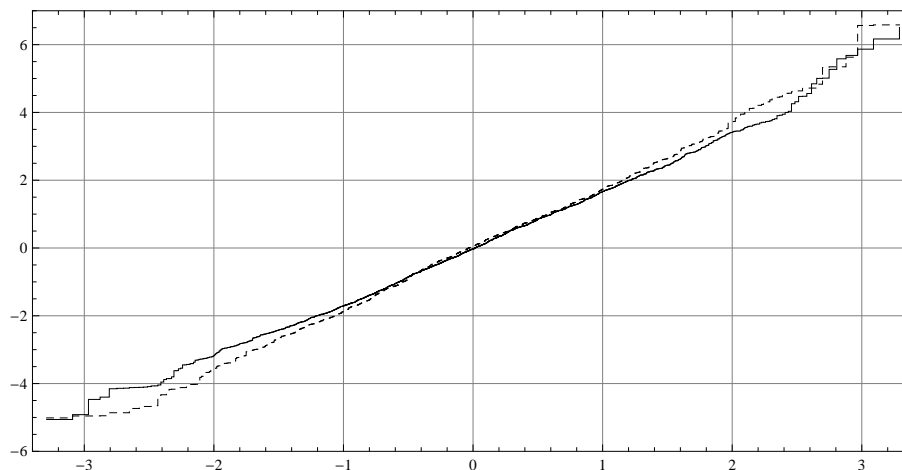


FIGURE 9.— This QQ -plot shows how two distributions of the ML-estimates of medium mean WTP-values deviates from their average values. The dashed line corresponds the true distribution of the deviates based on $C = 2000$ simulated copies of the true data $\mathbf{d2}_{n.2}^c$ and on usage of the approximated ML-estimates of WTP-medium mean values based on WTP-probabilities $\{\tilde{q}_1^{(5)}, \dots, \tilde{q}_k^{(5)}\}$. The solid line corresponds to the distribution of the similar deviates based on $C = 2000$ resampled copies $\mathbf{d2}_{n.2}^{*c}$ of the list with original true data $\mathbf{d2}_{n.2}$.

The granulation is based on a certain kind of division, where the division of the interval is based on the data. We also propose a stopping rule for sampling of respondents, which is quite practical. The investigator is to suggest a coverage probability, say 0.95, which when reached means that there is a 5% chance that the next sampled self-selected interval will be different from the ones already stated. Theorem 1 proves that the proposed non-parametric MLE is consistent, Theorem 2 shows that the accuracy of the estimator can be consistently estimated by resampling. Thus, our proposed statistical model has a number of attractive properties.

The R-programs that have been developed should be useful in applying the suggested approach. The programs will be available on CRAN. Inter-

ested readers can also use an initially developed Mathematica version of the programs.

Our approach is currently silent about how to add background variables (barring a simple stratification of the data). This is work in progress. Overall, we would be very grateful for any comments or suggestions.

7. APPENDIX

We consider an approach to obtaining consistent estimators of the probabilities $q_{trj} = P[X_i \in \mathbf{v}_j]$, $\mathbf{v}_j \in \mathcal{V}_k$, $j = 1, \dots, k$, by maximizing the log likelihood function (4.7). The vector, of true parameters $\mathbf{q}_{trk} = \{q_{tr1}, \dots, q_{trk}\}$, is a point in the $(k-1)$ -dimensional polyhedron \mathcal{S}_{k-1} defined in (4.4). Remind that we consider the case with all $q_{trj} > 0$, $j = 1, \dots, k$. The set

$$(7.1) \quad \check{\mathcal{S}}_{k-1} = \{\mathbf{q}_k = \{q_1, \dots, q_k\} : 0 < q_j < 1, \sum_{j=1}^k q_j = 1\}$$

is the open convex set of all inner points of \mathcal{S}_{k-1} . The set with boundary points $\partial\mathcal{S}_{k-1} = \mathcal{S}_{k-1} \setminus \check{\mathcal{S}}_{k-1} = \{\mathbf{q}_k = \{q_1, \dots, q_k\} : \min_{1 \leq j \leq k} q_j = 0, \sum_{j=1}^k q_j = 1\}$. We say that $n_{.2}$ is sufficiently large if in the collected on the second step data $\mathbf{d}\mathbf{2}_{n_{.2}}$ all $t_{pj} > 0$, $t_{sh} > 0$, $j = 1, \dots, k$, $h = 1, \dots, m$.

THEOREM 7.1 *For every sufficiently large $n_{.2}$ the log likelihood (4.7) is concave on $\check{\mathcal{S}}_{k-1}$.*

Proof. Let us take any two different points $\mathbf{q}_{ki} \in \check{\mathcal{S}}_{k-1}$, $i = 1, 2$. Then all points

$$\mathbf{q}_k[t] = (1-t)\mathbf{q}_{k1} + t\mathbf{q}_{k2} \in \check{\mathcal{S}}_{k-1}, \quad 0 \leq t \leq 1,$$

constitute the interval $\mathcal{J}[\mathbf{q}_{k1}, \mathbf{q}_{k2}] \subset \check{\mathcal{S}}_{k-1}$ connecting these two points. Let $f_j[t] = \text{Log}[q_j[t]]$, $j = 1, \dots, k$, $g_h[t] = \text{Log}[\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j[t]]$, $h = 1, \dots, m$. Their second derivatives are

$$(7.2) \quad \frac{d^2}{dt^2} f_j[t] = -\frac{(q_{j2} - q_{j1})^2}{q_j[t]^2}, \quad \frac{d^2}{dt^2} g_h[t] = -\frac{(\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} (q_{j2} - q_{j1}))^2}{(\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j[t])^2}.$$

From (7.2) it follows that the second sum in (4.7) has negative second derivative

$$(7.3) \quad \frac{d^2}{dt^2} \left(\frac{n_{p2}}{n_{\cdot 2}} \sum_{j=1}^k \frac{c_{pj}}{n_{p2}} \text{Log}[q_j[t]] \right) < 0, \quad 0 \leq t \leq 1.$$

Similarly, we obtain that the second derivative, of the first sum in (4.7), is non-positive

$$(7.4) \quad \frac{d^2}{dt^2} \left(\frac{n_{s2}}{n_{\cdot 2}} \sum_{h=1}^m \frac{t_{sh}}{n_{s2}} \text{Log} \left[\sum_{j \in \mathcal{E}_h} \hat{w}_{hj} q_j[t] \right] \right) \leq 0, \quad 0 \leq t \leq 1.$$

The second derivative of the last sum in (4.7) is zero. The second derivative of the log likelihood (4.7) is the sum of (7.3) and (7.4). Therefore, the second derivative is negative along intervals with any terminal points $\mathbf{q}_{k1}, \mathbf{q}_{k2}$ in $\check{\mathfrak{S}}_{k-1}$ defined in (7.1). Hence, the log likelihood (4.7) is concave on $\check{\mathfrak{S}}_{k-1}$. \square

COROLLARY 7.1 *The log likelihood (4.7) attains its maximum at a point $\check{\mathbf{q}}_{kM} = \{\check{q}_{1M}, \dots, \check{q}_{kM}\} \in \check{\mathfrak{S}}_{k-1}$ and $\check{\mathbf{q}}_{kM}$ is the unique stationary point of log likelihood (4.7) in $\check{\mathfrak{S}}_{k-1}$.*

Proof. Let us write in short $\text{lik}[\mathbf{q}_k] = \text{lik}[\mathbf{q}_k \mid \mathbb{W}_{mk}, \mathbf{d}_{2n_{\cdot 2}}], \mathbf{q}_k = \{q_1, \dots, q_k\}$. The function $\text{lik}[\mathbf{q}_k] < 0$ at any \mathbf{q}_k in $\check{\mathfrak{S}}_{k-1}$. Let \mathbf{q}_{k0} be a point in $\check{\mathfrak{S}}_{k-1}$, $l_0 = \text{lik}[\mathbf{q}_{k0}]$ and $Q(\mathbf{q}_{k0}) = \{\mathbf{q}_k : \text{lik}[\mathbf{q}_k] \geq l_0, \mathbf{q}_k \in \check{\mathfrak{S}}_{k-1}\}$. From concavity of $\text{lik}[\mathbf{q}_k]$ function on $\check{\mathfrak{S}}_{k-1}$ it follows that $Q(\mathbf{q}_{k0})$ is convex and compact. The function $\text{lik}[\mathbf{q}_k]$ is continuous on $Q(\mathbf{q}_{k0})$. If $\mathbf{q}_k \in \check{\mathfrak{S}}_{k-1} \setminus Q(\mathbf{q}_{k0})$ then $\text{lik}[\mathbf{q}_k] < l_0$. Hence, the global maximum of $\text{lik}[\mathbf{q}_k]$ on $\check{\mathfrak{S}}_{k-1}$ is attained on $Q(\mathbf{q}_{k0})$. From concavity of $\text{lik}[\mathbf{q}_k]$ on $Q(\mathbf{q}_{k0})$ it follows that the point $\check{\mathbf{q}}_{kM} = \check{\mathbf{q}}_{kM}(n_{\cdot 2})$, where $\text{lik}[\check{\mathbf{q}}_{kM}] = \max_{\mathbf{q}_k \in Q(\mathbf{q}_{k0})} \text{lik}[\mathbf{q}_k]$, is unique, and all first order partial derivatives $\left. \frac{\partial \text{lik}[\mathbf{q}_k]}{\partial q_j} \right|_{\mathbf{q}_k = \check{\mathbf{q}}_{kM}} = 0, j = 1, \dots, k$. Hence, $\check{\mathbf{q}}_{kM}$ is the unique stationary point of $\text{lik}[\mathbf{q}_k]$ in $\check{\mathfrak{S}}_{k-1}$. \square

Now, we want to find a convenient method for search of the stationary point $\check{\mathbf{q}}_{kM}$ corresponding to the maximum value of (4.7) in the multidimensional set $\check{\mathfrak{S}}_{k-1}$. We will use the method of Lagrange multipliers, see e.g. [Brinkhuis and Tihomirov \(2005\)](#). We introduce an auxiliary parameter λ and consider the following *Lagrange function*

$$(7.5) \quad \varphi_L[\mathbf{q}_k, \lambda] = \text{llik}[\mathbf{q}_k \mid \hat{\mathbb{W}}_{mk}, \mathbf{d}_{2n_2}] + \lambda(q_1 + \dots + q_k).$$

The first order partial derivatives of (7.5) w.r.t. components in \mathbf{q}_k at a stationary point in $\check{\mathfrak{S}}_{k-1}$, are zeros

$$(7.6) \quad \frac{n_{s2}}{n_{\cdot 2}} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}} + \frac{n_{p2}}{n_{\cdot 2}} \frac{t_{pj}}{n_{p2}} \frac{1}{q_j} + \lambda = 0, \quad j = 1, \dots, k.$$

Note that

$$(7.7) \quad \sum_{h=1}^m \frac{t_{sh}}{n_{s2}} = \sum_{j=1}^k \frac{t_{pj}}{n_{p2}} = 1,$$

$$(7.8) \quad \sum_{j=1}^k \left(\sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}} \right) = \sum_{h=1}^m \frac{t_{sh}}{n_{s2}} \sum_{j \in \mathcal{C}_h} \left(\frac{\hat{w}_{hj} q_j}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}} \right) = 1,$$

Then from (7.7), (7.8) by taking sum of (7.6), multiplied by q_j , over $j = 1, \dots, k$, it follows that $\lambda = -1$. Hence, each of k equalities (7.6) can be rewritten as follows

$$(7.9) \quad q_j = \frac{n_{p2}}{n_{\cdot 2}} \frac{t_{pj}}{n_{p2}} + \frac{n_{s2}}{n_{\cdot 2}} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}}, \quad j = 1, \dots, k.$$

If q_1, \dots, q_k are solutions of (7.9) then all first order partial derivatives $\frac{\partial \text{llik}[\mathbf{q}_k]}{\partial q_j} = 0, j = 1, \dots, k$, and $\mathbf{q}_k = \{q_1, \dots, q_k\}$ is a stationary point. From [Corollary 7.1](#) we have $\mathbf{q}_k = \check{\mathbf{q}}_{kM}$.

From (7.9) one may suggest the following recursion with $g_j^{(1)} = t_{pj}/n_{pj}$

$$(7.10) \quad q_j^{(r+1)} = \frac{n_{p2}}{n_{\cdot 2}} q_j^{(1)} + \frac{n_{s2}}{n_{\cdot 2}} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j^{(r)}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}^{(r)}}, \quad r = 1, 2, \dots$$

COROLLARY 7.2 For any sufficiently large $n_{\cdot 2}$ recursion (7.10) generate the sequence of points $\mathbf{q}_k^{(r)} = \{q_1^{(r)}, \dots, q_k^{(r)}\} \in \check{\mathfrak{S}}_{k-1}, r = 1, 2, \dots$, which converge to the unique stationary point $\check{\mathbf{q}}_{kM} = \{\check{q}_{1M}, \dots, \check{q}_{kM}\}$.

Proof. We can take as a start point $\mathbf{q}_k^{(1)}$ with $q_j^{(1)} = \frac{t_{pj}}{n_{p2}}, j = 1, \dots, k$. We have

$$(7.11) \quad \begin{aligned} q_j^{(1)} &\rightarrow q_{trj} > 0, \quad \hat{w}_{hj} \rightarrow w_{hj} = P[H_i = h \mid X_i = x \in \mathbf{v}_j], \\ \sum_{j' \in \mathcal{C}_h} \hat{w}_{hj} q_j^{(1)} &\rightarrow w_{trh} = \sum_{j \in \mathcal{C}_h} w_{hj} q_{trj} > 0, \quad \text{a.s.} \end{aligned}$$

as $n_{\cdot 2} \rightarrow \infty$. From (7.10) we have $q_j^{(r)} \geq \frac{n_{p2}}{n_{\cdot 2}} q_j^{(1)} > \frac{1}{2} \frac{n_{p2}}{n_{\cdot 2}} q_{trj} > 0$ a.s. for all sufficiently large $n_{\cdot 2}$. From (7.7), (7.8), (7.11) and (7.10) it follows

$$(7.12) \quad \sum_{j=1}^k q_j^{(r+1)} = \frac{n_{p2}}{n_{\cdot 2}} \sum_{j=1}^k q_j^{(1)} + \frac{n_{s2}}{n_{\cdot 2}} \sum_{j=1}^k \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j^{(r)}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_j^{(r)}} = 1.$$

If in (7.12) we start with $r = 1$, then $\sum_{j=1}^k q_j^{(2)} = 1$. We can apply the same argument with $r = 2$ in (7.12). Then we obtain $\sum_{j=1}^k q_j^{(3)} = 1$. Similarly we have $\sum_{j=1}^k q_j^{(r+1)} = 1$ if $\sum_{j=1}^k q_j^{(r)} = 1$. From (7.11) it follows that for any sufficiently large $n_{\cdot 2}$, and $r \geq 1$, $\mathbf{q}_k^{(r)} \in \mathfrak{S}_{k-1}(\mathbf{q}_{trk}) = \{\mathbf{q}_k : q_j \geq \frac{1}{2} q_{trj}, \sum_{j=1}^k q_j = 1\}$ which is a compact set. Hence, there exists a limit point $\mathbf{q}_k^{(\infty)}$, $q_j^{(r)} \rightarrow q_j^{(\infty)}$, and $q_k^{(r+1)} - q_k^{(r)} \rightarrow 0, j = 1, \dots, k$, a.s. as $r \rightarrow \infty$. The partial derivatives $\frac{\partial \text{lik}[\mathbf{q}_k]}{\partial q_j}, j = 1, \dots, k$, are uniformly continuous on $\mathfrak{S}_{k-1}(\mathbf{q}_{trk})$. Then from (7.10) if $n_{\cdot 2}$ is sufficiently large we have that $\frac{\partial \text{lik}[\mathbf{q}_k^{(r)}]}{\partial q_j} \rightarrow 0$ as $r \rightarrow \infty$. It follows that the limit point $\mathbf{q}_k^{(\infty)}$ is a stationary point. Due to concavity of (4.7) the stationary point $\mathbf{q}_k^{(\infty)}$ is unique, Therefore, from Corollary 7.1 it follows $\mathbf{q}_k^{(\infty)} = \check{\mathbf{q}}_{kM}$ and we have $q_j^{(r)} \rightarrow \check{q}_{jM}, j = 1, \dots, k$, as $r \rightarrow \infty$, for any sufficiently large $n_{\cdot 2}$. \square

If $n_{\cdot 2} \rightarrow \infty$ then all below following estimates a.s. converge to their limits

$$(7.13) \quad \begin{aligned} \frac{n_{p2}}{n_{\cdot 2}} &\rightarrow \alpha_{p2} > 0, \quad \frac{n_{s2}}{n_{\cdot 2}} \rightarrow \alpha_{s2} > 0, \quad \alpha_{p2} + \alpha_{s2} = 1, \\ \frac{t_{sh}}{n_{s2}} &\rightarrow w_{trh}, \quad \frac{t_{pj}}{n_{p2}} \rightarrow q_{trj}, \quad \frac{t_{phj}}{n_{p2}} \rightarrow w_{hj} q_{trj}, \end{aligned}$$

$j = 1, \dots, k, h \in \mathcal{D}_j$. Then the log likelihood (4.7) on compact $\mathcal{S}_{k-1}(\mathbf{q}_{trk})$ uniformly converges, as $n_{.2} \rightarrow \infty$, to the non-random function of \mathbf{q}_k

$$(7.14) \quad \text{llik}^{(\infty)}[\mathbf{q}_k] = \alpha_{p2} \sum_{j=1}^k q_{trj} \text{Log}[q_j] + \alpha_{s2} \sum_{h=1}^m w_{trh} \text{Log} \left[\sum_{j \in \mathcal{C}_h} w_{hj} q_j \right] + C_p,$$

where the constant $C_p = \alpha_{p2} \sum_{j=1}^k \sum_{j \in \mathcal{D}_j} q_{trj} \text{Log}[w_{hj}]$. In short we call $\text{llik}^{(\infty)}[\mathbf{q}_k]$ as a *limit log likelihood*. The limit of concave functions (4.7) is a concave function. Hence, the limit log likelihood (7.14) is concave. As above, by using the corresponding to (7.14) Lagrange function $\varphi_L^{(\infty)}(\mathbf{q}_k, \lambda) = \text{llik}^{(\infty)}[\mathbf{q}_k] + \lambda(q_1 + \dots + q_k)$, we find that the similar to (7.9) equations at the unique stationary point are

$$(7.15) \quad q_j = \alpha_{p2} q_{trj} + \alpha_{s2} \sum_{h \in \mathcal{D}_j} w_{trh} \frac{w_{hj} q_j}{\sum_{j' \in \mathcal{C}_h} w_{hj'} q_{j'}}, \quad j = 1, \dots, k.$$

From (4.1) we obtain that the true probabilities $q_{trj}, j = 1, \dots, k$ satisfy equations (7.15). Hence, \mathbf{q}_{trk} is the unique stationary point of the limit log likelihood where

$$(7.16) \quad \max_{\mathbf{q}_k \in \mathcal{S}_{k-1}(\mathbf{q}_{trk})} \text{llik}^{(\infty)}[\mathbf{q}_k] = \text{llik}^{(\infty)}[\mathbf{q}_{trk}].$$

COROLLARY 7.3 *Estimates $\check{\mathbf{q}}_{kM} = \check{\mathbf{q}}_{kM}(n_{.2})$ are strongly consistent estimators of \mathbf{q}_{trk} as $n_{.2}$ is unboundedly growing.*

Proof. The concave functions (4.7) uniformly converge on the compact set $\mathcal{S}_{k-1}(\mathbf{q}_{trk})$ to the limit log likelihood (7.14) which has negative second order derivatives at \mathbf{q}_{trk} as $n_{.2} \rightarrow \infty$. Then it follows from (7.13) and (7.16) that the sequence of stationary points $\check{\mathbf{q}}_{kM}(n_{.2}) \rightarrow \mathbf{q}_{trk}$. \square

Corollary 7.2 and 7.3 justify the possibility to use the sequence of $\mathbf{q}_k^{(r)}, r = 1, 2, 3, \dots$, obtained by recursion (7.9), as consistent estimates of \mathbf{q}_{trk} .

Let us show that the regularity properties **RP1 - RP5** stated in Section 4 are valid. Let as in Section 4 $\{H_i, J_i\}, i = 1, \dots, n_{.2}$, be independent pairs of

r.v.s: $H_i = h$ if the i th respondent stated interval $\mathbf{u}_h \in \mathcal{U}_m$, and $J_i = j$ if its WTP-point $X_i = x_i \in \mathbf{v}_j \in \mathcal{C}_h$. Let $\tilde{\mathbf{q}}_k = \{q_1, \dots, q_{k-1}, q_k\}$, $q_k = 1 - \sum_{j=1}^{k-1} q_j$. We introduce the following r.v.s

$$(7.17) \quad \text{LL}_{is}[\tilde{\mathbf{q}}_k] = \frac{1}{n.2} \text{Log} \left[\sum_{j \in \mathcal{C}_{H_i}} w_{H_i j} q_j \right] \quad \text{if } \mathbf{z}_i = \{i, \mathbf{u}_{H_i}, NA\},$$

and

$$(7.18) \quad \text{LL}_{ip}[\tilde{\mathbf{q}}_k] = \frac{1}{n.2} \text{Log}[w_{H_i J_i} q_{J_i}] \quad \text{if } \mathbf{z}_i = \{i, \mathbf{u}_{H_i}, \mathbf{v}_{J_i}\},$$

$i = 1, \dots, n.2$, $\tilde{\mathbf{q}}_k \in \mathcal{S}_{k-1}(\mathbf{q}_{trk})$. Then the normed log likelihood (4.7) can be considered as the observed value of the sum of values of r.v.s (7.17) and (7.18).

$$(7.19) \quad \begin{aligned} \text{LLik}[\tilde{\mathbf{q}}_k \mid \hat{\mathbb{W}}_{mk}, \mathbf{d}\mathbf{2}_{n.2}] &= \frac{1}{n.2} \sum_{i=1}^{n.2} \left(\text{LL}_{is}[\tilde{\mathbf{q}}_k] I[\mathbf{Z}_i = \{i, \mathbf{u}_{H_i}, NA\}] \right. \\ &\quad \left. + \text{LL}_{ip}[\tilde{\mathbf{q}}_k] I[\mathbf{Z}_i = \{i, \mathbf{u}_{H_i}, \mathbf{v}_{J_i}\}] \right). \end{aligned}$$

The $\mathcal{Q}_{k-1}(\mathbf{q}_{trk}) = \{\mathbf{q}_{k-1} = \{q_1, \dots, q_{k-1}\} : \tilde{\mathbf{q}}_k = \{q_1, \dots, q_{k-1}, 1 - q_1 - \dots - q_{k-1}\} \in \mathcal{S}_{k-1}(\mathbf{q}_{trk})\}$ is a compact set in \mathbb{R}^{k-1} . We denote by $\mathcal{C}_h \setminus k$ the exclusion of index k from the set $\mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}$. For each $h = 1, \dots, m$, $\max_{j \in \mathcal{C}_h} w_{hj} > 0$, because if for some h all $w_{hj} = 0$, $j \in \mathcal{C}_h$, then $w_h = 0$ and the probability of \mathbf{u}_h is zero. Then it would contradict to the observed \mathbf{u}_h on the first step of data collection. Hence, for each $h = 1, \dots, m$ and each $\tilde{\mathbf{q}}_k \in \mathcal{S}_{k-1}(\mathbf{q}_{trk})$ we have $\sum_{j \in \mathcal{C}_h} w_{hj} q_j > 0$. Then $\min_{\tilde{\mathbf{q}}_k \in \mathcal{S}_{k-1}(\mathbf{q}_{trk})} \sum_{j \in \mathcal{C}_h} w_{hj} q_j > 0$.

The first and second order partial derivatives, of the terms of log likelihood (7.19) corresponding to singles and pairs, are the following bounded and continuous functions on $\mathcal{S}_{k-1}(\mathbf{q}_{trk})$

$$(7.20) \quad \frac{\partial \text{LL}_{is}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} = \frac{w_{H_i j_1} I[j_1 \in \mathcal{C}_{H_i} \setminus k] - w_{H_i k} I[k \in \mathcal{C}_{H_i}]}{\sum_{j \in \mathcal{C}_{H_i}} w_{H_i j} q_j},$$

$$(7.21) \quad \frac{\partial \text{LL}_{ip}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} = \frac{I[J_i = j_1]}{q_{j_1}} - \frac{I[J_i = k]}{q_k}, \quad 1 \leq j_1 \leq k-1.$$

From (7.20) and (7.21) it follows that

$$(7.22) \quad \frac{\partial^2 \text{LL}_{is}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1} \partial q_{j_2}} = - \left(\frac{\partial \text{LL}_{is}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} \right) \left(\frac{\partial \text{LL}_{is}[\tilde{\mathbf{q}}_k]}{q_{j_2}} \right), \quad 1 \leq j_1, j_2 \leq k-1,$$

$$(7.23) \quad \begin{aligned} \frac{\partial^2 \text{LL}_{ip}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1} \partial q_{j_2}} &= - \left(\frac{I[J_i = j_1 \in \mathcal{C}_{H_i}]}{q_j^2} I[j_1 = j_2] + \frac{I[J_i = k \in \mathcal{C}_{H_i}]}{q_k^2} \right) \\ &= - \frac{\partial \text{LL}_{ip}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1}} \frac{\partial \text{LL}_{ip}[\tilde{\mathbf{q}}_k]}{\partial q_{j_2}} \quad 1 \leq j_1, j_2 \leq k-1. \end{aligned}$$

Hence, all first and second order partial derivatives, defined in (7.20) - (7.23), are continuous and bounded functions, and the regularity property **RP1** holds.

On $Q_{k-1}(\mathbf{q}_{trk})$ the maxima of the partial derivatives (7.20) - (7.23) are finite numbers. Hence, the above derivatives possess finite moments of any order. Therefore; the regularity property **RP2** holds. From boundedness, uniform continuity of all second order partial derivatives on $Q_{k-1}(\mathbf{q}_{trk})$ and from the limit Lebesgue theorem it follows that for any inner point $\mathbf{q}_{k-1} \in Q_{k-1}(\mathbf{q}_{trk})$

$$(7.24) \quad E_{\mathbf{q}_{k-1}} \left[\overline{\lim}_{\tilde{\mathbf{q}}'_{k-1} \rightarrow \tilde{\mathbf{q}}_k} \left| \frac{\partial^2 \text{LL}_{ig}[\tilde{\mathbf{q}}'_k]}{\partial q_{j_1} \partial q_{j_2}} - \frac{\partial^2 \text{LL}_{ig}[\tilde{\mathbf{q}}_k]}{\partial q_{j_1} \partial q_{j_2}} \right| > 0 \right] = 0,$$

$g = s, p$, $1 \leq j_1, j_2 \leq k-1$. Relation (7.24) corresponds to the regularity property **RP3**. To check the regularity property **RP4** we may use a reversible method. From relation $\sum_{h=1}^m \sum_{j \in \mathcal{C}_h} w_{hj} q_j = 1$, (7.18) and (7.20) we have

$$(7.25) \quad \begin{aligned} 0 &= \frac{1}{n_{\cdot 2}} \frac{\partial}{\partial q_{j_1}} \sum_{h=1}^m \sum_{j \in \mathcal{C}_h} w_{hj} q_j \\ &= \frac{1}{n_{\cdot 2}} \sum_{h=1}^m \left(\frac{w_{hj_1} - w_{hk} I[k \in \mathcal{C}_h]}{\sum_{j \in \mathcal{C}_h} w_{hj} q_j} \right) \sum_{j \in \mathcal{C}_h} w_{hj} q_j \\ &= \frac{1}{n_{\cdot 2}} \sum_{h=1}^m \frac{\partial}{\partial q_{j_1}} \text{Log} \left[\sum_{j \in \mathcal{C}_h} w_{hj} q_j \right] \sum_{j \in \mathcal{C}_h} w_{hj} q_j = E_{\tilde{\mathbf{q}}_k} \left[\frac{\partial \text{LL}_{is}[\mathbf{q}_k]}{\partial q_{j_1}} \right], \end{aligned}$$

$1 \leq j_1 \leq k-1$. One may read (7.25) from right to left. Similarly we obtain

$$(7.26) \quad E_{\tilde{\mathbf{q}}_k} \left[\frac{\partial \text{LL}_{ip}[\mathbf{q}_k]}{\partial q_{j_1}} \right] = 0, \quad 1 \leq j_1 \leq k-1.$$

From (7.22) and (7.23) we obtain

$$(7.27) \quad E_{\mathbf{q}_k} \left[\frac{\partial^2 \text{LL}_{ig}[\mathbf{q}_k]}{\partial q_{j_1} \partial q_{j_2}} \right] = -E_{\mathbf{q}_k} \left[\frac{\partial \text{LL}_{is}[\mathbf{q}_k]}{\partial q_{j_1}} \frac{\partial \text{LL}_{is}[\mathbf{q}_k]}{\partial q_{j_2}} \right], \quad g = s, p.$$

Validity of (7.25) - (7.27) means that the regularity property **RP4** holds.

Let us introduce two $(k-1) \times (k-1)$ matrices

$$(7.28) \quad \mathbb{F}_{sn.2}[\mathbf{q}_{k-1}] = -\frac{n_{s2}}{n.2} \left(E_{\mathbf{q}_{k-1}} \left[\frac{\partial^2 \text{LL}_{1s}[\mathbf{q}_{k-1}]}{\partial q_{j_1} \partial q_{j_2}} \right] \right)_{1 \leq j_1, j_2 \leq k-1},$$

$$(7.29) \quad \mathbb{F}_{pn.2}[\mathbf{q}_{k-1}] = -\frac{n_{p2}}{n.2} \left(E_{\mathbf{q}_{k-1}} \left[\frac{\partial^2 \text{LL}_{1p}[\mathbf{q}_{k-1}]}{\partial q_{j_1} \partial q_{j_2}} \right] \right)_{1 \leq j_1, j_2 \leq k-1}.$$

The Fisher Information (FI-)matrix corresponding the log likelihood (4.7) is

$$(7.30) \quad \mathbb{F}_{n.2}[\mathbf{q}_{k-1}] = \mathbb{F}_{sn.2}[\mathbf{q}_{k-1}] + \mathbb{F}_{pn.2}[\mathbf{q}_{k-1}], \quad \mathbf{q} \in Q_{k-1}(\mathbf{q}_{trk}).$$

From (7.27) - (7.29) it follows that matrix $\mathbb{F}_{sn.2}[\mathbf{q}_{k-1}]$ is non-negative and $\mathbb{F}_{pn.2}[\mathbf{q}_{k-1}]$ is positive definite with full rank $(n-1)$. Then defined in (7.30) the FI-matrix $\mathbb{F}_{n.2}[\mathbf{q}_{k-1}]$ has full rank. Hence the regularity property **RP5** also holds.

We summarize obtained results in the following

THEOREM 7.2 *The regularly properties **RP1** - **RP5** are valid for the statistical model with the log likelihood (4.7).*

Let $\mathfrak{M} = \{\mathcal{M}\}$ be a class of statistical models \mathcal{M} which possess the following properties:

PD: the set of data in the model \mathcal{M} is a list $\mathbf{d}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ which components, $\mathbf{x}_i \in \mathbb{R}^d$, are values of independent, r.v.s, and their distributions can belong to different families of parametric distributions however with the same set $\Theta = \{\boldsymbol{\theta}\} \subseteq \mathbb{R}^k$ of parameters,

PE: there exists a consistent ML-estimate $\check{\theta}_n$ of the true parameter $\theta_0 \in \Theta$;

PR: the log likelihoods of components $\mathbf{x}_i \in \mathbf{d}_n$ possess the standard regularity properties, (similar with **PR1** - **PR5** stated in Section 4 and checked above in this Appendix.)

PROPOSITION 1 *Let the properties **PD**, **PE**, **PR** be valid for a model $\mathcal{M} \in \mathfrak{M}$. Then the distribution of normed deviations $\sqrt{n}(\check{\theta}_n - \theta_0)$ approaches, as $n \rightarrow \infty$, with d -dimensional Normal distribution with the covariance matrix equal to the inverse FI-matrix.*

This result is a convenient slight generalisation of a well known classical result of R. Fisher, see [Lehmann and Casella \(1998\)](#), [Ferguson \(1996\)](#).

Let us $\mathbf{d}_n^{*c} = \{\mathbf{x}_{i_1^{*c}}, \dots, \mathbf{x}_{i_r^{*c}}\}$ be resampled copy of the original data \mathbf{d}_n . Then we can restate properties **PD**, **PE** and **PR** by using the resampled copies of data \mathbf{d}_n^{*c} instead of \mathbf{d}_n . We denote these properties by **PD**^{*c}, **PE**^{*c} and **PR**^{*c}. For example, we have

PE^{*c}: there exists a consistent ML-estimate $\check{\theta}_n^{*c}$ of the true parameter $\theta_0 \in \Theta$.

PROPOSITION 2 *Let the properties **PD**, **PE**, **PR** and **PD**^{*c}, **PE**^{*c}, **PR**^{*c} be valid for a model $\mathcal{M} \in \mathfrak{M}$. Then the distributions of $\sqrt{n}(\check{\theta}_n - \theta_0)$ and $\sqrt{n}(\check{\theta}_n^{*c} - \check{\theta}_n)$ are approaching each other in probability as $n \rightarrow \infty$.*

The Proposition 2 is valid for the considered statistical model with two-step sampling design with self-selected rounded interval data.

The Proposition 1 and 2 are proved in [Belyaev and Nilsson \(1997\)](#), see there Theorem 5 and 6, in [Nilsson \(1998\)](#) there are also several applications and detailed proofs. The asymptotic properties of resampled sums of independent r.v.s are considered in [Belyaev and Sjöstedt-de Luna \(2000\)](#); [Belyaev \(2003, 2007\)](#).

REFERENCES

- 1
2
3 BELYAEV, Y. (2003): Necessary and sufficient conditions for consistency of resampling.
4 Research report, Center of Biostochastics. Swedish University of Agricultural Sciences.
5 <http://biostochastics.slu.se/publikationer>.
6
7 BELYAEV, Y. (2007): Resampling for Lifetime Distributions. In *Encyclopedia of Statistics*
8 *in Quality and Reliability*, Ed. Rupper, F., Kennet, R., Faltin, F.W. John Wiley,
9 Chichester, UK, 1653–1657.
10
11 BELYAEV, Y. AND KRISTRÖM, B. (2010): Approach to Analysis of Self-selected Interval
12 Data. CERE Working Paper #2/2010.
13
14 BELYAEV, Y. AND KRISTRÖM, B. (2012): Analysis of contingent valuation data with self-
15 selected ounded WTP-intervals collected by two-steps sampling plans. Proceedings of
16 the 9th Conference on Multivariate Statistics. Tartu, 2011. (In print).
17
18 BELYAEV, Y. AND NILSSON, L. (1997): Parametric Maximum Likelihood Estimators
19 and Resampling. Research report, N15. Department of Mathematical Statistics.
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 BRINKHUIS, J. AND TIHOMIROV, V. (2005): *Optimization: Insights and Applications*.
25 Princeton University Press, Princeton and Oxford.
26
27 BROBERG, T. AND BRÄNNLUND, R. (2008): An alternative interpretation of multiple
28 bounded WTP data-Certainty dependent payment card intervals. *Resources Energy*
29 *Econ*, doi:10.1016/j.reseneeco.2008.09.001.
30
31 DAVISON, A.C. AND HINKLEY, D. (1997): *Bootstrap Methods and their Applications*.
32 Cambridge University Press. Cambridge.
33
34 FERGUSON, T.S. (1996): *A Course in Large Sample Theory*. Chapman & Hall, London.
35
36 FINK, A. (1985): *How to ask survey questions*. Thousand Oaks, CA: SAGE Publications.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

- 1 survey response bias, in D. Wise (Ed.) *Frontiers in the Economics of Aging*, 353–387, 1
Chicago, IL: University of Chicago Press. 2
- 3 HUTTENLOCHER, J., HEDGES, L.V. AND BRADBURN, N.M. (2008): Reports of elapsed 3
time: Bounding and rounding processes in estimation. *Journal of Experimental Psy-* 4
chology: Learning, Memory, and Cognition **16**, 196–213. 4
- 5 HÅKANSSON, C. (2007): Cost-Benefit Analysis and Valuation Uncertainty. Doctoral 5
Thesis, No. 2007-41, Swedish University of Agricultural Sciences, Umeå. 6
- 7 HÅKANSSON, C. (2008): A new valuation question - Analysis of and insights from interval 7
open ended data in contingent valuation. *Environmental and Resource Economics* 8
39,2,175–188. 8
- 9 JAMMALAMADAKA, S.R. AND MANGALAM, V. (2003): Non-parametric estimation for 9
middle-censored data. *J. Nonparametr. Statist.* **15**, 253–265. 10
- 11 KLEIN, J.P. AND MOESCHBERGER, ML (1997): *Survival Analysis: Techniques for Cen-* 11
sored and Truncated Data, Springer-Verlag. 12
- 13 LEHMANN, E.L. AND CASELLA, G. (1998): *Theory of point estimation*. 2nd. ed., N. Y.: 13
Springer-Verlag. 14
- 15 MANSKI, C.F. AND F.MOLINARI (2010) ROUNDING PROBABILISTIC EXPECTATIONS IN 15
SURVEYS. *Journal of Business and Economic Statistics.* **28**, 219-231. 16
- 16 MCFADDEN, D., BEMMAOR, A., CARO, F., DOMINITZ, J., JUN, B-H., LEWBEL, A., 16
MATZKIN, R., MOLINARI, F., SCHWARZ, N., WILLIS, R. AND WINTER, J. (2005): 17
Statistical Analysis of Choice Experiments and Surveys, *Marketing Letters*, **16**, 183– 18
196. 18
- 19 NILSSON, L. (1998): Parametric Estimation Using Computer Intensive Methods. Ph. 19
Dissertation, ISBN 91-7191-438-8, Dept. of Mathematical Statistics, Umeå University. 20
- 21 ROSCH, E. (1975): Cognitive reference points. *Cognitive Psychology* **7**, 532–547. 21
- 22 SCHAEFFER, N.C. AND BRADBURN, N.M. (1989): Respondent behaviour in magnitude 22
estimation. *Journal of the American Statistical Association* **84**, 402–413. 23
- 24 TURNBULL, B.W. (1976): The empirical distribution function with arbitrarily grouped, 24
censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, pp. 290–295. 25
- 25 VAN EXEL, W.B.F., BROUWER, B., VAN DEN BERG, B. AND KOOPMANSCHAP, M.A. 25
(2006): With a little help from an anchor discussion and evidence of anchoring effects 26
in contingent valuation, *The Journal of Socio-Economics*, **35**, 836–853. 26
- 27 WENCHAO, Z., BELYAEV, Y. AND KRISTRÖM, B. (2012): Package Software for Analysis 27
of Rounded Self-Selected Interval Data (PSARSSID). Swedish University of Agricul- 28
tural Sciences. (In preparation). <http://www.cere.se>. 28
29 29