Two-step approach to analysis of statistical interval data models.

Yuri Belyaev and Bengt Kriström
Department of Mathematics and Mathematical Statistics
Umeå University
Department of Forest Economics SLU, Umeå
Centre for Environmental & Resource Economics (CERE)
Umeå, Sweden

Keywords: interval rounded data, coverage probability, likelihood, maximization, recursion, resampling, accuracy of estimable characteristics.

21 - 23 June 2011, Ulvön Conference

In many applications one can meet statistical data sets which elements are intervals. The respondent may freely self-select any interval of choice that contains the point. The notion of utility function is used, and the respondent's true wiliness-to-pay (WTP)-point is the related to the utility function as compensating variation. For the collected data, we found that presence of strong rounding is a typical feature. The self-selected intervals can be considered as censoring the true WTP-points. Usually in the Survival Analysis it is assumed that the censoring intervals are independent of such points and cover only some of the points. But here these intervals can depend on the unobserved positions of their WTP-points, and all WTP-points are covered.

Our aim: find consistent estimates related to the distribution of these WTP-points. We propose statistical models which admit dependency the self- selected WTP-intervals on the positions of their WTP-points.

Our approach is general, but we will frame the analysis within a contingent valuation experiment, where the individual is to report his valuation of a public good. This is a natural application of self-selected intervals, given the many difficulties individuals demonstrably face when trying to put a value on a public good. Indeed, introspection, common sense and a bulk of empirical evidence does suggest that individuals find it difficult to report on their willingness-to-pay (WTP) as a point. We denote the WTP-points $\{x_i\}$. Our approach is based on the three assumptions, of which we first discuss two (the third being statistical in nature):

We collect basic assumptions in the following:

Assumption 1. *Each respondent might not be aware of the exact location of the true WTP-point. The respondents may freely state 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).*

Assumption 2 is crucial in data collecting on the two-step approach, in that we assume that the second step has no impact on the distribution of WTP; the only effect of the second step is to increase information. From some perspectives, this assumption is very strong, given the evidence that exists on the so-called double bounded valuation question. In this case, the individual is to accept/reject a suggested price (for a public good) and, depending on the answer, a lower/higher price is then offered in the second round. It is clearly a delicate matter to formulate the question in the second round, because delivery might have been promised at the suggested price in the first round. We simply assume here that there is a logical way out of the second step dilemma, the most natural one being uncertainty; the cost of providing the public good might not be known with certainty.

We now turn to the statistical model and begin by explaining the third and final assumption needed.

First step of data collecting

We consider the following 1st step of *two-step plan* of data collecting. On the *first step* randomly sampled individuals will be suggested to state self-selected intervals containing their true WTP-values. The decision to stop collection of self-selected intervals can be based on the idea to estimate introduced below a coverage probability. Let n respondents have stated intervals $\mathbf{y}_1^n = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$, $\mathbf{y}_i = (y_{Li}, y_{Ri}]$ containing their WTP-points. Due to rounding the same intervals can be repeatedly stated by different respondents.

We can consider stated intervals $\mathbf{y}_1^n = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ as a realization of a multinomial random process $\{Y_i\}_{i\geq 1}$ with the discrete time parameter $i = 1, 2, ...$ The r.v.s $\{Y_i\}_{i\geq 1}$ are i.i.d. and the set of their values is a set containing finite number of all possible to be stated self-selected intervals $\mathcal{U}_{all} = \{\mathbf{u}_\alpha, \alpha \in A\}$. There is a discrete probability distribution $p_\alpha = P[Y_i = \mathbf{u}_\alpha], \alpha \in A$. The sets $A, \mathcal{U}_{all}$ and the probability distribution $\{p_\alpha, \alpha \in A\}$ are not known.

Empirical data: self-selected Intervals with WTP-values stated by respondents.

The following list $\mathbf{d1}_{n_1}$ contains the ends of all different, freely stated by $n_1 = 241$ respondents, WTP-intervals $\mathbf{u}_h = (u_{Lh}, u_{Rh}], h = 1, ..., 46,$ t_h is the number of cases when $\mathbf{u}_h$ has been stated by respondents, $\mathbf{d1}_{n_1} = \{..., \{h, \{u_{Lh}, u_{Rh}\}, t_h\}, ...\}$. The ordering of $\mathbf{u}_h$ was done by the values of left ends u_{Lh} and if they are the same then by the values of right ends u_{Rh}. Each interval $\mathbf{u}_h$ is the union of disjoint division intervals $\mathbf{v}_j = (v_{Lj}, v_{Rj}], j = 1, ..., 23.$

Let $\mathcal{U}_m = \{\mathbf{u}_1, ..., \mathbf{u}_m\}, m = 46$, and $\mathcal{V}_k = \{\mathbf{v}_1, ..., \mathbf{v}_k\}, k = 23.$

$\mathbf{d1}_{n_1} = \{\{1, \{0, 20\}, 2\}, \{2, \{0, 50\}, 3\}, \{3, \{0, 100\}, 4\}, \{4, \{0, 200\}, 1\},$
$\{5, \{5, 10\}, 1\}, \{6, \{5, 50\}, 1\}, \{7, \{10, 15\}, 1\}, \{8, \{10, 20\}, 5\}, \{9, \{10, 30\}, 5\},$
$\{10, \{10, 40\}, 1\}, \{11, \{10, 50\}, 6\}, \{12, \{10, 100\}, 2\}, \{13, \{20, 30\}, 3\},$
$\{14, \{20, 40\}, 5\}, \{15, \{20, 50\}, 39\}, \{16, \{20, 60\}, 1\}, \{17, \{20, 70\}, 1\},$
$\{18, \{20, 80\}, 1\}, \{19, \{20, 100\}, 11\}, \{20, \{20, 200\}, 1\}, \{21, \{25, 50\}, 3\},$
$\{22, \{30, 50\}, 1\}, \{23, \{30, 70\}, 1\}, \{24, \{30, 75\}, 1\}, \{25, \{50, 70\}, 2\},$
$\{26, \{50, 75\}, 3\}, \{27, \{50, 100\}, 69\}, \{28, \{50, 150\}, 3\}, \{29, \{50, 200\}, 3\},$
$\{30, \{50, 500\}, 1\}, \{31, \{75, 100\}, 1\}, \{32, \{100, 150\}, 8\}, \{33, \{100, 170\}, 1\},$
$\{34, \{100, 200\}, 23\}, \{35, \{100, 250\}, 3\}, \{36, \{100, 300\}, 4\}, \{37, \{100, 500\}, 5\},$
$\{38, \{100, 1000\}, 1\}, \{39, \{150, 200\}, 2\}, \{40, \{200, 300\}, 3\}, \{41, \{200, 400\}, 1\},$
$\{42, \{300, 500\}, 1\}, \{43, \{300, 600\}, 1\}, \{44, \{400, 500\}, 1\}, \{45, \{500, 1000\}, 4\},$
$\{46, \{500, 2000\}, 1\}\};$

The following list contains ends of division intervals $\mathbf{v}_j = (v_{Lj}, v_{Rj}]$

$\mathbf{dv} = \{\{0, 5\}, \{5, 10\}, \{10, 15\}, \{15, 20\}, \{20, 25\}, \{25, 30\}, \{30, 40\},$
$\{40, 50\}, \{50, 60\}, \{60, 70\}, \{70, 75\}, \{75, 80\}, \{80, 100\}, \{100, 150\},$
$\{150, 170\}, \{170, 200\}, \{200, 250\}, \{250, 300\}, \{300, 400\}, \{400, 500\},$
$\{500, 600\}, \{600, 1000\}, \{1000, 2000\}\}.$

The sequence of random numbers h of registered self-selected intervals collected in $\mathbf{d1}_{n_1}$, by $n_1 = 241$ respondents. Below the first-fourth respondents have stated $\mathbf{u}_{27}$, the fifth respondent stated $\mathbf{u}_{28}$ etc.

{ 27, 27, 27, 27, 28, 29, 27, 15, 27, 27, 44, 15, 27, 27, 27, 15, 27, 26, 15, 11, 26, 27, 3, 35, 34, 27, 34, 15, 23, 1, 34, 28, 27, 27, 32, 27, 14, 9, 27, 15, 34, 45, 33, 34, 6, 34, 21, 34, 15, 27, 34, 14, 43, 15, 40, 15, 15, 37, 27, 27, 34, 27, 27, 8, 36, 42, 15, 27, 3, 32, 34, 15, 19, 29, 34, 7, 35, 9, 37, 27, 8, 27, 15, 15, 15, 27, 11, 27, 27, 27, 41, 40, 15, 30, 9, 32, 27, 19, 45, 34, 27, 11, 15, 27, 13, 39, 45, 10, 34, 27, 14, 8, 32, 27, 27, 27, 27, 15, 29, 27, 27, 15, 34, 27, 19, 3, 15, 27, 27, 15, 35, 15, 19, 19, 25, 15, 19, 18, 24, 27, 2, 14, 34, 2, 34, 15, 21, 27, 27, 34, 27, 32, 17, 28, 38, 11, 19, 27, 27, 12, 46, 27, 1, 36, 15, 27, 15, 27, 8, 34, 20, 34, 21, 27, 37, 25, 15, 37, 19, 27, 27, 27, 27, 13, 27, 11, 27, 22, 34, 15, 19, 34, 40, 27, 15, 15, 19, 4, 27, 36, 27, 9, 15, 12, 34, 27, 14, 26, 37, 27, 15, 15, 3, 32, 8, 15, 19, 27, 32, 32, 15, 39, 15, 5, 27, 27, 36, 13, 9, 27, 15, 16, 34, 27, 15, 27, 2, 15, 11, 31, 45}

For each i we obtain the conditional expectation of the indicator $I[h_i \text{ has } ct_{h_i} \geq 2]$ of the event that $\mathbf{u}_{h_i}$ has been already stated by $\mathbf{ct}_{h_i}$ respondents with numbers in $\{1, ..., i\}$, given sufficient statistics $\{\{h_1, ct_{h_1}\}, ..., \{h_i, ct_{h_i}\}\}$. We have

$$\hat{p}_c[i] = \frac{i - \#\{ct_{h_{i'}} = 1, i' \leq i\}}{i}. \tag{1}$$

For $i = 197, ..., 209$ we have the following values $\{i, \hat{p}_c[i]\}$

$\{\{$ 196, 177/196$\}$, $\{$ 197, 178/197$\}$, $\{$ 198, 178/198$\}$, $\{$ 199, 179/199$\}$, $\{$ 200, 180/200$\}$, $\{$ 201, 181/201$\}$, $\{$ 202, 182/202$\}$, $\{$ 203, 183/203$\}$, $\{$ 204, 185/204$\}$, $\{$ 205, 186/205$\}$, $\{$ 206,187/206$\}$, $\{$ 207, 188/207$\}$, $\{$ 208, 189/208$\}$, $\{$ 209, 190/209$\}\}$

If $i = 241, \hat{p}_c[241] = \frac{241-21}{241} = \frac{220}{241} \approx 91.29\%.$

$p_c[i]$ is an unbiased estimate of the coverage probability $p_c[i]$ that the $\mathbf{u}_{h_i}$ interval has been already stated earlier.

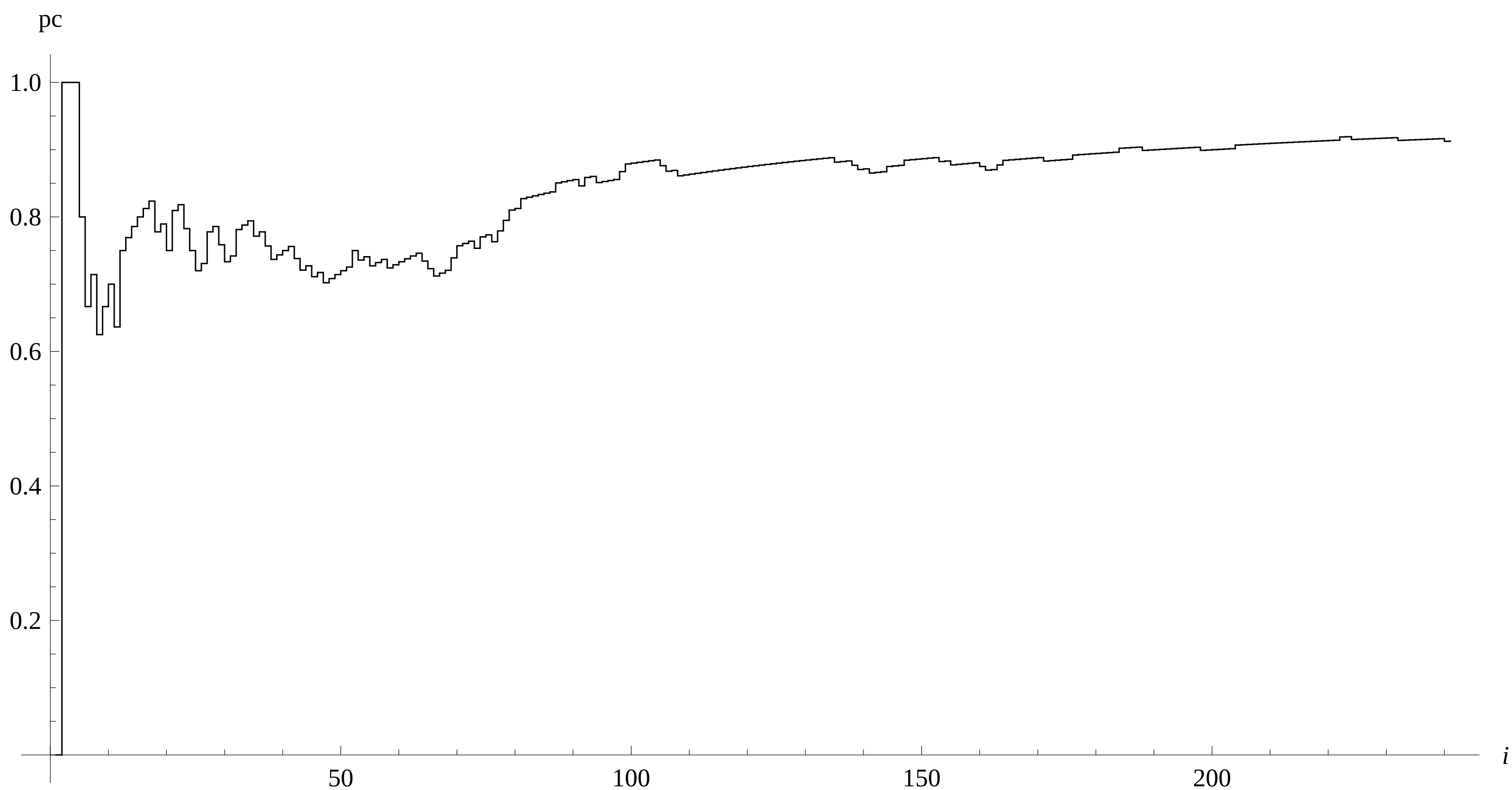

Figure 1: Dynamic of the coverage probability estimates $\hat{p}_c[i]$ as a function of i sequentially selected respondents. The coverage probability characterizes the part of respondents in the whole population who being selected would state self-selected WTP-intervals already stated by the first i selected respondents.

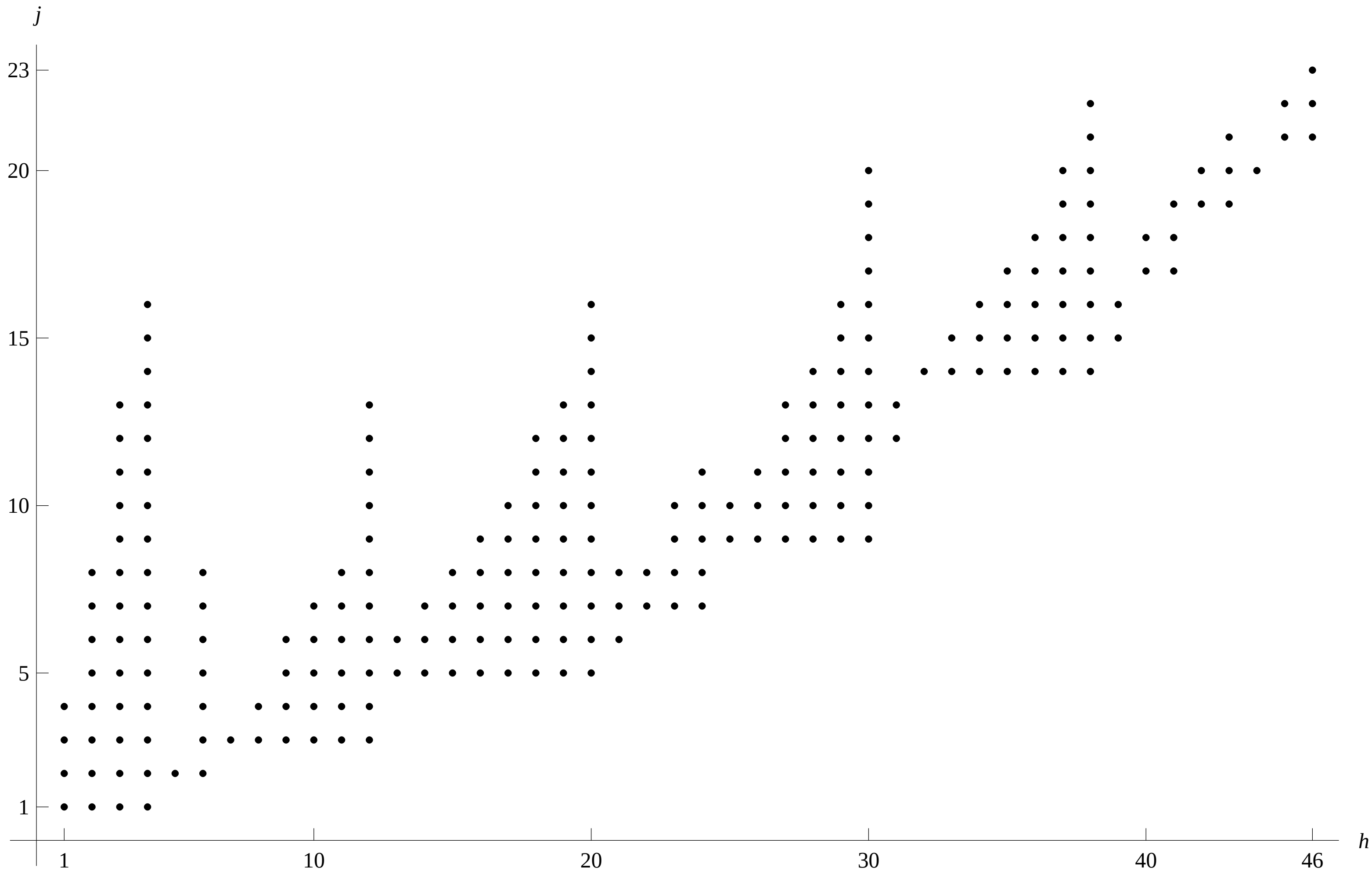

Figure 2: The set of all compatible indexes $\{\{h,j\} : h = 1,\dots,46, j \in \mathcal{C}_h\} = \{\{h,j\} : h \in \mathcal{D}_j, j = 1,\dots,23\}$. The sets $\mathcal{C}_h = j, \mathbf{v}_j \subseteq \mathbf{u}_h\}$, $\mathcal{D}_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}$ are h-cuts and j-cuts of the shown set.

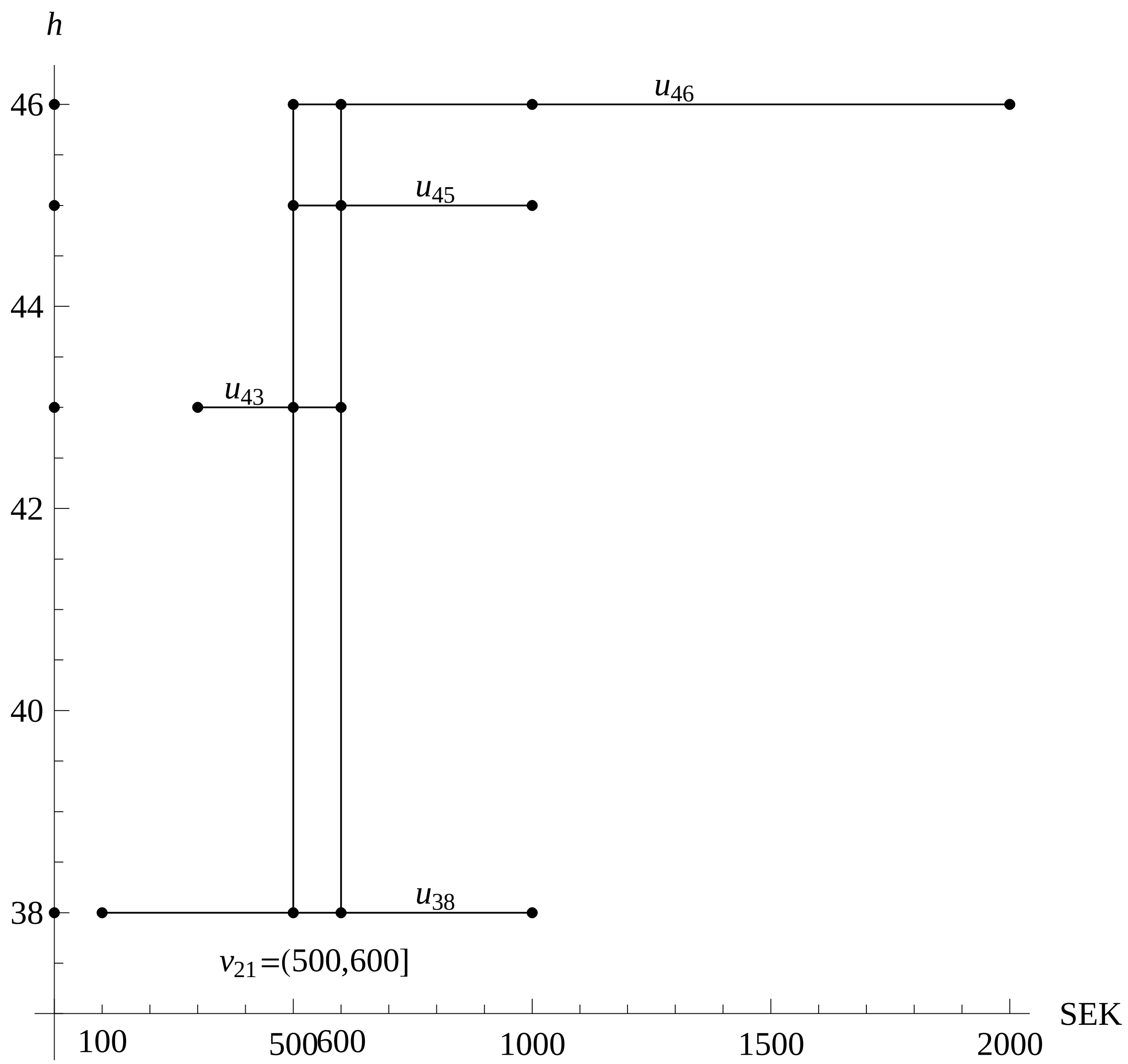

Figure 3: $D_{21} = \{38, 43, 45, 46\}$, i.e. $\mathbf{v}_{21} = (500, 600]$ is a part in $\mathbf{u}_{38} = (100, 1000]$, $\mathbf{u}_{43} = (300, 600]$, $\mathbf{u}_{45} = (500, 1000]$ and in $\mathbf{u}_{46} = (500, 2000]$. If WTP-point $X \in \mathbf{v}_{21}$, then the i-th respondents has to state $\mathbf{u}_h, h \in D_{21}$.

We need to estimate the d.f. $F[x]$, or the same, to estimate the s.f. $S[x]$ of the true WTP-distribution. Let $H_i, i = 1, ..., n_1 (= 241)$ be i.i.d. r.v.s, $H_i = h$ if the i-th respondent has stated interval $\mathbf{u}_h$ containing his/her WTP-point x_i. We use notation $w_h = P[\{H_i = h\} \cap \{X_i \in \mathbf{u}_h\}], w_h \leq p_h = P[X_i \in \mathbf{u}_h], h = 1, ..., m$. The event $\{X_i \in \mathbf{u}_h\}$ is not observable. The probability to obtain numbers of times $t_1, ..., t_h$ is $\prod_{h=1}^{m} w_h^{t_h}$, i.e the observed number of times have multinomial distribution. The corresponding normed log likelihood llik is

$$\mathrm{llik}[w_1, ..., w_m | t_1, ..., t_m] = \sum_{h=1}^{m} \frac{t_h}{n_d} \mathrm{Log}[w_h], \quad \sum_{h=1}^{m} t_h = n_d.$$

The maximum of this llik over all possible values $w_h \geq 0, \sum_{h=1}^{m} w_h = 1$, is attained at $\hat{w}_h = t_h / n_d, h = 1, ..., m$. Note that $-\sum_{h=1}^{m} \hat{w}_h \mathrm{Log}[\hat{w}_h]$ is the empirical Entropy of the multinomial distribution with probabilities $\{\hat{w}_1, ..., \hat{w}_m\}$.

There are many hypothetical behavior models of respondents, e.g.

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

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

$\mathbf{BM3}$ is the behavior model of respondents who, with $\mathbf{v}_j$ containing their WTP-point, select to state $\mathbf{u}_h, h \in \mathcal{D}_j$, proportionally to frequencies $\hat{w}_h = \frac{t_h}{n}, h = 1, ..., m$. Here

$$\hat{w}_{hj} = \frac{t_h}{\left(\sum_{h' \in \mathcal{D}_j} t_{h'}\right)}.$$

For identification the true WTP-distribution is necessary consistently estimate conditional probabilities $\mathbb{W} = (w_{hj})$ by using extended empirical data. It is not possible to identify the true WTP-distribution if we have only the data $\mathbf{d1}_{n_1}$. We obtain this possibility on the second step of data collecting.

Second step of data collection.

We have need to take a random sample of size $n_{\cdot 2}$ of new (not yet sampled) individuals from the population $\mathfrak{P}$. There are several variants of requests to the sampled individuals. In the second step each individual has the request freely to state an interval containing his/her WTP-value. If the stated self-selected interval belongs to $\mathcal{U}_m$ then we submit to this individual an additional request to select from the division $\mathcal{V}_k$ an interval $\mathbf{v}_j \in \mathcal{V}_k, \mathbf{v}_j \subseteq \mathbf{u}_h,$ containing his/her true WTP-value. The respondent has right to avoid selection of such division interval. The collected data will be $\mathbf{d2}_{n_{\cdot 2}} = \{\mathbf{z}_1, ..., \mathbf{z}_{n_{\cdot 2}}\}$, $\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, NA\}$ or $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$, NA is "no answer". In short we call these triples as *singles* and *pairs.* We suppose that $n_{\cdot 2}$ is sufficiently large and any $\mathbf{v}_j \in \mathcal{V}_k$ was stated many times.

The collected data with pairs $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ admit simple estimates of $q_{trj} = P[X_i \in \mathbf{v}_j], j = 1, ..., k.$

The ML-estimates are

$$\check{q}_{pj} = \frac{c_{pj}}{n_{p2}}, \quad j = 1, ..., k. \tag{2}$$

We can use part of data with pairs $\{\mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$. We obtain the following unbiased and consistent estimates of w_{hj}

$$\hat{w}_{hj} = \frac{c_{phj}}{c_{pj}}, \tag{3}$$

where $c_{pj} = \sum_{i=1}^{n.2} I[\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_j\}], \;\; c_{phj} = \sum_{i=1}^{n.2} I[\mathbf{z}_i = \{i, \mathbf{u}_h, \mathbf{v}_j\}].$

We can write the estimate of the normed by $n_{\cdot 2}$ log likelihood function of $\mathbf{q}_k$ corresponding to the all data, with pairs and singles, collected on the second step

$$\begin{aligned} \mathrm{llik}[\mathbf{q}_k \quad &| \quad \hat{W}_{mk}, \mathbf{d2}_{n_{\cdot 2}}] = \frac{n_{s2}}{n_{\cdot 2}} \sum_{h=1}^{m} \frac{t_{sh}}{n_{s2}} \mathrm{Log}\left[\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j \right] \\ &+ \quad \frac{n_{p2}}{n_{\cdot 2}} \sum_{j=1}^{k} \frac{c_{pj}}{n_{p2}} \mathrm{Log}[q_j] + \frac{1}{n_{\cdot 2}} \sum_{i=1}^{n_2} \mathrm{Log}[\hat{w}_{h_i j_i}], \end{aligned} \quad (4)$$

where t_{sh} is the number of times to state $\mathbf{u}_h$ in the part with singles $\{i, \mathbf{u}_{h_i}, NA\}$ and $n_{s2} = \sum_{h=1}^{m} t_{sh}$, $n_{\cdot 2} = n_{s2} + n_{p2}$.
The set of all possible values $q_j, j = 1, ..., k$, is the $(k-1)$-dimensional polyhedron

$$S_{k-1} = \{\mathbf{q}_k : \ 0 \leq q_j \leq 1, \ \sum_{j=1}^{k} q_j = 1\}. \quad (5)$$

A variant of algorithm in form of recursive iterations can be applied for obtaining ML-estimators in (4) of probabilities $\mathbf{q}_{tr}$.

Here, we have the following contractive recursion

$$q_j^{(r+1)} = \frac{n_{p2}}{n_{s2}+n_{p2}} q_j^{(1)} + \frac{n_{s2}}{n_{s2}+n_{p2}} \sum_{h \in \mathcal{D}_j} \frac{\hat{w}_{hj} q_j^{(r)}}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}^{(r)}} \hat{w}_{sh}, \qquad (6)$$

$$j = 1, ..., k, \quad q_j^{(1)} = \frac{c_{pj}}{n_{p2}}, \quad \hat{w}_{sh} = \frac{1}{n_{2s}} \sum_{i=1}^{n_{\cdot 2}} I[\mathcal{E}_i = \{\mathbf{u}_{h_i}, NA\}].$$

The recursion (6) can be used if for each $\hat{w}_{sh} > 0$ *also* $\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_j^{(1)} > 0$. This restriction implies that $\sum_{j=1}^{k} q_j^{(r)} = 1$ for any $r = 1, 2, ...$ and the stationary point exists. This restriction holds if n_{p2} and n_{s2} are sufficiently large.

Theorem 1. *Suppose that Assumptions 1 - 3 are valid, the sizes* n_{p2} *and* n_{sp} *of collected pairs and singles are growing unboundedly and*

$$0 < \underline{\lim}_{n_{2\cdot}\to\infty}\frac{n_{p2}}{n_{\cdot 2}} \leq \overline{\lim}_{n_{\cdot 2}\to\infty}\frac{n_{p2}}{n_{\cdot 2}} \leq 1. \tag{7}$$

Then for any sufficiently large n_{p2} *and* n_{s2} *the based on the data with singles and pairs the ML-estimator* $\check{\mathbf{q}}_{n_{\cdot 2}k}$ *exists, is consistent, and can be found as the stationary point of recursion (6).*

Hence, we can find consistent ML-estimates $\check{q}_{n_{\cdot 2}j}, j = 1, ..., k$, of the projection WTP-distribution on the divisions intervals $\mathbf{v}_j \in \mathcal{V}_k$, i.e. $\check{q}_{n_{\cdot 2}j} \to q_{trj}, j = 1, ..., k$, if n_{p2} and n_{s2} are growing.

It is reasonably to introduce approximating m_{1tr} the following estimable true *medium mean* of WTP-distribution

$$mm_{1tr} = v_{L1} + \sum_{j=1}^{k} \left(q_{tr,j+1}^{k} + \frac{1}{2} q_{trj} \right) (v_{Rj} - v_{Lj}), \tag{8}$$

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

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

$$\check{mm}_{1n.2} = v_{L1} + \sum_{j=1}^{k} \left(\check{q}_{n.2,j+1}^{k} + \frac{1}{2} \check{q}_{n.2j} \right) (v_{Rj} - v_{Lj}). \tag{9}$$

$\check{q}_{n.2j}$ is the ML-estimator of q_{trj} with the likelihood given in (4), $\check{q}_{n.2,j+1}^{k} = \sum_{i=j+1}^{k} \check{q}_{n.2i}$, $j = 1, ..., k$, $\check{q}_{n.2k+1} = 0$.

Numerical experiment

We consider the data, collected and used in the cost-benefit analysis by Håkansson (2008), as if they were realized on the fist step of data collecting. All self-selected intervals are given in the set $\mathcal{U}_m$, and all division intervals in $\mathcal{V}_k, m = 46, k = 23$. The estimate of the coverage probability is $\hat{p}_c \backsimeq 91\%$. In the second step we use the subsets $\mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}$ and $D_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}$. For a simulation of the data collected on the second step a computer has need to know a distribution of true WTP-points X_i and a behavior model describing decisions to state self-selected intervals. We suppose that the behavior model is BM3 and the WTP-distribution is the following p_{WE}-mixture WE of the Weibull $W(a, b)$ and the Exponential distribution $E(m_1), p_{WE} = 0.8160, a = 74.8992, b = 1.8374, m1 = 254.7344.$

$$sf_{WTP}[x] = p_{WE}e^{-(x/a)^b} + (1 - p_{WE})e^{-x/m_1}. \qquad (10)$$

The array $\mathbb{W} = (w_{hj})$ corresponds to BM3.

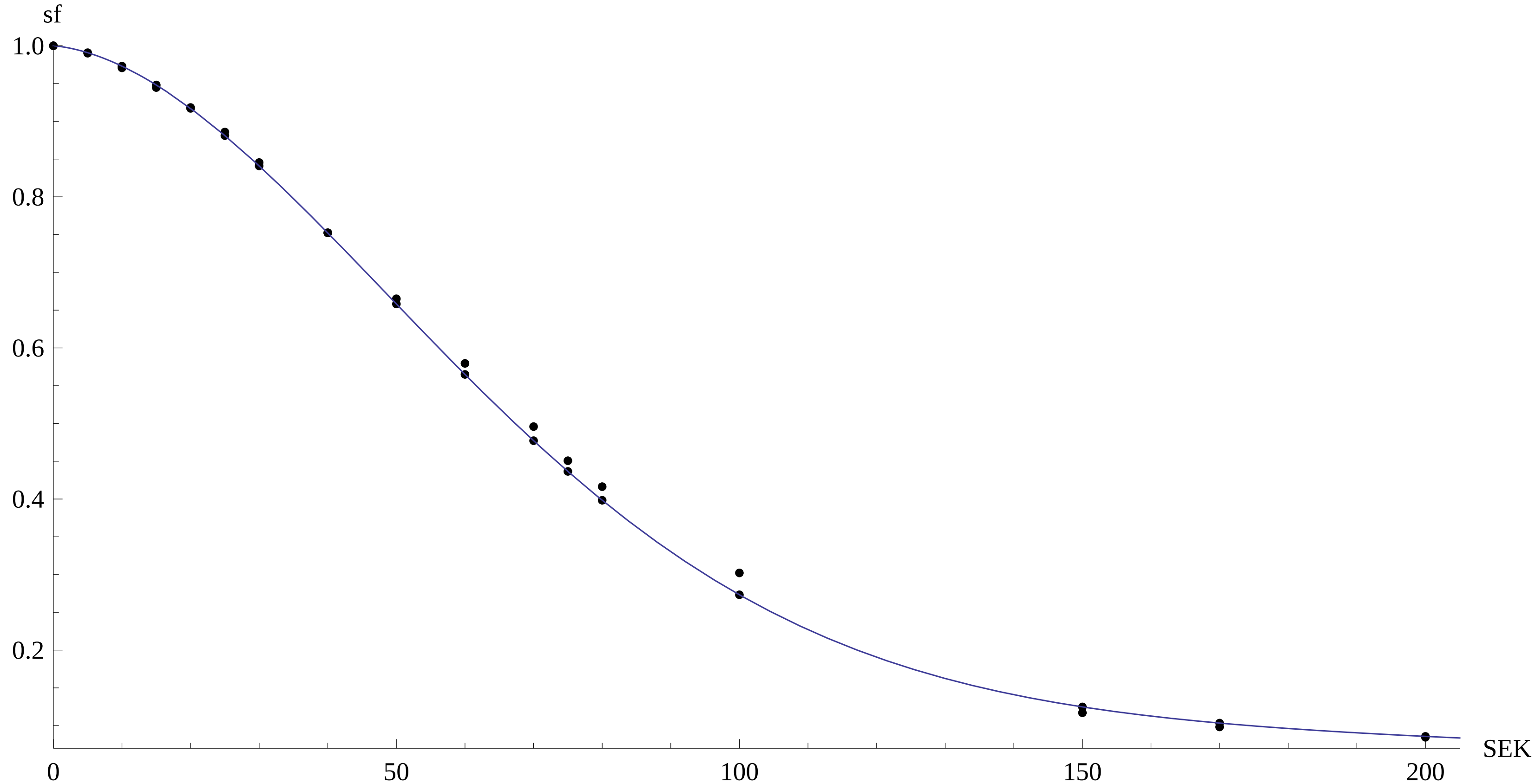

Figure 4: The part of the ”true” WTP-survival function is shown together with points corresponding to the values of the estimated survival function on the division points, if $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ in pairs were used. This part has size $n_{p2} = 3144$. The whole simulated ”true” data has size $n_{\cdot 2} = 9000$.

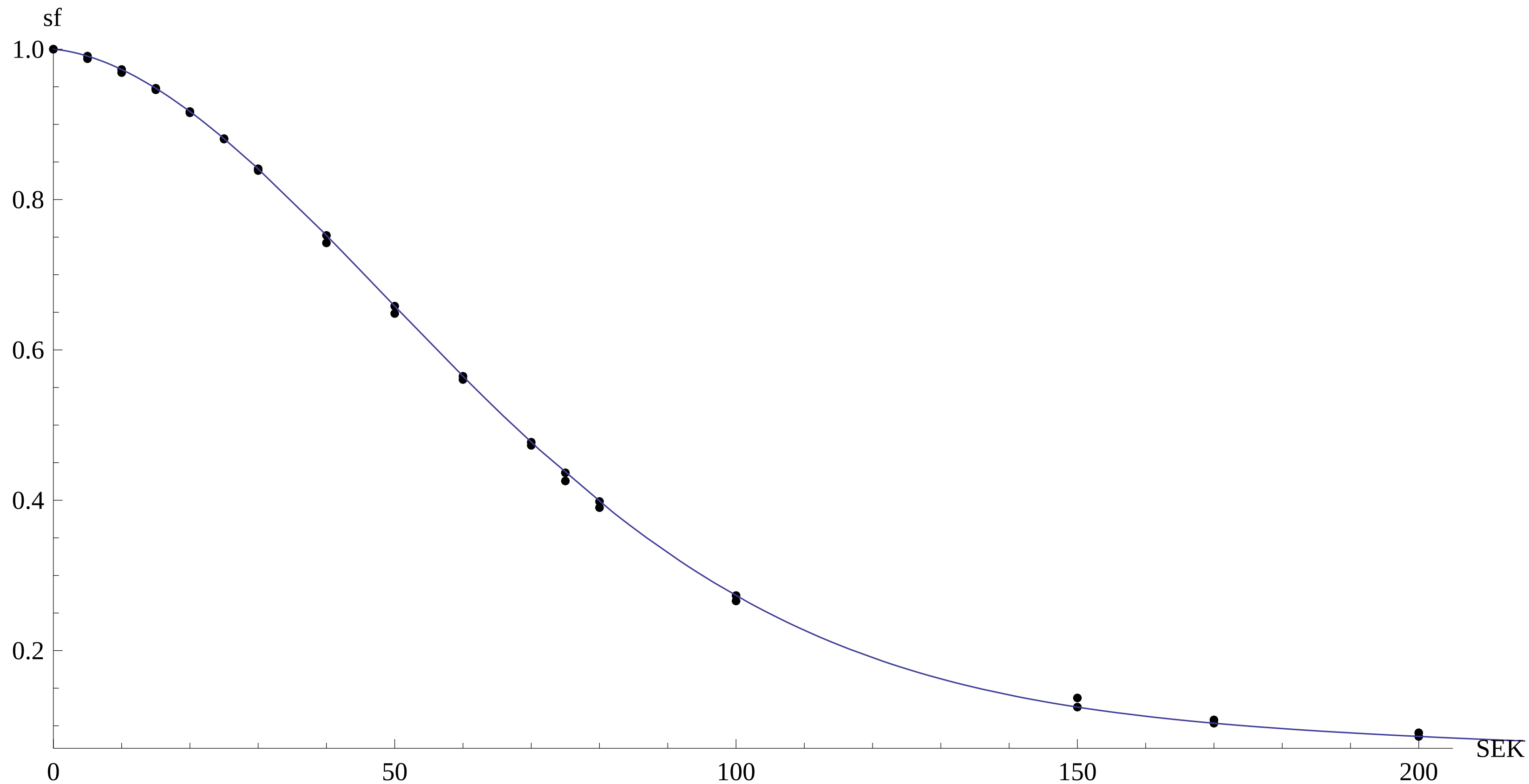

Figure 5: Part of the ”true” WTP-survival function together with points of the estimated survival function corresponding to the division points, if both $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ in the list of compatible pairs and $\{i, \mathbf{u}_{h_i}, NA\}$ in singles in the simulated ”true” data of size $n_{\cdot 2} = 9000$. 5 times iterated recursions were used.

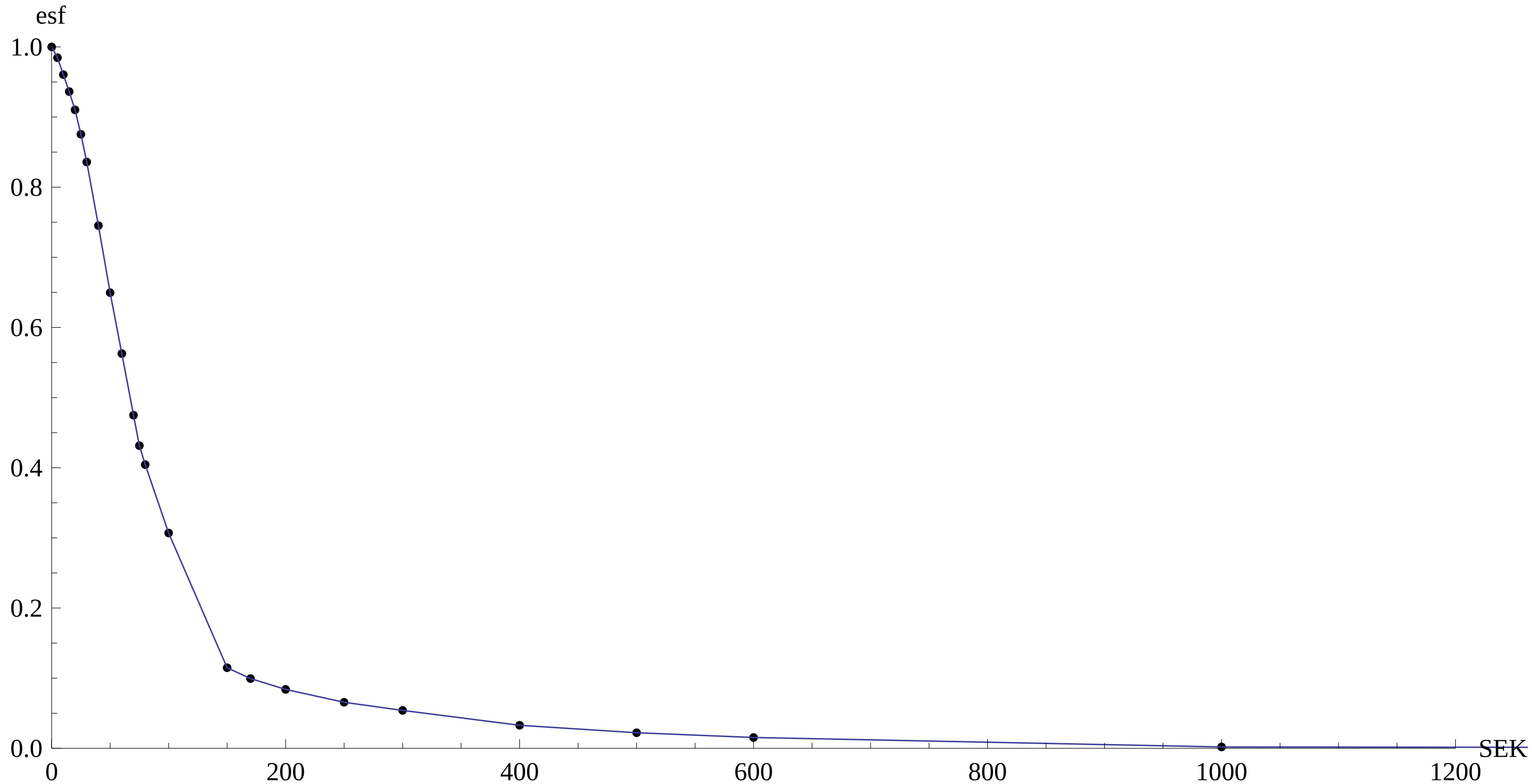

Figure 6: The broken line joins the estimated values $\check{sf}(5)[v_{Rj}]$ of WTP-survival function at ends of division intervals. The medium mean estimate $\hat{mm}_{1n_{\cdot 2}}$ is the area below the broken line.

QQ-plots are popular for graphical representations of empirical distributions. We use normal QQ-plots where normal distribution are shown as straight lines. In the following Figure 7 we see that accuracy of estimates obtained by iterated recursions based on resampling copies of original data (shown by solid line) can be considered as rather well approximation to the non observed accuracy of the true WTP-distribution corresponding to the estimate $m\check{m}_{1n\cdot 2}^{(5)}$ (shown by dashed line). For the dashed line $9000 \times 2000 = 18\,000\,000$ self-selected pairs and singles were simulated. For the solid line only 9000 pairs and singles were simulated.

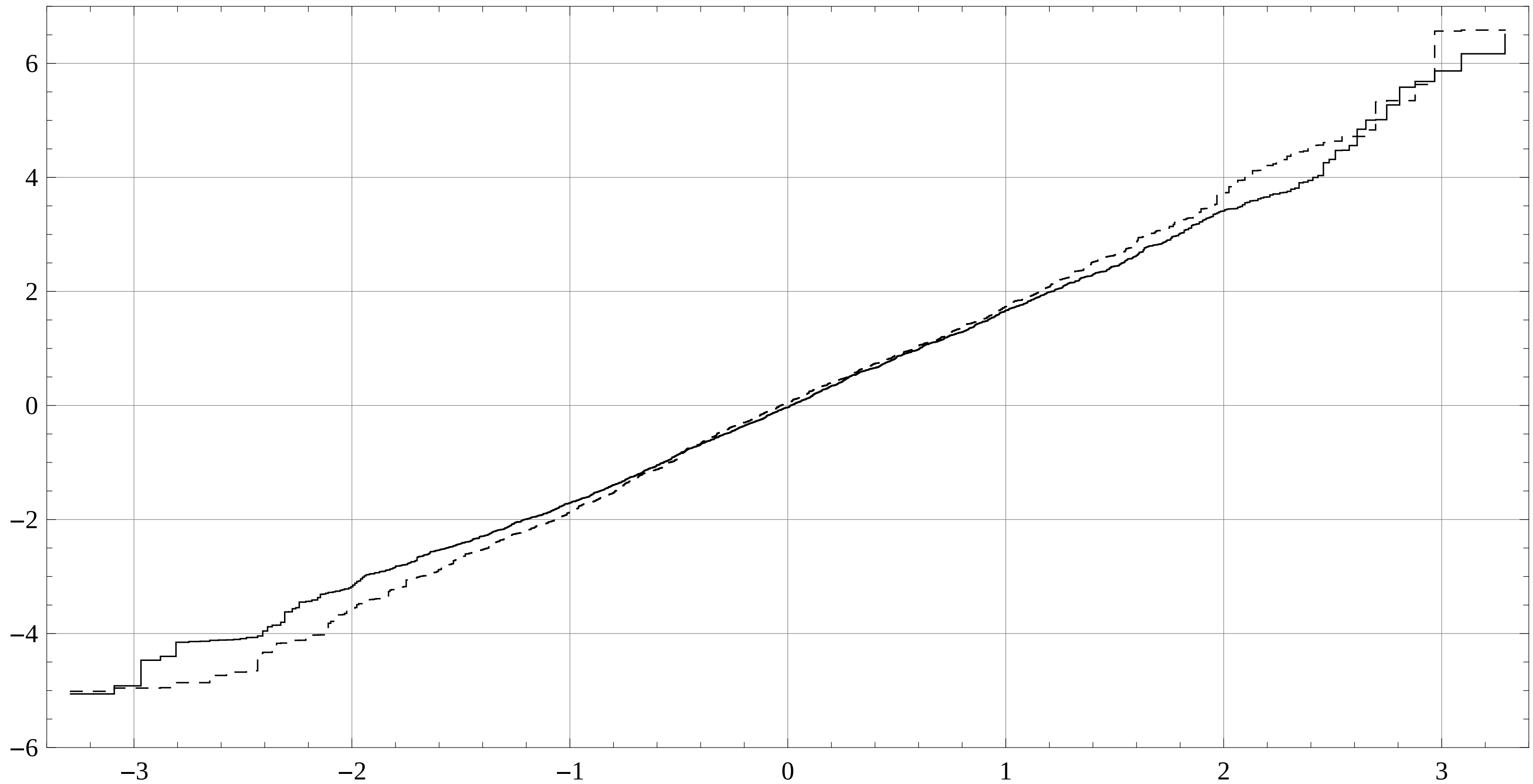

Figure 7: Two distributions, of the deviations of medium mean WTP-values from their average values. The dashed line corresponds to case with $C = 2000$ copies of true deviations $\check{mm}_{1n.2}^{(5)c} - \frac{1}{C}\sum_{c'=1}^{C} \check{mm}_{1n.2}^{(5)c'}$. The solid line corresponds to case with $C = 2000$ copies of deviations $\check{mm}_{1n.2}^{\star(5)c} - \frac{1}{C}\sum_{c'=1}^{C} \check{mm}_{1n.2}^{\star(5)c'}$ obtained via resamplings.

Conclusion.

Provide complete empirical approach for elicitation in surveys. Only 3 relatively mild assumptions needed for WTP-application.

Self-selected intervals allow for rounding, etc. The approach differs from standard bracketing and it has several advantages.

Propose two-step approach to self-selected intervals. Advantage: can identify nonparametric MLE. Disadvantage: Assumption 2 is crucial and not yet subject to empirical scrutiny.

Propose an approach to solve a sampling stoppage problem given selected coverage probability.

Propose on the second step estimate of ”psychological model” on how point related to interval and use all collected data.

Identifications and accuracies. Theorem 1 and Theorem 2 on MLE and resampling.

Extensive computing model checking. All needed programs were developed in R and collected in a software package.

Where do we go from here?

A) Need empirical tests (lab, field)

B) Allow background (explanatory) variables.