Second Nonparametric Approach for Double-Bound Dichotomous Choice Contingent Valuation*

Taku Terawaki†

Abstract

In this paper, I suggest the second nonparametric estimator in the double-bound dichotomous choice contingent valuation. It is different from the Turnbull estimator and is represented by a closed form. I compare the performances of the second estimator and the Turnbull estimator through Monte Carlo experiments and clarify that there is hardly any difference in the performances of both estimators. Also it is found that the even sub-sample design is desirable in the nonparametric estimation for the double-bound format.

1 Introduction

Recently, the most frequently used question formats for contingent valuation (CV) survey are single-bound and double-bound dichotomous choice formats. In the single-bound format each respondent is asked once whether she would be willing to pay a specified bid amount, and in the double-bound format, after the single-bound question, she is asked once again whether she would be willing to pay another bid amount. The single-bound format is incentive compatible theoretically and has the advantage of making the responses easy since it is similar to our real purchase actions.[1] Although the double-bound format has not such advantages, it is more efficient in estimating willingness to pay (WTP) since it makes the respondent's WTP more restrictive.

As for the econometric models used to estimate the WTP distributions for each question format, they are roughly classified into two groups. One is the parametric approach, the other is the nonparametric approach. The parametric approach is based on the assumption of a parametric distribution and employed earlier in history for each format.[2] However, since the true WTP distribution is unknown, the distribution assumption is a heavy constraint obviously. If we mis-specify the distribution assumption, the maximum likelihood estimators of parameters in the distribution are inconsistent in general, and therefore the estimates of WTP may have a large error. In the single-bound format, Kriström[9] first noted this point and suggested the nonparametric method by applying the estimator developed by Ayer *et al.*[2]. Since the Ayer *et al.* nonparametric estimator is represented by a closed and simple form, this method has the advantage of not requiring the complex calculation for optimization like the parametric approach, and therefore has the attraction of being a "convenient method".

On the other hand, as to the nonparametric approach in the double-bound format, it is usual to apply the Turnbull[12][13] estimator as McFadden[10] and Carson, Wilks, and Imber[4], extended the Ayer *et al.* estimator, which treats a right or left censored data, to a interval censored data.[3] However, the nonparametric estimator in this more complex question format is not represented by a closed form unlike that in the single-bound format, and therefore requires the computer calculation for optimization in the same way as the parametric approach. Accordingly, this method is inconvenient.

*Discussion Paper No.02005, College of Economics, Ritsumeikan University, Version: January 27, 2003.

†Faculty of Economics, Ritsumeikan University
E-mail: `masala.terawaki@nifty.ne.jp`
URL: `http://www.taku-t.com/`

[1] Although this incentive compatibility was proved by Hoehn and Randall[8] theoretically, Cummings, Harrison and Rutström[5] later demonstrated that the single-bound format was not incentive compatible by using the methodology of experimental economics.

[2] I'm not concerned in this paper with the parametric approach. It is surveyed in Hanemann and Kanninen[6] in detail.

[3] According to Hanemann and Kanninen[6], Carson and Steinberg[3] first applied the Turnbull estimator to double-bound CV.

This paper suggests the second nonparametric estimator in the double-bound format, which is different from the Turnbull estimator and is represented by a closed form.[4] The basic idea of the nonparametric approaches discussed above is that the researchers treat respondent's answers of "yes" or "no" as independent Bernoulli trials respectively. Following this idea, in this paper I would like to extend the nonparametric approach in the single-bound format simply. Moreover, I will clarify the case when the second estimator is equivalent to the Turnbull estimator, as well as the case when both estimators are not equivalent; in the latter, I will compare the performances of both estimators through Monte Carlo experiments.

This paper is structured as follows. Section 2 introduces the basic idea of the second nonparametric estimator in the double-bound format. Section 3 clarifies the differences between the second estimator and the Turnbull estimator. Section 4 describes the basic design of the Monte Carlo experiments. Section 5 gives the simulation result for the design described in Section 4 and some extensions. Section 6 offers the conclusion of this paper.

2 Second Nonparametric Estimator

As a beginning, I would like to describe the nonparametric approach in the single-bound format. Let n_k denote the number of respondents who meet the k^{th} bid amount, b_k $(k = 1 \cdots M)$, and let m_k denote the number of respondents who answer "yes" to b_k. If the respondents are drawn randomly, we treat respondent's answers of "yes" or "no" as independent Bernoulli trials associated with each bid amount respectively, and therefore the proportion of yes-answers for b_k, m_k/n_k is the natural estimator of the probability of yes-answers for b_k, π_k, which represents the survival probability of WTP at b_k, $\text{Prob}\{b_k \leq WTP\}$. We can simply demonstrate that this estimator, m_k/n_k is a maximum likelihood estimator (MLE) for any k by solving the likelihood equation of the following log-likelihood function for vector parameter, $(\pi_1, \pi_2, \cdots, \pi_M)$.

$$\ln L = \sum_{k=1}^{M} \{m_k \ln \pi_k + (n_k - m_k)\ln(1 - \pi_k)\} \tag{2.1}$$

The survival function of WTP can be estimated by plotting the estimates of the probability of yes-answers for all the bid amounts on the plane of $b - \text{Prob}\{b \leq WTP\}$ and interpolating between all the points in some way like Figure 1.[5] However, in case the sequence of m_k/n_k is not monotonically non-increasing, for example, in the case of $m_k/n_k < m_{k+1}/n_{k+1}$ for an arbitrary k, we need to replace the estimates of two probabilities, m_k/n_k, m_{k+1}/n_{k+1} with $(m_k + m_{k+1})/(n_k + n_{k+1})$, and repeat this procedure until the sequence is monotonically non-increasing. This algorithm is developed by Ayer *et al.*, and is known as pool-adjacent-violator-algorithm (PAVA) in general. The estimate of mean WTP is represented by the area under the survival curve, and the estimate of median WTP is represented by the b-coordinate of the intersection point of the line for $\text{Prob}\{b \leq WTP\} = 0.5$ and the survival curve.

[4] This nonparametric estimator is introduced by the author in Japanese only in Terawaki[11]. In Japan, recently, this approach is applied in the field of agricultural policies mainly. This paper focuses on the difference between the second estimator and the Turnbull estimator which is not described in Terawaki[11].

[5] Although Figure1 shows the case of linear interpolation applied by Kriström[9], there is also the way to estimate step functions for the purpose of producing more conservative mean WTPs. See Hanemann and Kanninen[6], pp.398-402. Since this paper focuses on estimating the survival probability at bid amounts, and does not discuss how to derive the welfare measurements from those, I am not concerned here with the methodology of the interpolation.

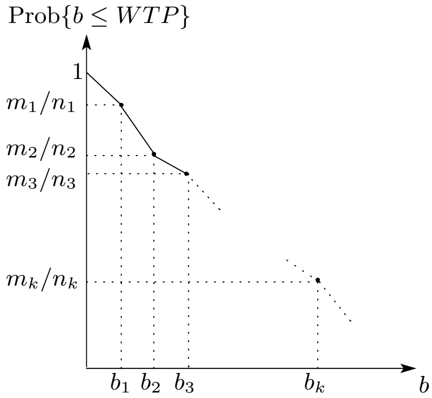


Figure 1: The Nonparametric Survival Curve in Single-bound Format

In this paper I would like to extend the nonparametric approach in the single-bound format simply on the basis of the idea to treat respondent's answers of "yes" or "no" as independent Bernoulli trials respectively, and call it the second nonparametric estimator. As describe below, the crucial difference between the second estimator and the Turnbull estimator is that the former focuses on only one scalar parameter, that is only one survival probability for a bid amount, and estimates only the parameter by maximum likelihood, while the latter estimates the vector parameter which consists of all the survival probabilities for all the bid amounts, by maximum likelihood.

The basic idea is as follows. First, as for each bid amount, we select the observations which can be useful for the test of "pay" or "don't pay", that is the respondents who have the possibility to meet the bid amount. Since each respondent can meet not only the initial bid amount, but also the second higher and lower bid amount in the double-bound format, the answer of one respondent has the information of "pay" or "don't pay" for three bid amounts simultaneously. If the sample is drawn randomly, we can treat the selected observations as independent Bernoulli trials. Second, we calculate the proportion of respondents who answer "yes" or will reasonably answer "yes" for the bid amount in the selected observations. That is the estimate of survival probability for the bid amount.

Let b_k denote the initial bid amount offered the k^{th} group, and let b_k^d denote the second lower bid offered when a respondent in the k^{th} group answers "no" for b_k, and let b_k^u denote the second higher bid offered when a respondent in the k^{th} group answers "yes" for b_k. And then, let m_k^{nn}, m_k^{ny}, m_k^{yn}, m_k^{yy} respectively denote the number of respondents who answer "no" for both b_k and b_k^d, the number of respondents who answer "no" for b_k and "yes" for b_k^d, the number of respondents who answer "yes" for b_k and "no" for b_k^u, the number of respondents who answer "yes" for both b_k and b_k^u, and let n_k denote the the selected observations, that is $m_k^{nn}+m_k^{ny}+m_k^{yn}+m_k^{yy}$. Here, as Figure2 shows, the number of observations for $b_{kd} \leq WTP$ is $m_k^{ny}+m_k^{yn}+m_k^{yy}$, and the number of observations for $b_k \leq WTP$ is $m_k^{yn}+m_k^{yy}$, and the number of observations for $b_{ku} \leq WTP$ is m_k^{yy}. Accordingly, the estimators of the survival probabilities are represented by

$$\begin{aligned} \hat{\pi_k^d} &= \frac{m_k^{ny}+m_k^{yn}+m_k^{yy}}{n_k}, \\ \hat{\pi_k} &= \frac{m_k^{yn}+m_k^{yy}}{n_k}, \\ \hat{\pi_k^u} &= \frac{m_k^{yy}}{n_k}, \end{aligned} \tag{2.2}$$

where $\pi_k^d :=$ Prob$\{b_k^d \leq WTP\}$, $\pi_k :=$ Prob$\{b_k \leq WTP\}$ and $\pi_k^u :=$ Prob$\{b_k^u \leq WTP\}$. Also $\hat{A}$ represents the estimator of A. Surely these estimators are the MLEs for each survival probability.

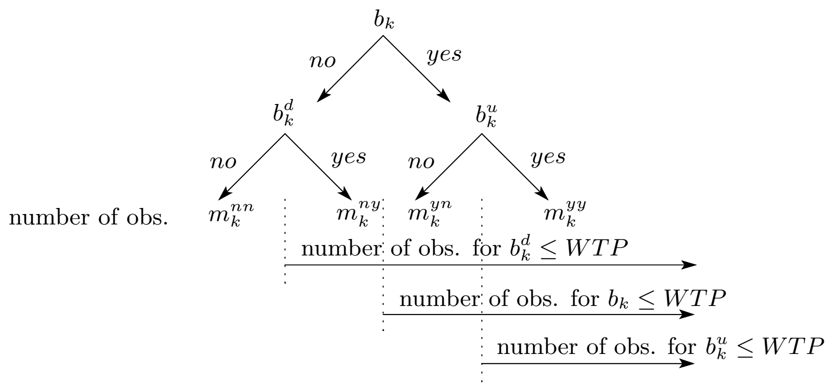


Figure 2: The Number of Respondents who Answer "yes" for the Bid Amounts

3 Difference from the Turnbull Estimator

So far the Turnbull estimator, developed in the field of survival analysis, has been mainly applied to the nonparametric approach in the double-bound format. In fact, this second nonparametric estimator is not equivalent to the Turnbull estimator in general. This depends on the former estimating only one scalar parameter, that is only one survival probability, by maximum likelihood, while the latter estimates the vector parameter which consists of all the survival probabilities, by maximum likelihood. The two estimators are equivalent only in the case where the bid amount offered one group does not coincide with the bid amount offered another group.

Suppose that $b_i \neq b_j^d$, $b_i \neq b_j^u$, for any i, j under the notation defined in the previous section. Then, the log-likelihood function is represented by

$$\ln L = \sum_{k=1}^{M} \left\{ m_k^{nn} \ln(1 - \pi_k^d) + m_k^{ny} \ln(\pi_k^d - \pi_k) + m_k^{yn} \ln(\pi_k - \pi_k^u) + m_k^{yy} \ln \pi_k^u \right\}, \tag{3.1}$$

where M is the number of groups. We can obtain the Turnbull nonparametric estimators by solving the likelihood equations for the log-likelihood function (3.1), which are represented by

$$\begin{aligned} \frac{\partial \ln L}{\partial \pi_k^d} &= -\frac{m_k^{nn}}{1 - \pi_k^d} + \frac{m_k^{ny}}{\pi_k^d - \pi_k} = 0 \quad \text{for } k = 1 \cdots M, \\ \frac{\partial \ln L}{\partial \pi_k} &= -\frac{m_k^{ny}}{\pi_k^d - \pi_k} + \frac{m_k^{yn}}{\pi_k - \pi_k^u} = 0 \quad \text{for } k = 1 \cdots M, \\ \frac{\partial \ln L}{\partial \pi_k^u} &= -\frac{m_k^{yn}}{\pi_k - \pi_k^u} + \frac{m_k^{yy}}{\pi_k^u} = 0 \quad \text{for } k = 1 \cdots M. \end{aligned} \tag{3.2}$$

The solutions of the simultaneous equations are just represented by equation (2.2).[6]

On the other hand, The two estimators are not equivalent in the case where the bid amount offered one group coincides with the bid amount offered another group. Even then, the second estimator is, however, represented by a simple and closed form like equation (2.2), while the Turnbull estimator is not represented by a closed form, and therefore requires the computer calculation for optimization.

Suppose that the initial bid b_k for the k^{th} group coincides the second higher bid b_{k-1}^u for the $(k-1)^{th}$ group, and also coincides the second lower bid b_{k+1}^d for the $(k+1)^{th}$ group, that is $b_k = b_{k-1}^u = b_{k+1}^d$ for any k.[7] Then, the set of observations which have the information of "pay" or "don't

[6] For example, we can confirm this result by implementing the following program under Mathematica (Version 3.0) which is the product of Wolfram Research, Inc..

```
LLF=m0*Log[1-p1]+m1*Log[p1-p2]+m2*Log[p2-p3]+m3*Log[p3]
dLLF=D[LLF, p1]==0, D[LLF, p2]==0, D[LLF, p3]==0
MLE=Simplify[Solve[dLLF, p1, p2, p3]]
```

[7] Carson, Wilks, and Imber[4] employs such a bid design, which is very popular in the CV survey using the double-bound format.

pay" for b_k consists of the set of observations of k^{th} group, $(k-1)^{th}$ group, and $(k+1)^{th}$ group. Accordingly, the total number is $n_{k-1}+n_k+n_{k+1}$ and the number of observations for $b_k \leq WTP$ is $m_{k-1}^{yy}+(m_k^{yn}+m_k^{yy})+(m_{k+1}^{ny}+m_{k+1}^{yn}+m_{k+1}^{yy})$, as shown in Figure 3. As a result, the second estimator of survival probability for b_k, $\hat{\pi_k}$ is represented by

$$\hat{\pi_k} = \frac{m_{k-1}^{yy}+(m_k^{yn}+m_k^{yy})+(m_{k+1}^{ny}+m_{k+1}^{yn}+m_{k+1}^{yy})}{n_{k-1}+n_k+n_{k+1}}. \tag{3.3}$$

Of course, as for the estimators of survival probabilities for the minimum initial bid, b_1, the minimum second lower bid, b_{1d}, the maximum initial bid, b_M and the maximum second higher bid, b_{Mu}, we can not apply equation (3.3) as it is. Those estimators are represented by the following equations (3.7).[8]

$$\hat{\pi_1} = \frac{(m_1^{yn}+m_1^{yy})+(m_2^{ny}+m_2^{yn}+m_2^{yy})}{n_1+n_2} \tag{3.4}$$

$$\hat{\pi_1^d} = \frac{m_1^{ny}+m_1^{yn}+m_1^{yy}}{n_1} \tag{3.5}$$

$$\hat{\pi_M} = \frac{m_{M-1}^{yy}+(m_M^{yn}+m_M^{yy})}{n_{M-1}+n_M} \tag{3.6}$$

$$\hat{\pi_M^u} = \frac{m_M^{yy}}{n_M} \tag{3.7}$$

Although the survival function of WTP can be estimated by plotting the estimates of the survival probabilities for all the bid amounts on the plane of $b-\text{Prob}\{b \leq WTP\}$ and interpolating between all the points in some way, the sequence of the estimates is not always monotonically non-increasing. For example, consider the case of

$$\begin{aligned}\frac{m_{k-1}^{yy}+(m_k^{yn}+m_k^{yy})+(m_{k+1}^{ny}+m_{k+1}^{yn}+m_{k+1}^{yy})}{n_{k-1}+n_k+n_{k+1}} \\ < \frac{m_k^{yy}+(m_{k+1}^{yn}+m_{k+1}^{yy})+(m_{k+2}^{ny}+m_{k+2}^{yn}+m_{k+2}^{yy})}{n_k+n_{k+1}+n_{k+2}},\end{aligned} \tag{3.8}$$

that is $\hat{\pi_k} < \hat{\pi_{k+1}}$ for an arbitrary k. In this case, although we can employ the PAVA in the Turnbull estimator, we cannot simply employ the PAVA in the second estimator. The reason is that the PAVA is derived from the maximization of the likelihood function for the vector parameter subject to the constraint that the sequence is monotonically non-increasing, and is not related to the calculation structure to estimate only one scalar parameter by maximum likelihood in the second nonparametric approach. However, since there is the constraint of $\pi_k \geq \pi_{k+1}$ for any k, in the case of $\hat{\pi_k} < \hat{\pi_{k+1}}$, it is natural that we consider π_k and π_{k+1} to be equivalent, replace $\hat{\pi_k}$ and $\hat{\pi_{k+1}}$ with another equivalent estimator. In this paper, I suggest replacing these estimators with the proportion of the number of respondents who answer "yes" or will reasonably answer "yes" for b_k, or b_{k+1} in the number of respondents who can meet b_k or b_{k+1}, that is

$$\hat{\pi_k} = \hat{\pi_{k+1}} = \frac{m_{k-1}^{yy}+(m_k^{yn}+m_k^{yy})+(m_{k+1}^{ny}+m_{k+1}^{yn}+m_{k+1}^{yy})+(m_{k+2}^{ny}+m_{k+2}^{yn}+m_{k+2}^{yy})}{n_{k-1}+n_k+n_{k+1}+n_{k+2}}. \tag{3.9}$$

As described above, since the second nonparametric estimator is represented by simple and closed form, we can calculate the estimates of the survival probabilities easily. Such a convenience of calculation is a great advantage of this approach.

[8] As described above, although the estimators represented by equations (3.3), (3.7) are MLEs for each survival probability, those are not MLEs for all the survival probabilities. Thus these second nonparametric estimators are not equivalent to the Turnbull nonparametric estimators. For example, we can confirm this by implementing the following program under Mathematica.

```
LLF=m00*Log[1-p1]+m01*Log[p1-p2]+m02*Log[p2-p3]+m03*Log[p3]
+m10*Log[1-p2]+m11*Log[p2-p3]+m12*Log[p3-p4]+m13*Log[p4]
+m20*Log[1-p3]+m21*Log[p3-p4]+m22*Log[p4-p5]+m23*Log[p5]
dLLF=D[LLF, p1]==0, D[LLF, p2]==0, D[LLF, p3]==0,
D[LLF, p4]==0, D[LLF, p5]==0
MLE=Simplify[Solve[dLLF, p1, p2, p3, p4, p5]]
```

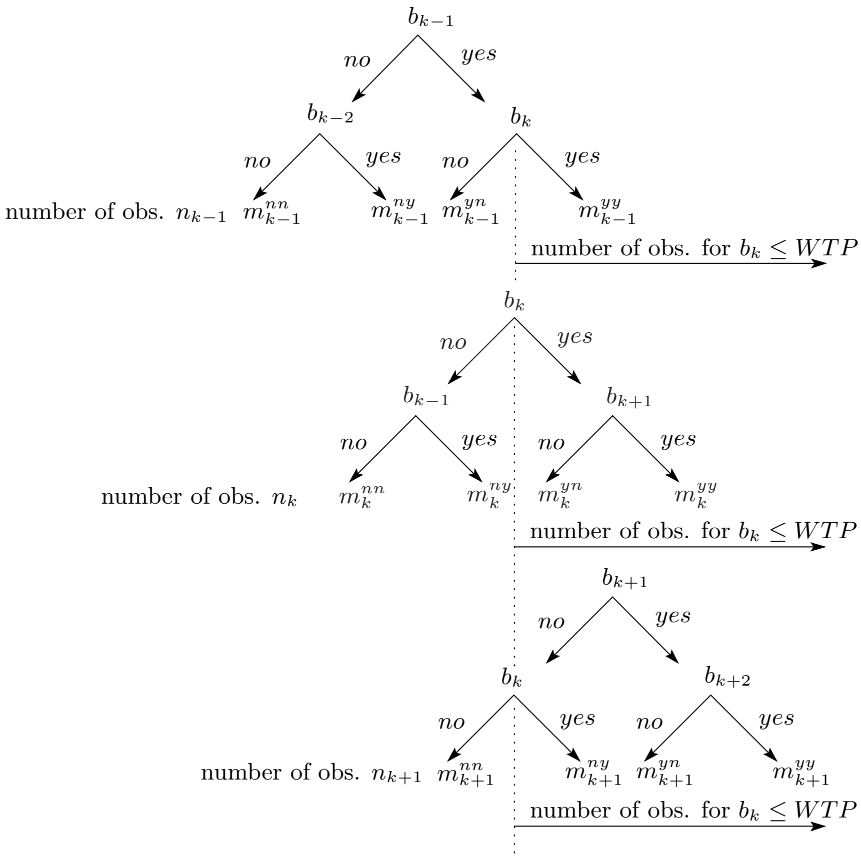


Figure 3: The Number of Observations for $b_k \leq WTP$ in Popular Bid Design

4 Monte Carlo Experiments

In the previous section, I explained that the second nonparametric estimator is not equivalent to the Turnbull estimator in general. Although this estimator has the advantage as a convenient method, there is a strong possibility that it is inefficient, since the estimators are not the MLEs for all the parameters. In the following section, I compare the performances of the two estimators by performing the Monte Carlo experiments for the various combination of the number of initial bid amounts and the sub-sample sizes assigned for each initial bid, and examine whether the second nonparametric estimator is well worth using as a convenient method. In this section, I would like to describe the basic and simple design of the Monte Carlo experiments.

WTP distributions In these experiments, I assume the two WTP distributions as the true distributions. The observations of WTP are randomly drawn from these distributions.

As for the first distribution, from the viewpoint of using the distribution which is as close as possible to the real distribution as the true distribution, I employ the WTP distribution estimated for the agricultural multifunctionality in Terawaki[11]. Here, the WTP is assumed to follow the following log-logistic density function.

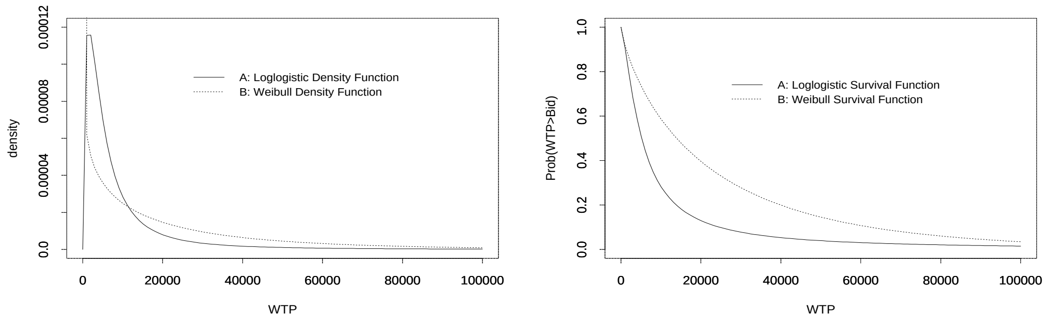
$$f(WTP) = \begin{cases} \dfrac{1.4065\exp(-12.027)WTP^{0.4065}}{\{1+\exp(-12.027)WTP^{1.4065}\}^2} & \text{if} \quad WTP > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

The mean of this distribution is 14,656, and the median is 5,172. The units are yen.

Moreover, as for the second distribution, I employ another distribution that has a different shape from log-logistic distributions. Here, the WTP is assumed to follow the following Weibull density function, which is not unimodal unlike log-logistic distributions.

$$f(WTP) = \begin{cases} \dfrac{0.8}{22026.47}\left(\dfrac{WTP}{22026.47}\right)^{-0.2}\exp\left\{-\left(\dfrac{WTP}{22026.47}\right)^{0.8}\right\} & \text{if} \quad WTP > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

The mean of this distribution is 24,956, and the median is 13,931. Figure 4 shows the both density functions and survival functions. We see from the density functions that the Weibull distribution is not unimodal, while the log-logistic distribution is unimodal, and from the survival functions that although the both survival probabilities converge to 0 around 100,000, the decreasing patterns are different obviously.



(a) density function (b) survival function

Figure 4: WTP Distributions

Bid Designs The number of initial bid amounts M is assumed to be 2, 4, 6 or 8. Also, the second bid amounts are assumed to be $b_k^u = b_{k+1}$ and $b_k^d = b_{k-1}$ for any k, and as for the maximum and minimum initial bid, I assume the special second bids, since higher and lower initial bids than those do not exist respectively. Accordingly, the total number of bids offered the respondents is $M+2$. The bid amounts are designed by using $M+2$ equi-probability quantiles, and the initial bids are designed to be the M bids except the maximum and minimum bids.[9] In the following section, I also perform experiments relaxing this constraint of the equi-probability bid design.

Sub-sample Designs The sub-sample sizes n are assumed to be assigned evenly for each initial bid, and to be 25, 50, 100 or 200. In the following section, I also perform the experiments in the situation where the sub-sample sizes increase or decrease with the rising initial bid amount.

Procedure The Monte Carlo experiments are performed as follows. First, we randomly draw the WTP observations $M \times n$ times from each distribution, and transform them into the double-bound data according to each bid and sub-sample design. Second, using the obtained sample, we estimate the survival probabilities by the second nonparametric method and the Turnbull method. Third, we repeat this operation 1,000 times, and calculate the biases, the standard deviations and mean squared errors (MSE) of each survival probability.[10]

5 Results

5.1 Results for Simple Design

Here, the 32 types of experiment are performed, since I assume two types of distribution, four types of the number of initial bids and four types of sub-sample size in the previous section. Table 1 shows the biases, the standard deviations and the MSEs of the second nonparametric estimators and those of the Turnbull estimators on each experiment. Although the $M+2$ survival probability parameters are estimated in each experiment, and therefore the $M+2$ biases and the $M+2$ standard deviations and the $M+2$ MSEs are calculated respectively, I present only the averages of those not to be complex in Table 1.

[9]This design sets the bid amounts at the quantiles which divide the area under the WTP density function into equal areas. For example, see Alberini[1], p.295. In this case, the bids divide the area into equal $M+3$ areas.

[10]The MSE is an error index which consists of the bias and the variance of estimator. The MSE of the estimator $\hat{\tau}$ is defined by the following equations.

$$\begin{aligned}\mathrm{MSE}(\hat{\tau}) &:= \mathrm{E}[(\hat{\tau}-\tau)^2]\\ &= \mathrm{E}[\{\hat{\tau}-\mathrm{E}(\hat{\tau})\}^2]+[\mathrm{E}(\hat{\tau})-\tau]^2\\ &= \mathrm{Var}(\hat{\tau})+\mathrm{Bias}(\hat{\tau})^2\end{aligned}$$

Table 1: Simulation Results in Simple Designs

	$n = 25$		$n = 50$		$n = 100$		$n = 200$	
	SECOND	TURNBULL	SECOND	TURNBULL	SECOND	TURNBULL	SECOND	TURNBULL
Log-logistic								
$M = 2$								
BIAS	0.307	0.328	0.127	0.138	0.054	0.049	0.057	0.061
S.D.	7.327	7.083	5.347	5.089	3.809	3.632	2.662	2.539
MSE	0.541	0.504	0.287	0.259	0.146	0.132	0.071	0.065
$M = 4$								
BIAS	0.151	0.143	0.157	0.114	0.060	0.069	0.077	0.076
S.D.	6.315	5.831	4.515	4.022	3.195	2.826	2.262	2.014
MSE	0.401	0.343	0.206	0.163	0.103	0.080	0.052	0.041
$M = 6$								
BIAS	0.206	0.193	0.075	0.061	0.048	0.057	0.052	0.046
S.D.	5.716	5.027	4.138	3.570	2.941	2.512	2.093	1.793
MSE	0.328	0.254	0.172	0.128	0.087	0.063	0.044	0.032
$M = 8$								
BIAS	0.142	0.084	0.096	0.071	0.064	0.050	0.043	0.036
S.D.	5.298	4.585	3.908	3.280	2.793	2.305	1.965	1.637
MSE	0.281	0.211	0.153	0.108	0.078	0.053	0.039	0.027
Weibull								
$M = 2$								
BIAS	0.160	0.168	0.117	0.137	0.100	0.077	0.026	0.021
S.D.	7.326	7.014	5.269	5.042	3.738	3.570	2.640	2.520
MSE	0.541	0.493	0.280	0.255	0.141	0.128	0.070	0.064
$M = 4$								
BIAS	0.137	0.177	0.168	0.135	0.092	0.059	0.036	0.027
S.D.	6.247	5.678	4.404	3.951	3.184	2.859	2.250	2.014
MSE	0.393	0.325	0.196	0.157	0.102	0.082	0.051	0.041
$M = 6$								
BIAS	0.117	0.074	0.112	0.059	0.077	0.068	0.090	0.066
S.D.	5.767	5.149	4.094	3.501	2.923	2.519	2.038	1.732
MSE	0.333	0.266	0.168	0.123	0.086	0.064	0.042	0.030
$M = 8$								
BIAS	0.174	0.105	0.089	0.081	0.042	0.038	0.054	0.039
S.D.	5.375	4.622	3.859	3.211	2.759	2.294	1.975	1.631
MSE	0.290	0.215	0.149	0.103	0.076	0.053	0.039	0.027

[a] The values of BIAS, S.D. and MSE are the average for all the survival probabilities, and multiplied by 10^2. Also the averages of biases are calcurated by using the absolute values of the biases.

As a beginning, note the averages of MSEs. We see from Table 1 that the MSEs of the Turnbull estimators are smaller than those of the second estimators in any combination of the number of initial bids and the sub-sample sizes, and the Turnbull estimator performs better than the second estimator. This expected result seems to depend on the second nonparametric estimator not being the MLE for the vector parameter.

Since MSE is a relative measure, we can not analyze only the scale. However, Table 1 seems to show that the differences between the MSE values of both estimators are small empirically. Here, consider to divide the MSEs into the biases and the standard deviations. First, we see from Table 1 that the maximum difference of the averages of the biases is at most 0.00069 in the experiment of $M = 8$, $n = 25$ and the Weibull distribution, and nothing up to the third decimal place. Also, there are seven experiments in which the biases of the second estimators are smaller, especially in the cases of the small sub-sample. Accordingly, we can conclude that there is hardly any difference in the performances of both estimators as long as we see only the biases.

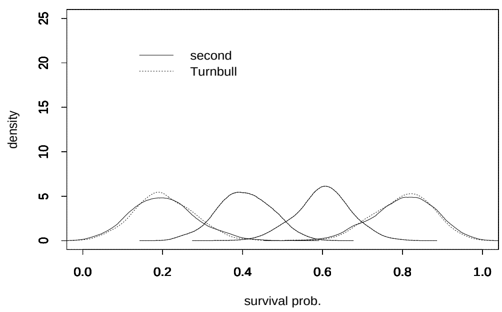
Second, note the standard deviations. We see from Table 1 that the averages of the standard deviations of the Turnbull estimators are smaller than those of the second estimators generally. However, the maximum difference of the averages of the standard deviations is 0.00753 in the experiment of $M = 8$, $n = 25$ and the Weibull distribution. If the distribution of a estimator is approximated to a normal distribution, the 95% confidence interval is presented by the inside of the double of standard deviation from the mean approximately. Accordingly, this implies that the 95% confidence intervals of the Turnbull estimators are at most 0.03 narrower than those of the second estimators on average. Although it is certain that the Turnbull estimator is more efficient for the second estimator, the difference seems not to be large.

Figure 5-8 shows the density estimation results for all the survival probabilities in each experiment. These are smoothed by using normal kernel functions.[11] We see from these figures that the tails of the distributions of the Turnbull estimators are fatter. This implies the difference of the standard deviations. Also we can confirm visually that there is hardly any difference in both the biases, since both the distributions are symmetrical and are very close in mode.

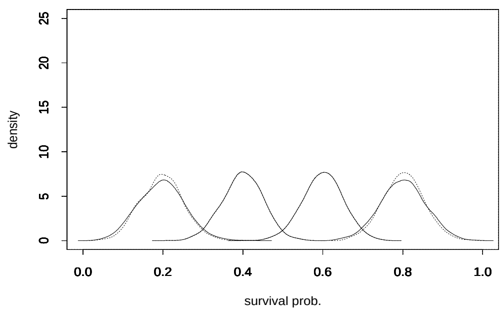
[11] The kernel density estimators are defined by the following $\hat{f}_h(x)$.

$$\hat{f}_h(x) = \frac{1}{Nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{N}\right)$$

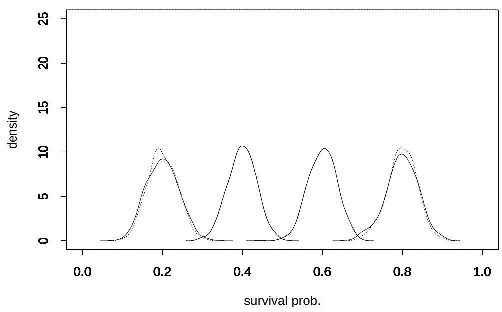
where h is a window or a bandwidth, N is the number of observations, $K(\cdot)$ is a kernel function and X_i is the i^{th} observation. A normal kernel function means the $K(u)$ being a standard normal density function, $\frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})$. To draw the figures, I use the command, `density` in S-PLUS 2000 that is the product of Insightful Corporation. Also, I use the default prepared in `density` for h.



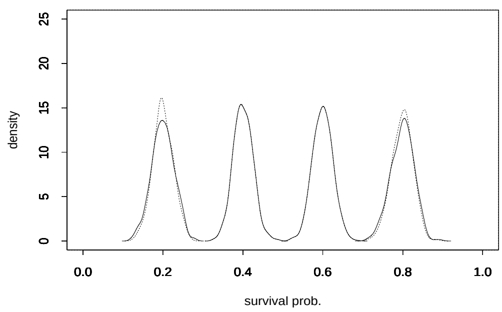
(a) Log-logistic, sub-sample size 25



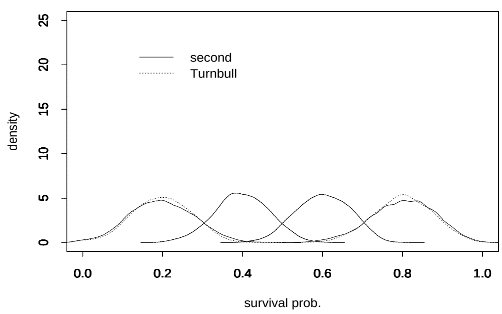
(b) Log-logistic, sub-sample size 50



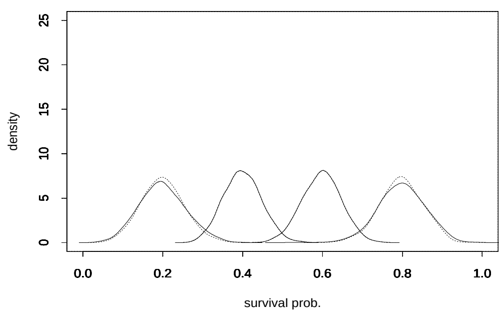
(c) Log-logistic, sub-sample size 100



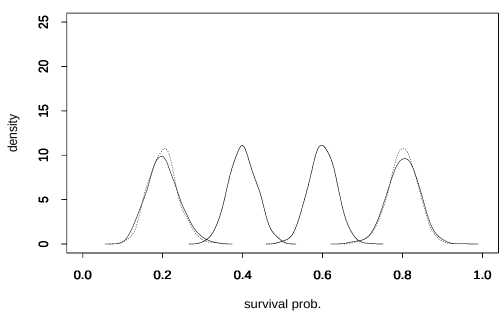
(d) Log-logistic, sub-sample size 200



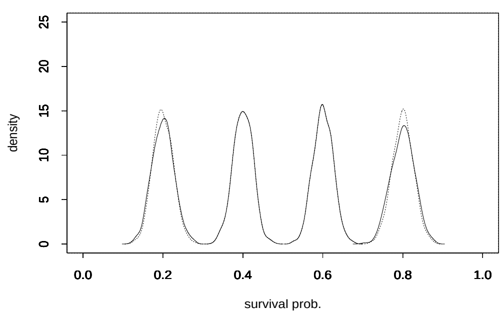
(e) Weibull, sub-sample size 25



(f) Weibull, sub-sample size 50

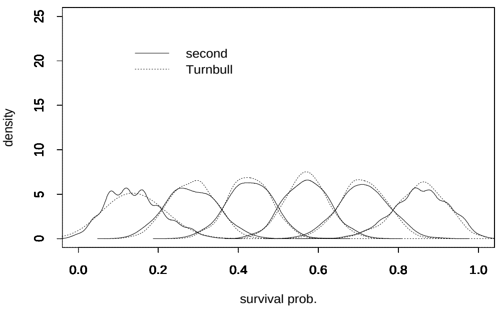


(g) Weibull, sub-sample size 100

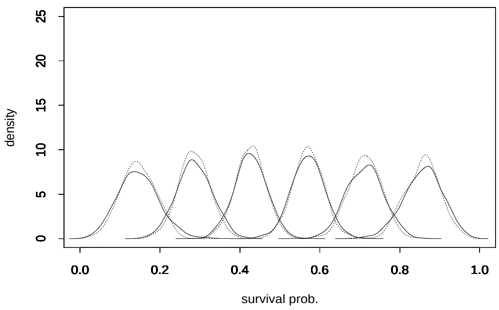


(h) Weibull, sub-sample size 200

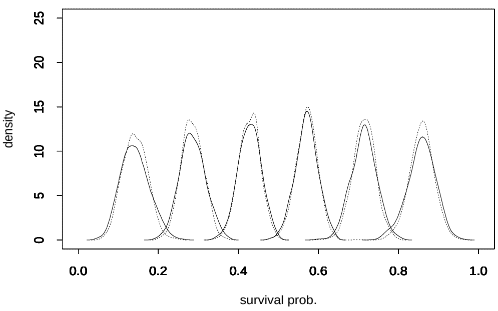
Figure 5: Density Estimation Result in Simple Design with the Number of Initial Bids 2



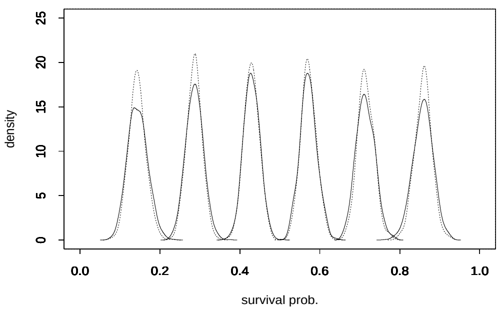
(a) Log-logistic, sub-sample size 25



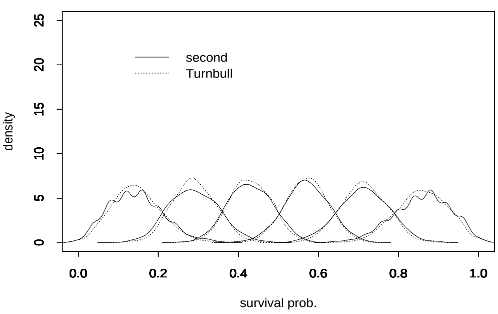
(b) Log-logistic, sub-sample size 50



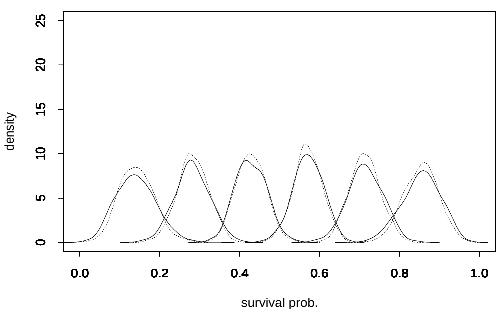
(c) Log-logistic, sub-sample size 100



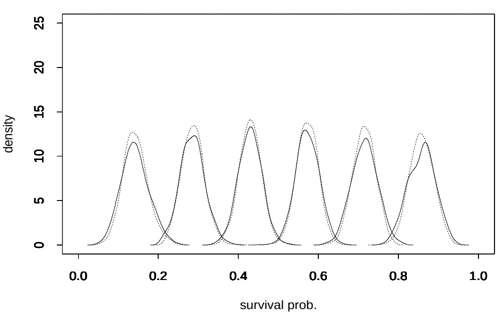
(d) Log-logistic, sub-sample size 200



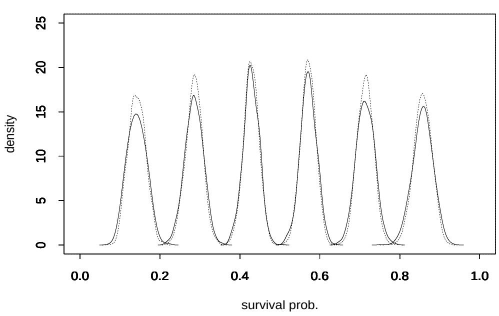
(e) Weibull, sub-sample size 25



(f) Weibull, sub-sample size 50

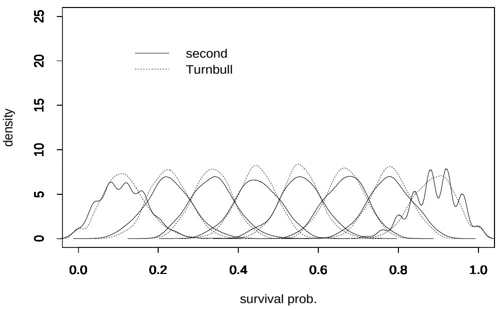


(g) Weibull, sub-sample size 100

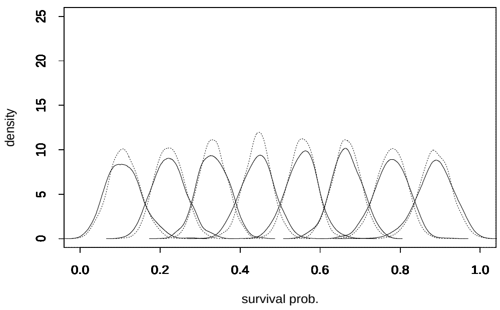


(h) Weibull, sub-sample size 200

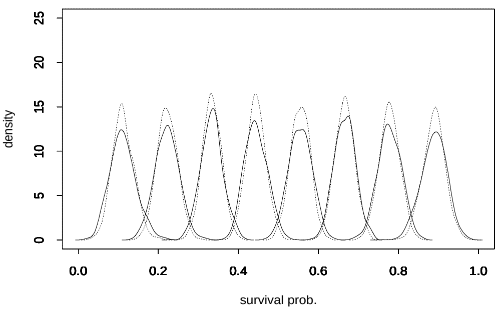
Figure 6: Density Estimation Result in Simple Design with the Number of Initial Bids 4



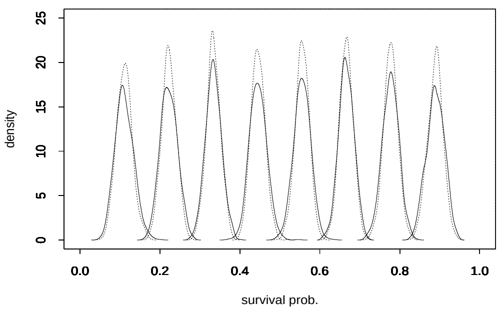
(a) Log-logistic, sub-sample size 25



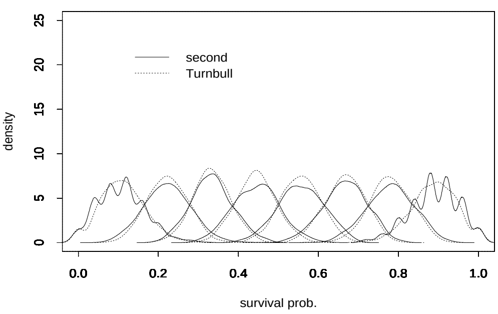
(b) Log-logistic, sub-sample size 50



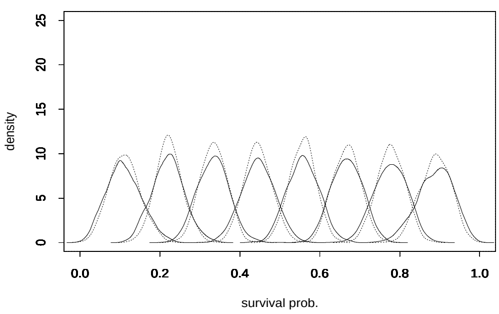
(c) Log-logistic, sub-sample size 100



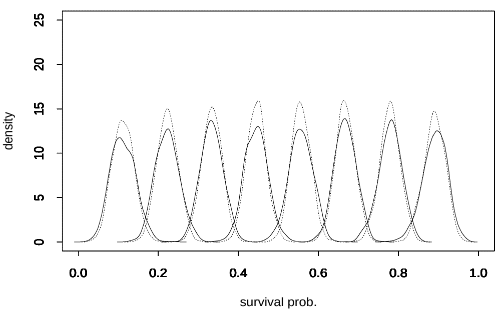
(d) Log-logistic, sub-sample size 200



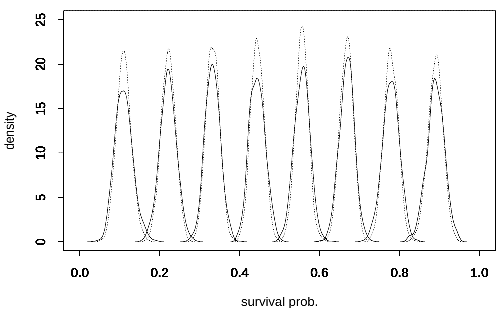
(e) Weibull, sub-sample size 25



(f) Weibull, sub-sample size 50

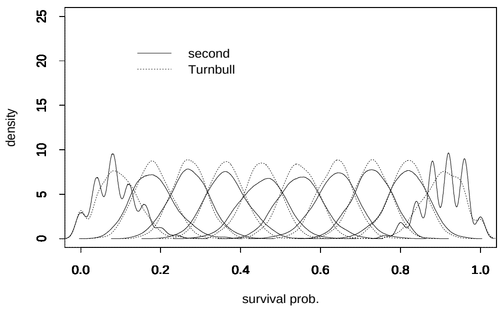


(g) Weibull, sub-sample size 100

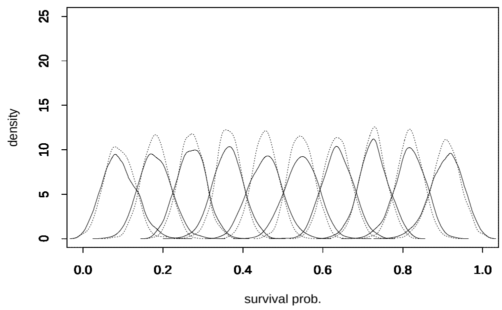


(h) Weibull, sub-sample size 200

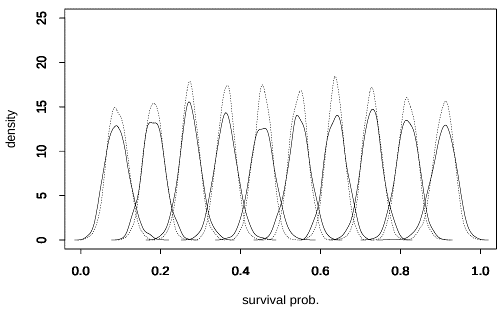
Figure 7: Density Estimation Result in Simple Design with the Number of Initial Bids 6



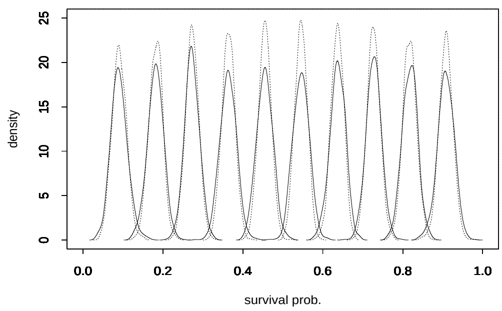
(a) Log-logistic, sub-sample size 25



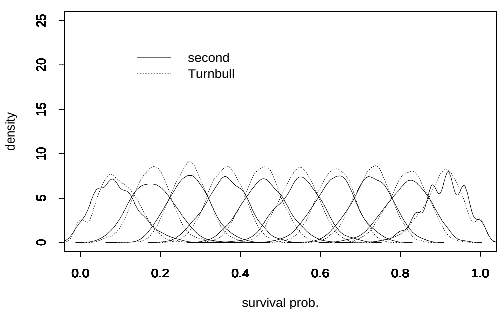
(b) Log-logistic, sub-sample size 50



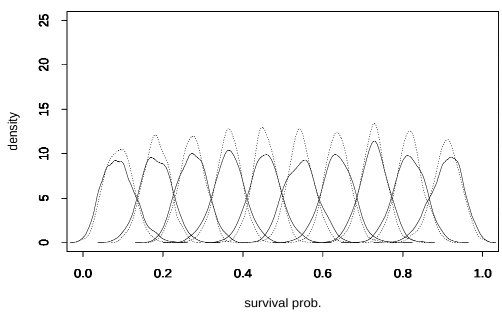
(c) Log-logistic, sub-sample size 100



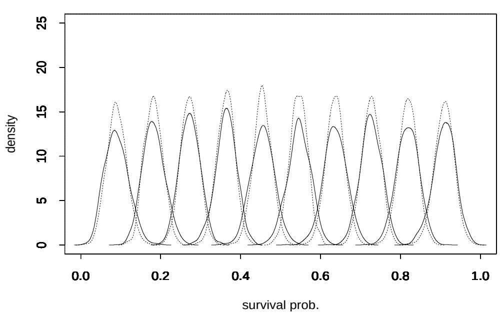
(d) Log-logistic, sub-sample size 200



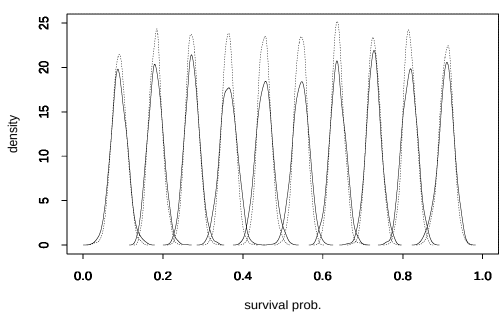
(e) Weibull, sub-sample size 25



(f) Weibull, sub-sample size 50



(g) Weibull, sub-sample size 100



(h) Weibull, sub-sample size 200

Figure 8: Density Estimation Result in Simple Design with the Number of Initial Bids 8

5.2 Results for Complex Design

In the previous section, I analyzed the experimental results based on equi-probability bid designs and even sub-sample designs, and clarified that the performances of the second estimators bore comparison with that of the Turnbull estimators on such designs. However it is difficult to collect such ideal data in the real surveys, because of mis-prediction of WTP distributions and response rates. Moreover, we have no warrant that the equi-probability bid designs and even sub-sample designs, which seem to be desirable intuitively, are optimal designs. In this section, I perform experiments in various situations where the probabilities divided by the bid amounts increase or decrease with rising bid amount, and the sub-sample sizes increase or decrease with the rising initial bid amount. I then compare the performances of the second estimators and the Turnbull estimators. Also I compare such complex designs with simple designs, namely equi-probability bid designs and even sub-sample designs, and investigate an efficient design.

In this paper I call the designs in which divided probabilities increase with rising bid amounts IP bid designs and the designs in which divided probabilities decrease with rising bid amounts DP bid designs. Here, I fix the number of initial bid amounts M to 4 and employ the quantiles of the probability 1/22, 3/22, 6/22, 10/22, 15/22, 21/22 as bid amounts of the IP designs and the quantiles of the probability 6/22, 11/22, 15/22, 18/22, 20/22, 21/22 as bid amounts of the DP designs. The first differences of sequence in the IP designs compose the arithmetic sequence with the tolerance 1/22 and those in the DP designs compose the arithmetic sequence with the tolerance -1/22.

In the same way I call the designs in which sub-sample sizes increase with rising initial bid amounts IS designs, and the designs in which sub-sample sizes decrease with rising initial bid amounts DS designs. Here, I employ 40, 80, 120, and 160 as the sub-sample sizes. Since the total sample sizes are 400 in these designs, we can compare these designs with the designs in which the number of initial bids is 4 and the sub-sample sizes are 100 evenly from the perspective of efficient design. In this paper, I call equi-probability bid designs EP designs and even sub-sample designs ES designs.

Table 2 shows the experimental results in various combinations of bid designs and sub-sample designs. I employ the log-logistic distributions as true distributions. As described above, the result in EP-ES design is the same as that in the simple design of $M = 4$, $n = 100$ and the log-logistic distribution, which is shown in Table 1.

Table 2: Simulation Results in Complex Designs

	ES		IS		DS	
	Second	Turnbull	Second	Turnbull	Second	Turnbull
EP						
bias	0.060	0.069	0.049	0.058	0.102	0.090
s.d.	3.195	2.826	3.540	3.120	3.473	3.029
mse	0.103	0.080	0.136	0.103	0.131	0.098
IP						
bias	0.076	0.066	0.062	0.033	0.076	0.072
s.d.	2.589	2.429	2.705	2.545	2.775	2.583
mse	0.069	0.061	0.076	0.067	0.084	0.072
DP						
bias	0.071	0.039	0.125	0.092	0.049	0.045
s.d.	2.855	2.599	3.309	2.932	2.898	2.683
mse	0.089	0.073	0.147	0.111	0.085	0.073

[a] The values of bias, s.d. and mse are the averages for all the survival probabilities, and multiplied by 10^2. Also the averages of biases are calculated by using the absolute values of the biases.

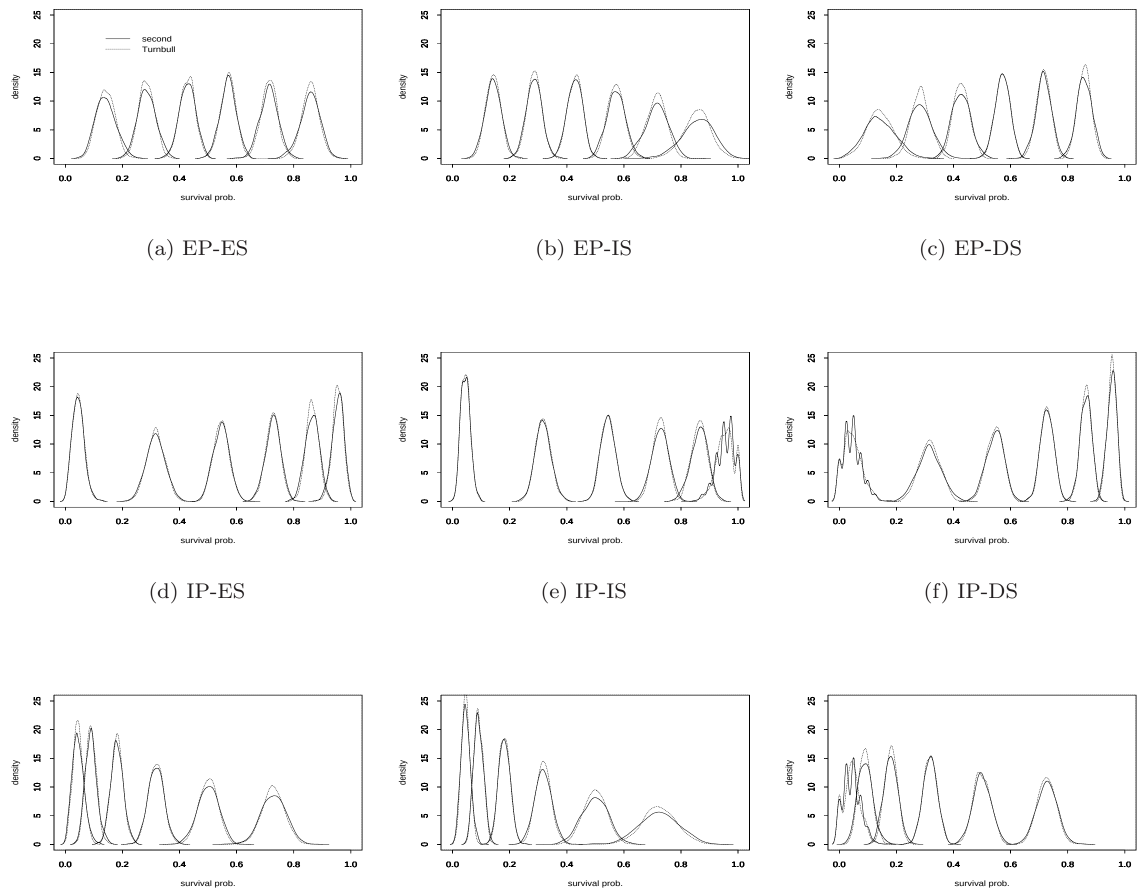
We see from Table 2 that the MSEs of the Turnbull estimators are smaller than those of the second estimators in such complex designs as well as simple designs, that is EP-ES design. However the differences between the MSEs of two estimators seem to be almost the same as those in the EP-ES design. To clarify this intuition, I show the MSEs of the second estimators divided by those of the Turnbull estimators in Table 3. We see from this table that the ratio values in the complex design except for the EP-ES design are very close to 1.28 in EP-ES design or less than that. It follows from this result that the conclusion in the previous section doesn't change in such complex designs.

Table 3: Ratios of the Average MSEs in Second and Turnbull's Estimator

	ES	IS	DS
EP	1.283	1.321	1.340
IP	1.135	1.135	1.169
DP	1.215	1.316	1.164

[a] The values are multiplied by 10^2.

Also, in the same way as the previous section, I show the density estimation results for all the survival probabilities in each experiment in Figure 9. We can confirm visually from Figure 9 that there is hardly difference in the performances of two estimators.



(a) EP-ES (b) EP-IS (c) EP-DS

(d) IP-ES (e) IP-IS (f) IP-DS

(g) DP-ES (h) DP-IS (i) DP-DS

Figure 9: Density Estimation Results in Complex Designs

Next we show the efficient design. As for the bid designs, we cannot generally conclude which is desirable, since the conclusion depends on the representative value which researchers want and the methodology of the interpolation. However, as for the sub-sample designs, we can see from Table 2 that the averages of MSEs in ES designs for each estimator are smallest in EP or IP design. Also, the average of MSEs in ES designs for the Turnbull estimator is smallest in DP design and that in ES designs for the second estimator is close to the smallest value in DP design, although it is larger than that in DS design. We can conclude from this result that the even sub-sample (ES) design is desirable.

6 Application

Finally I show some applications of the second estimators. Table 4 shows the results of the two real CV surveys with double-bound dichotomous choice formats. These CV surveys were practiced for the purpose of measuring the amenity effects of "Water Environment Improvement Project" in 1998.[12] Although these surveys were practiced in 24 districts, here I select two districts with almost same sample sizes and different sub-sample allocations. We can see from Table 4 that both the sample sizes or the averages of sub-sample sizes are very close and the standard deviations are different. Also, the first column in Table 4 shows the initial bid amounts. The second lower and higher bids are set as described in Section 2, and the second lower bid for the minimum initial bid is set at 250; the second higher bid for the maximum initial bid is set at 100,000.

Table 4: Results of the Two Real CV Surveys with Double-bound Formats

initial bids	Data Set I					Data Set II				
	m_k^{nn}	m_k^{ny}	m_k^{yn}	m_k^{yy}	n_k	m_k^{nn}	m_k^{ny}	m_k^{yn}	m_k^{yy}	n_k
500	0	3	18	13	34	0	4	21	12	37
1000	1	9	15	13	38	0	13	22	16	51
3000	2	20	14	2	38	0	19	16	9	44
5000	9	13	4	2	28	2	8	14	4	28
10000	16	10	3	0	29	7	13	10	3	33
30000	26	9	4	0	39	7	10	7	1	25
50000	26	4	1	0	31	11	1	3	0	15
sample size					237					233
average					33.86					33.29
s.d.					4.26					11.17

Table 5 shows the survival probabilities estimated by using the data sets shown in Table 4. We see from Table 5 that the average of the differences between the survival probabilities estimated by the two methods is smaller in the data set with almost even sub-sample. This result supports the conclusion in the previous section that the even sub-sample design is desirable.

Table 5: Survival Probabilities Estimated by the Data Sets shown in Table 4

bids	Data Set I			Data Set II		
	Second(a)	Turnbull(b)	(a)-(b)	Second(a)	Turnbull(b)	(a)-(b)
250	1.000	1.000	0.000	1.000	1.000	0.000
500	0.944	0.970	−0.025	0.955	0.979	−0.025
1000	0.700	0.766	−0.066	0.712	0.805	−0.093
3000	0.462	0.483	−0.021	0.545	0.577	−0.032
5000	0.221	0.247	−0.026	0.505	0.424	0.081
10000	0.188	0.137	0.051	0.407	0.241	0.166
30000	0.091	0.053	0.038	0.205	0.098	0.107
50000	0.014	0.006	0.008	0.100	0.033	0.067
100000	0.000	0.000	0.000	0.000	0.000	0.000
average			0.026			0.063

[a] The averages of (a)-(b) are calcurated by using the absolute values of the differences.

[12] "Water Environment Improvement Project" means a project that conserve or recover the functions of agricultural water supply facilities, such as dams, reservoirs, and provide the waterside spaces to be rest and recreation spaces for people by maintaining in consideration of waterside accessibility and ecosystem. It is a project by ministry of agriculture, forestry and fisheries of Japan.

7 Conclusion

In this paper, I recommend the second nonparametric estimator in the double-bound format, which is represented by a closed form, and compare the performances of this estimator and the Turnbull estimator through Monte Carlo experiments. We can see from the experimental results that the Turnbull estimator is better than the second estimator from the viewpoint of estimation efficiency. However, since the performances are very close, we can conclude that the second estimator is available as a "convenient method". Also it is found that the even sub-sample design is desirable in the nonparametric estimation for the double-bound format.

References

[1] Alberini, A., Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound and Bivariate Models, *Journal of Environmental Economics and Management*, Vol.28, 1995, pp.287-306.

[2] Ayer, M., H.D.Brunk, G.M.Ewing, and E.Silverman, An Empirical Distribution Function for Sampling with Incomplete Information., *Annals of Mathematical Statistics*, Vol.26, 1955, pp.641-647.

[3] Carson, R.T. and D. Steinberg, Experimental Design for Discrete Choice Voter Preference Surveys, in *1989 Proceeding of the Survey Methodology Section of the American Statistical Association.* Washington, D.C.: American Statistical Association, 1990.

[4] Carson, R.T., L.Wilks, and D.Imber, Valuing the Preservation of Australia's Kadadu Conservation Zone, *Oxford Economic Papers*, Vol.46, 1994, pp.727-749.

[5] Cummings, R., G. Harrison, and E. Rutström, Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?, *American Economic Review*, Vol.85, 1995, pp.260-266.

[6] Hanemann, W. M. and B. Kanninen, The Statistical Analysis of Discrete-Response CV Data, In I. Bateman and K. G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, 1999, pp.302-441.

[7] Härdle, W., *Smoothing Techniques with Implementation in S*, Springer-Verlag, 1991.

[8] Hoehn, J. P. and A. Randall, A Satisfactory Benefit Cost Indicator from Contingent Valuation, *Journal of Environmental Economics and Management*, Vol.14, 1987, pp.226-247.

[9] Kriström, B., A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies, *Land Economics*, Vol.66, 1990, pp.135-139.

[10] McFadden, D., Contingent Valuation and Social Choice, *American Journal of Agricultural Economics*, Vol.76, 1994, pp.689-708.

[11] Terawaki, T., Toshikinko-Nogyo no Gaibu-Keizaikoka no Keisoku: Nidankai-Nishisentaku CVM niokeru WTP-Bunpu no Nonparametric-Suitei (Measuring the Positive Externality of Agricultural Production in Sub-urban Region: Nonparametric Estimation of WTP Distribution in Double-Bound Dichotomous Choice CVM), *Nogyo Keizai Kenkyu (Japanese Journal of Rural Economics)*, Vol.69, 1998, pp.201-212.

[12] Turnbull, B.W., Nonparametric Estimation of Survivorship Function with Doubly Censored Data, *Journal of the American Statistical Association*, Vol.69, 1974, pp.169-173.

[13] Turnbull, B.W., The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data, *Journal of the Royal Statistical Society B*, Vol.38, 1976, pp.290-295.