Introduction to contingent valuation using Stata

Alejandro Lopez-Feldman

Centro de Investigación y Docencia Económicas (Cide)

2012

Online at http://mpra.ub.uni-muenchen.de/41018/
MPRA Paper No. 41018, posted 4. September 2012 19:36 UTC
Introduction to Contingent Valuation using Stata*

Alejandro López-Feldman
Centro de Investigación y Docencia Económicas, CIDE
alejandro.lopez@cide.edu

1 Introduction

Cost benefit analysis is a key input for the ex-ante evaluation of public projects and policies. An ideal cost-benefit analysis incorporates all the social costs and benefits of a project for all members of a society. Boardman et al. (2006). One of the big challenges to achieve this is the need to place monetary values on non-marketed goods and services. The objective of this chapter is to provide the reader with the basic tools to obtain estimates of these values using the contingent valuation method. A basic introduction to the method is provided but the focus of the chapter is the use of Stata to do basic empirical analysis. The use of the doubleb command is illustrated with an example.

2 Contingent Valuation

Valuation methods for non-marketed goods can be divided in direct and indirect methods. With indirect methods the estimations are based on the observed behaviour of individuals in the market of a good or service related to the one of interest. An example of this is the estimation of the use value of the services provided by a national park using information on the costs that travellers must incur in order to get to that place. Among indirect methods we have the travel cost method, hedonic pricing and averting behaviour and defensive expenditures. Direct methods, on the other hand, try to elicit information about the value of the non-marketed good or service directly.

---

*Taken from the book "Aplicaciones en Economía y Ciencias Sociales con Stata" (forthcoming) and translated with permission of the publisher, Stata Press (College Station, TX: StataCorp).
from the individual. Among these methods we have contingent valuation and choice modelling.\(^1\)

Contingent valuation implies asking to a sample of the population about their willingness to pay (WTP) for the provision of a given good or service.\(^2\) The name of the method comes from the fact that the elicited values are contingent to the hypothetical scenario that is presented to those being interviewed Portney (1994). This is a very flexible method since one can obtain estimations for public policies or projects that have not been implemented. Furthermore, with this method it is possible to obtains estimates of non-use values.\(^3\)

Originally contingent valuation was basically an intellectual exercise with limited practical relevance. That changed when, by request of the State of Alaska, Carson et al. (1992) did a contingent valuation study to get an estimate of the non-use value loss associated with the Exxon Valdez oil spill. Not long before that, the US government had approved a federal act in which contingent valuation was accepted as a valid method to measure the loss associated with environmental disasters. After the study of Carson et al. (1992) Exxon and other companies sponsored a series of research efforts looking to discredit the method in order to avoid its future use. As a response to this the National Oceanic and Atmospheric Administration (NOAA) commissioned a group of experts to do a report in the validity of contingent valuation. The panel concluded that contingent valuation was capable of generating estimates reliable enough to be used in court and made some recommendations and guidelines for the application of contingent valuation studies (Arrow et al., 1993).

2.1 Contingent valuation with dichotomous choice questions

The focus of the rest of this chapter is on how to econometrically analyse the data obtained from a contingent valuation survey. Nonetheless, it is

\(^1\)For more details on the different kind of direct and indirect methods see Hanley et al. (2007).

\(^2\)An alternative is to ask about the willingness to accept, nonetheless, in practice WTP is used more frequently.

\(^3\)Non-use values are those that the individual obtain without the direct consumption or use of the resource. Examples of this kind of value are existence values (e.g., the existence of polar bears in the wilderness provide utility or satisfaction to some individuals even though they have never seen them in their natural habitats nor they have plans to do so) and bequest values (the value comes from knowing that the good or service will be available for future generations). Direct methods are the only ones capable of capturing these kind of values.
important to emphasize that, as concluded by the panel commissioned by NOAA, the design of the questionnaire and its application is the fundamental part of any contingent valuation study.\textsuperscript{4}

Generally speaking there are three ways in which WTP can be elicited using contingent valuation. The first one is via open-ended questions. In this case the individual is asked how much is he/she willing to pay for a good or service that has been previously described along with a hypothetical scenario. Another approach is to use payment cards; the individuals are presented with a series of amounts for possible payments and they choose the one that is closer to their individual valuation. The last approach is to use dichotomous choice questions. In the simplest case the individual is asked (after the description of a hypothetical scenario): will you be willing to pay \(X\), yes or no? This last method, which was the one mentioned as the most adequate of the three by the NOAA panel, is the one discussed in this chapter.

2.2 Econometric estimation of the dichotomous model

The information that is directly elicited from individual \(i\), when a contingent valuation questionnaire is applied using the dichotomous choice model, is simply a dichotomous answer (\(y_i = 0\) if the individual answers no and \(y_i = 1\) if the answer is yes), given a question about paying a previously determined amount (\(t_i\), that varies randomly across individuals). It is possible to estimate the WTP assuming that it can be modelled as the following linear function:

\[
WTP_i(z_i, u_i) = z_i\beta + u_i
\]  

(1)

where \(z_i\) is a vector of explanatory variables, \(\beta\) is a vector of parameters and \(u_i\) is an error term. It is expected that the individual will answer yes when his WTP is greater than the suggested amount, i.e., when \(WTP_i > t_i\). In that case, the probability of observing a positive response given the values of the explanatory variables is given by:

\[
Pr(y_i = 1|z_i) = Pr(WTP_i > t_i) = Pr(z_i\beta + u_i > t_i) = Pr(u_i > t_i - z_i\beta)
\]

If we assume that $u_i \sim N(0, \sigma^2)$ we have that: \(^5\)

$$Pr(y_i = 1|z_i) = Pr\left(v_i > \frac{t_i - z'_i \beta}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{t_i - z'_i \beta}{\sigma}\right)$$

$$Pr(y_i = 1|z_i) = \Phi\left(z'_i \frac{\beta}{\sigma} - t_i \frac{1}{\sigma}\right) \quad (2)$$

where $v_i \sim N(0, 1)$ and $\Phi(x)$ is the standard cumulative normal. This is very similar to what is traditionally known as the probit model. The difference with the traditional probit model is that in this case in addition to the explanatory variables we have the variable $t_i$.

There are two ways in which one could estimate this model. The first one is to use Equation (2) and maximum likelihood estimation solving for $\beta$ and $\sigma$. \(^6\) The other option is to directly use the `probit` command available in Stata. The probit model assumes that variance is equal to one since in the traditional case there is not enough information available to estimate that parameter. In this case that assumption is not necessary as we have the additional variable $t_i$. \(^7\) Therefore, we can use the `probit` command in Stata, including $t_i$ as an additional explanatory. Doing so we obtain estimates of $\beta/\sigma$ and $-1/\sigma$. That is to say, the results that we get from the `probit` command are: $\hat{a} = \frac{\hat{\beta}}{\hat{\sigma}}$ (the vector of coefficients associated to each one of the explanatory variables) and $\hat{d} = -\frac{1}{\hat{\sigma}}$ (the coefficient for the variable capturing the amount of the bid).

Until now we have concentrated in the estimation of the parameters of the model. That information can be valuable in itself, nevertheless, what we actually want is to have an estimate of the willingness to pay. Starting with the normality assumption and using Equation (1) we have that the expected value for the willingness to pay is given by: $E(DAP|z_i, \beta) = z'_i \beta$.

Although we do not know the true value of $\beta$ we can get a consistent estimate for that vector using $\hat{a}$ and $\hat{d}$ (estimated with the `probit` command).

---

\(^5\)Other assumptions are possible for the distribution of the error term. This will lead to alternative econometric models, see Haab & McConnell (2003) for more details.

\(^6\)The Stata command `singleb`, created by the author of this chapter, is an option to apply that procedure in Stata. The example presented in section 2.3 is easy to replicate using `singleb`.

\(^7\)See Cameron & Trivedi (2005) page 476 for a more complete explanation about the identification conditions that are relevant for this case. Cameron & James (1987) were the first ones to propose the use of traditional probit computational routines to estimate the WTP in this context.
What we have then is that $\hat{\beta} = -\hat{\alpha}$. Given this, there are different ways in which we can estimate WTP depending on the values that we give to the vector $z$. Some options are to estimate the WTP for every individual, the WTP for individuals with certain characteristics and the WTP using the average of the explanatory variables. In general what we have is:

$$E(WTP|\tilde{z}, \beta) = \tilde{z}' \left[ -\frac{\hat{\alpha}}{\delta} \right]$$

(3)

where $\tilde{z}'$ is a vector with the values of interest for the explanatory variables (i.e., the value for each individual, the value for a certain group or the average).

### 2.3 Dichotomous model using Stata

A data set for a natural reserve in Portugal is used to illustrate the estimation of willingness to pay using the dichotomous model in Stata. The data set captures willingness to pay to avoid the development of commercial and tourist infrastructure inside the park. The questionnaire used dichotomous questions with follow-up, that is to say, two contingent valuation questions were asked to each individual (this method is described in more detail in section 2.4). Table 1 presents the definition of some of the variables included in the data.

One of the first things that we want to look from this data set is the distribution of the amount of the initial bid (i.e., $t_i$).

```
. tab bid1

<table>
<thead>
<tr>
<th>bid1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>76</td>
<td>24.36</td>
<td>24.36</td>
</tr>
<tr>
<td>12</td>
<td>77</td>
<td>24.68</td>
<td>49.04</td>
</tr>
<tr>
<td>24</td>
<td>82</td>
<td>26.28</td>
<td>75.32</td>
</tr>
<tr>
<td>48</td>
<td>77</td>
<td>24.68</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 312 | 100.00 |
```

We have a total of 312 observations divided in four groups, with approximately the same number of individuals on each one of them. Some modifications to the data set are needed before proceeding with the analysis of the data and the estimation of the WTP.

---


9For more details on the study see Nunes & Schokkaert (2003)).
**Table 1:**

<table>
<thead>
<tr>
<th>Name of the variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bid1</td>
<td>initial amount (bid) in euros</td>
</tr>
<tr>
<td>bidh</td>
<td>high bid in euros</td>
</tr>
<tr>
<td>bidl</td>
<td>low bid in euros</td>
</tr>
<tr>
<td>nn</td>
<td>= 1 if the answer to the willingness to pay questions was no, no</td>
</tr>
<tr>
<td>ny</td>
<td>= 1 if the answer to the willingness to pay questions was no, yes</td>
</tr>
<tr>
<td>yn</td>
<td>= 1 if the answer to the willingness to pay questions was yes, no</td>
</tr>
<tr>
<td>yy</td>
<td>= 1 if the answer to the willingness to pay questions was yes, yes</td>
</tr>
<tr>
<td>depvar</td>
<td>indicator variable with the following structure (=1 if nn=1, =2 if ny=1, =3 if yn=1 and =4 if yy=1)</td>
</tr>
<tr>
<td>age</td>
<td>age in 6 categories (&lt; 29, 29 – 39, 40 – 49, 50 – 59, 60 – 69, &gt; 69)</td>
</tr>
<tr>
<td>female</td>
<td>= 1 if the individual is a female</td>
</tr>
</tbody>
</table>

* We generate a variable that indicates the answer to the first question
  . generate answer1 = 0
  . replace answer1 = 1 if depvar ==3 | depvar ==4
  (171 real changes made)

* Now we put a label to each value of the variable answer1
  . label define dummy 0 "No" 1 "Yes"
  . label values answer1 dummy

Now that we have information about who gave a positive response to the first question we can calculate the fraction of respondents that answered yes to that question.

```
. tab answer1  

<table>
<thead>
<tr>
<th>answer1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>141</td>
<td>45.19</td>
<td>45.19</td>
</tr>
<tr>
<td>Yes</td>
<td>171</td>
<td>54.81</td>
<td>100.00</td>
</tr>
</tbody>
</table>
```

Almost 55% of those interviewed answered yes to the first contingent valuation question. An important aspect to check when using contingent valuation data is that individuals should be sensible to the bid amount, that is to say, we expect that as the bid amount goes up the proportion of
individuals that give a positive answer goes down. Let’s see what happens with the data used in this example.

```
.tabulate answer1 bid1, column nofreq

<table>
<thead>
<tr>
<th>answer1</th>
<th>bid1</th>
<th>6</th>
<th>12</th>
<th>24</th>
<th>48</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>34.21</td>
<td>44.16</td>
<td>48.78</td>
<td>53.25</td>
<td>45.19</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>65.79</td>
<td>55.84</td>
<td>51.22</td>
<td>46.75</td>
<td>54.81</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
```

As expected the proportion of positive answers goes down as the bid amount goes up. Now we proceed to econometrically estimate willingness to pay using the `probit` command.

```
.probit answer1 bid1
Iteration 0:  log likelihood = -214.81738
Iteration 1:  log likelihood = -212.39348
Iteration 2:  log likelihood = -212.39341
Iteration 3:  log likelihood = -212.39341

Probit regression                  Number of obs = 312
                                  LR chi2(1) = 4.85
                                  Prob > chi2 = 0.0277
                                  Log likelihood = -212.39341  Pseudo R2 = 0.0113

| answer1 | Coef.  | Std. Err. |  z     | P>|z|   | [95% Conf. Interval] |
|---------|--------|-----------|-------|-------|----------------------|
| bid1    | -.0098408 | 0.0044791 | -2.20 | 0.028 | -.0186197 to -0.0010619 |
| _cons   | .344227 | 0.1244284 | 2.77  | 0.006 | .1003517 to .5881022  |
```

The first thing that we can observe with these results is that the bid variable is statistically significant and that as the bid goes up the probability of a positive answer goes down. To calculate willingness to pay we use Equation (3). In this case in which no explanatory variables are included (recall that in this model $t_i$ is not considered strictly as an explanatory variable), we have that $\alpha$ is a scalar and $z$ is equal to one, therefore the formula is simply $-\hat{\alpha}/\hat{\delta}$.

```
.nlcom (WTP: -_b[_cons]/_b[bid1]), noheader

| answer1 | Coef.  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---------|--------|-----------|-------|-------|----------------------|
| WTP     | 34.97942 | 9.152495 | 3.82  | 0.000 | 17.04086 to 52.91798 |
```

When no control variables are included in the estimation we have that the average WTP is close to 35 euros.\(^{10}\)

\(^{10}\)Loomis \textit{et al.} (2000) present a good example of how to estimate the value of an
The next example includes the variables age and female.

```
. probit answer1 bid1 age female
Iteration 0: log likelihood = -214.81738
Iteration 1: log likelihood = -194.89582
Iteration 2: log likelihood = -194.87357
Iteration 3: log likelihood = -194.87357
Probit regression Number of obs = 312
LR chi2(3) = 39.89
Prob > chi2 = 0.0000
Log likelihood = -194.87357 Pseudo R2 = 0.0928

               Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------- -------- -------------- ------- -------- ------------------
bid1        -.0110408   .0046774   -2.36   0.018    -.0202083    -.0018733
age         -.253761   .0488002   -5.20   0.000    -.3494076    -.1581144
female      -0.3478191  .149391    -2.33   0.020    -.64062    -.0550181
_cons        1.339737   .2178649    6.15   0.000     .91273      1.766745
```

All the variables included in the model are statistically significant. Females are less likely to give a positive response, the same is true for older people.

Expanding Equation (3) what we have is $-(\hat{a}_0 + \hat{a}_{age} + \hat{a}_{female})/\hat{\delta}$, where instead of age and female we need to use the values of interest for each variable. Two possibilities are illustrated: using the mean values for the explanatory variables and estimate willingness to pay with certain characteristics.

```
* First get the mean values and create a scalar with that information
* summarize age, meanonly
* scalar age_m = r(mean)
* summarize female, meanonly
* scalar female_m = r(mean)
* WTP for mean values
  nlcom (WTP: -( _b[_cons] + age_m * _b[age] + female_m * _b[female]) / _b[bid1]), noheader
```

```
          WTP
Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------- -------- -------------- ------- -------- ------------------
WTP       34.2614    8.291197    4.13   0.000    18.01095    50.51185
```

The result (34 euros) shows that in this case WTP does not change too much when including control variables evaluated at their mean values. Now we estimate WTP separately for male and females that have between 40 and 49 year of age (category 3 of the age variable).

```
* WTP for males in age group 3
  nlcom (WTP: -( _b[_cons] + 3 * _b[age]) / _b[bid1]), noheader
```

environmental service for a given population using as one of the inputs the econometric estimation of WTP.
| answer1 | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|---------------------|
| WTP     | 52.39245 | 16.40545  | 3.19  | 0.001 | 20.23836 84.54655 |

* WTP for females in age group 3

```
. nlcom (WTP: -(_b[_cons]+3*_b[age]+_b[female])/_b[bid1]), noheader
```

| answer1 | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|---------------------|
| WTP     | 20.88937 | 9.000112 | 2.32  | 0.020 | 3.249474 38.52926 |

### 2.4 Contingent valuation using dichotomous questions with follow-up

A problem with the method that was just described is that each individual provides very few information with respect to his willingness to pay. Let’s see this with an example. Assume that an individual is asked if he is willing to pay an amount \( t_i \) for a given change in the provision of a public good. If the individual answers no then we can infer that \( 0 \leq WTP < t_i \). If he answers yes then \( t_i \leq WTP < \infty \). This implies that in order to obtain accurate estimations of WTP relatively large samples are needed. Hane
mann et al. (1991) suggest an alternative to improve the efficiency of the estimation. This alternative is know as dichotomous question with follow-up or double-bounded model. In this case, a follow-up dichotomous question is asked after the first dichotomous choice question. If the individual answers yes to the first question then he is asked about his WTP for a higher amount. If he answers no to the first question then a lower amount is offered. This implies that the second question is endogenous in the sense that the amount asked depends on the answer obtained for the first question (which is exogenous). With this method we have two answers for each individual, which provides us with more information but at the same time makes the econometric estimation slightly more complicated than before. Let’s look more carefully at the kind of information that is gathered with this kind of questionnaire before describing the econometric model.

Let’s call the first bid amount \( t^1 \) and the second one \( t^2 \) (for simplicity we skip the sub-index \( i \)), then each individual will be in one of the following categories:

1. The individual answers yes to the first question and no to the second, then \( t^2 > t^1 \). In this case we can infer that \( t^1 \leq WTP < t^2 \).
2. The individual answers yes to the first question and yes to the second,
then $t^2 \leq WTP < \infty$.

3. The individual answers no to the first question and yes to the second, then $t^2 < t^1$. In this case we have that $t^2 \leq WTP < t^1$.

4. The individual answers no to the first and second questions, then we have that $0 < WTP < t^2$.

In cases 1 and 3 we have well defined intervals for the willingness to pay for each individual, this was not possible with the method described in the previous section. The intervals for cases 2 and 4 are similar to what we get using a single question but in this case $t^2$ is closer to the true value of the willingness to pay than $t^1$. In this sense the dichotomous choice model with follow-up provides more information than the simpler format with one question.

### 2.5 Econometric estimation using the double-bounded or interval data model

The method known as the double-bounded or interval data model allows the efficient use of the data to estimate willingness to pay (under the assumption that there is a single valuation function behind both answers).\(^\text{11}\) Let’s define $y^1_i$ and $y^2_i$ as the dichotomous variables that capture the response to the first and second closed questions, then the probability that an individual answers yes to the first question and no to the second can be expressed as $Pr(y^1_i = 1, y^2_i = 0 | z_i) = Pr(s, n)$ (where to simplify notation the right hand side of the expression omits the fact that the probability is conditional on the values of the explanatory variables). Given this and under the assumption that $WTP(z_i, u_i) = z_i'\beta + u_i$ and $u_i \sim N(0, \sigma^2)$, we have that the probability of each one of the three cases is given by:

1. $y^1_i = 1$ and $y^2_i = 0$.

   $$Pr(s, n) = Pr(t^1 \leq WTP < t^2)$$

   $$= Pr(t^1 \leq z_i'\beta + u_i < t^2)$$

   $$= Pr\left(\frac{t^1 - z_i'\beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{t^2 - z_i'\beta}{\sigma}\right)$$

   $$= \Phi\left(\frac{t^2 - z_i'\beta}{\sigma}\right) - \Phi\left(\frac{t^1 - z_i'\beta}{\sigma}\right)$$

\(^\text{11}\)Cameron & Quiggin (1994) y Haab & McConnell (2003) discuss some situation in which the assumption made here might be problematic and suggest alternative estimation methods.
where the last expression follows from \( Pr(a \leq X < b) = F(b) - F(a) \). Therefore, using symmetry of the normal distribution we have that:

\[
Pr(s, n) = \Phi \left( z_i' \beta - \frac{t^1}{\sigma} \right) - \Phi \left( z_i' \beta - \frac{t^2}{\sigma} \right) \tag{4}
\]

2. \( y_i^1 = 1 \) and \( y_i^2 = 1 \).

\[
Pr(s, s) = Pr(WTP > t^1, WTP \geq t^2) = Pr(z_i' \beta + u_i > t^1, z_i' \beta + u_i \geq t^2)
\]

Using Bayes rule, which says that \( Pr(A, B) = Pr(A|B) \ast Pr(B) \), we have:

\[
Pr(s, s) = Pr(z_i' \beta + u_i > t^1 | z_i' \beta + u_i \geq t^2) \ast Pr(z_i' \beta + u_i \geq t^2)
\]

Here by definition \( t^2 > t^1 \) and then \( Pr(z_i' \beta + u_i > t^1 | z_i' \beta + u_i \geq t^2) = 1 \) which implies:

\[
Pr(s, s) = Pr(u_i \geq t^2 - z_i' \beta) = 1 - \Phi \left( \frac{t^2 - z_i' \beta}{\sigma} \right)
\]

so by symmetry we have:

\[
Pr(s, s) = \Phi \left( z_i' \beta - \frac{t^2}{\sigma} \right) \tag{5}
\]

3. \( y_i^1 = 0 \) and \( y_i^2 = 1 \).

\[
Pr(s, n) = Pr(t^2 \leq WTP < t^1) = Pr(t^2 \leq z_i' \beta + u_i < t^1)
\]

\[
= Pr \left( \frac{t^2 - z_i' \beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{t^1 - z_i' \beta}{\sigma} \right)
\]

\[
= \Phi \left( \frac{t^1 - z_i' \beta}{\sigma} \right) - \Phi \left( \frac{t^2 - z_i' \beta}{\sigma} \right)
\]

\[
Pr(s, n) = \Phi \left( z_i' \beta - \frac{t^2}{\sigma} \right) - \Phi \left( z_i' \beta - \frac{t^1}{\sigma} \right) \tag{6}
\]
4. $y_1^i = 0$ and $y_2^i = 0$.

$$
Pr(n, n) = Pr(WTP < t^1, WTP < t^2) \\
= Pr(z_i^1 \beta + u_i < t^1, z_i^2 \beta + u_i < t^2) \\
= Pr(z_i^1 \beta + u_i < t^2) \\
= \Phi \left( \frac{t^2 - z_i^1 \beta}{\sigma} \right)
$$

$$
Pr(n, n) = 1 - \Phi \left( \frac{z_i^1 \beta - t^2}{\sigma} \right) \quad (7)
$$

Contrary to Section 2.2 where we can use the probit model to estimate willingness to pay, equations (4) to (7) do not correspond directly to a pre-existent model. One way to proceed with the estimation is to construct a likelihood function to directly obtain estimates for $\beta$ and $\sigma$ using maximum likelihood estimation. The function that needs to be maximized in order to find the parameters of the model is:

$$
\sum_{i=1}^{N} \left[ d_{in}^{sn} \ln \left( \Phi \left( \frac{z_i^1 \beta - t^1}{\sigma} \right) - \Phi \left( \frac{z_i^2 \beta - t^2}{\sigma} \right) \right) + d_{ns}^{ss} \ln \left( \Phi \left( \frac{z_i^2 \beta - t^2}{\sigma} \right) \right) \\
+ d_{ns}^{ns} \ln \left( \Phi \left( \frac{z_i^1 \beta - t^2}{\sigma} \right) - \Phi \left( \frac{z_i^2 \beta - t^1}{\sigma} \right) \right) + d_{nn}^{nn} \ln \left( 1 - \Phi \left( \frac{z_i^2 \beta - t^2}{\sigma} \right) \right) \right]
$$

where $d_{in}^{sn}, d_{i}^{ss}, d_{i}^{ns}, d_{nn}^{nn}$ are indicator variables that take the value of one or zero depending on the relevant case for each individual, that is to say, a given individual contributes to the logarithm of the likelihood function in only one of its four parts. Contrary to what happens to the approach described in Section 2.2, here we obtain directly $\hat{\beta}$ y $\hat{\sigma}$. Once we have this information we can estimate WTP as we did it before.

### 2.6 Example of the double-bounded model using Stata

The command `doubleb`, created by the author of this chapter, allows the direct estimation of $\beta$ and $\sigma$ using maximum likelihood. To illustrate the

---

12 This model can also be seen as a modified ordered probit model (see Verbeek, 2008).

13 The same is true when we use the Stata command `singleb` instead of `probit`.

14 To install the command `doubleb` type in Stata `findit doubleb`, then click in the link and follow the installation instructions.
use of the command we will use the same data of the previous example. Before starting with the estimation we need to modify the data so it has the structure used by the `doubleb` command.\(^\text{15}\)

\[
\begin{align*}
\text{. * We generate a variable that captures the response to the second question} \\
\text{. generate answer2 = 0} \\
\text{. replace answer2 = 1 if depvar ==2 | depvar ==4} \\
(76 real changes made) \\
\text{. * We generate a variable for the second amount} \\
\text{. gen bid2 = .} \\
(312 missing values generated) \\
\text{. replace bid2 = bidh if answer1 ==1} \\
(171 real changes made) \\
\text{. replace bid2 = bid1 if answer1 ==0} \\
(141 real changes made)
\end{align*}
\]

Now that we have a variable that captures the answer to the second question as well as the second bid amount that was actually offered \((t^2)\), we can estimate the econometric model.

\[
\begin{align*}
\text{. * Model with no control variables} \\
\text{. doubleb bid1 bid2 answer1 answer2} \\
\text{initial: log likelihood = -\text{<inf>} (could not be evaluated)} \\
\text{feasible: log likelihood = -940.87306} \\
\text{rescale: log likelihood = -444.64525} \\
\text{rescale eq: log likelihood = -409.27306} \\
\text{Iteration 0: log likelihood = -409.27306} \\
\text{Iteration 1: log likelihood = -409.00743} \\
\text{Iteration 2: log likelihood = -409.00449} \\
\text{Iteration 3: log likelihood = -409.00449} \\
\text{Number of obs = 312} \\
\text{Wald chi2(0) = .} \\
\text{Log likelihood = -409.00449} \\
\text{Prob > chi2 = .}
\end{align*}
\]

\[
\begin{array}{lcccc}
\text{Coef.} & \text{Std. Err.} & z & P>|z| & [95\% \text{ Conf. Interval}] \\
\hline
\text{Beta} & \_cons & 18.73884 & 2.496957 & 7.50 & 0.000 & 13.84489 & 23.63278 \\
\text{Sigma} & \_cons & 38.61272 & 2.933311 & 13.16 & 0.000 & 32.86354 & 44.36191 \\
\end{array}
\]

\[
\begin{align*}
\text{First-Bid Variable:} & \quad \text{bid1} \\
\text{Second-Bid Variable:} & \quad \text{bid2} \\
\text{First-Response Dummy Variable:} & \quad \text{answer1} \\
\text{Second-Response Dummy Variable:} & \quad \text{answer2}
\end{align*}
\]

Since the `doubleb` command directly estimates \(\hat{\beta}\) the WTP formula is simply \(\tilde{\varepsilon}'\beta\). Therefore, in this case (with no control variables) WTP is simply the constant and is approximately equal to 19 euros. The next step is

\(^{15}\text{The syntax requires a variable for } t^1, \text{ one for } t^2 \text{ and a dummy for each one of the dichotomous questions. The names of the variables are not relevant but the order needs to be the same as in the example.}\)
to estimate the same WTP versions (including control variables) as in the previous example.

. * Model with explanatory variables
. doubleb bid1 bid2 answer1 answer2 age female
initial: log likelihood = -<inf> (could not be evaluated)
feasible: log likelihood = -940.87306
rescale: log likelihood = -444.64525
rescale eq: log likelihood = -394.56437
Iteration 0: log likelihood = -394.5571
Iteration 1: log likelihood = -394.5571
Iteration 2: log likelihood = -394.5571
Iteration 3: log likelihood = -394.5571
Iteration 4: log likelihood = -394.5571
Number of obs = 312
Wald ch2(2) = 26.28
Log likelihood = -394.5571 Prob > chi2 = 0.0000

| Beta              | Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------------------|-------|-----------|-------|-------|----------------------|
| age               | -8.047011 | 1.639399  | -4.91 | 0.000 | -11.26017 -4.833848 |
| female            | -6.237376 | 4.81779  | -1.29 | 0.195 | -15.68007 3.205319  |
| _cons             | 46.35356  | 5.83763   | 7.94  | 0.000 | 34.91202 57.79511  |

| Sigma             | Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------------------|-------|-----------|-------|-------|----------------------|
| _cons             | 36.90406 | 2.776473 | 13.29 | 0.000 | 31.46227 42.34585  |

First-Bid Variable: bid1
Second-Bid Variable: bid2
First-Response Dummy Variable: answer1
Second-Response Dummy Variable: answer2

. * WTP for mean values
. nlcom (WTP:(_b[_cons]+age_m*_b[age]+female_m*_b[female])), noheader

| Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|-----------|-------|-------|----------------------|
| WTP   | 18.52186  | 2.425411 | 7.64  | 0.000 | 13.76814 23.27558  |

The WTP evaluated using the average values for the explanatory variables is equal to 18.52. In this case the gender dummy is not statistically significant so estimating WTP separately for males and females might not be very relevant. Nevertheless, the results for that estimation are:

. * WTP for males in age category 3
. nlcom (WTP:(_b[_cons]+3*_b[age])), noheader

| Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|-----------|-------|-------|----------------------|
| WTP   | 22.21253  | 3.533588 | 6.29  | 0.000 | 15.28682 29.13824  |

. * WTP for females in age category 3
. nlcom (WTP:(_b[_cons]+3*_b[age]+_b[female])), noheader

| Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|-----------|-------|-------|----------------------|
|       | Coef.   | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|---------|-----------|------|-------|---------------------|
| WTP   | 15.97515| 3.299311  | 4.84 | 0.000 | 9.508623 to 22.44169 |

References


