

# Biases in Willingness-To-Pay measures from Multinomial Logit estimates due to unobserved heterogeneity

Vincent van den Berg<sup>1\*</sup>, Eric Kroes<sup>1,2</sup>, Erik Verhoef<sup>1,3</sup>

*1 Department of Spatial Economics, VU University, Amsterdam, The Netherlands*

*2 Significance BV, Leiden, The Netherlands*

*3 Tinbergen Institute, Amsterdam/Rotterdam, The Netherlands*

---

## Abstract

It is a common finding in empirical discrete choice studies that the estimated mean relative values of the coefficients (i.e. WTP's) from multinomial logit (MNL) estimations are different from those from mixed logit estimations, where the mixed logit has the better statistical fit. However, it is less clear under exactly which circumstances such differences arise, whether they are important, and if they can be seen as biases in the WTP estimates from MNL. A well known form of bias is the omitted variable bias. We discuss a number of cases where such an omitted variable problem can occur and argue that this endogeneity problem might be more common in discrete choice studies than is commonly thought. We use datasets created by simulation to test, in a controlled environment, the effects of the different possible sources of bias on the accuracy of WTP's estimated by MNL. We reproduce the known result from earlier Monte Carlo studies, that random heterogeneity in the marginal utilities in itself does not cause biased MNL estimates. However, we find that if two heterogeneous marginal utilities are correlated, that than the WTP's from MNL can be biased. If the correlation between the marginal utilities is negative, than the bias in the MNL estimate is negative, whereas if the correlation is positive the bias is positive.

*Keywords:* Discrete Choice, Biases in WTP Estimates from Multinomial Logit, Correlated Heterogeneous Marginal Utilities, Omitted Variable Bias

---

---

\* Corresponding author: Vincent van den Berg: (email: vberg@feweb.vu.nl)

# 1. Introduction

Multinomial Logit (MNL) is often used in empirical discrete choice studies. However, as for instance Bhat (1998a) notes, if there is heterogeneity in the marginal utilities and/or in the alternative specific constants, than ignoring this, by estimating a MNL, could lead to biased parameter estimates and choice probabilities. With MNL it is only possible to control for observed heterogeneity. For instance, that the marginal utility of a price attribute can depends on income. To control for unobserved heterogeneity, a mixed logit or a probit can be used. However, ignoring heterogeneity by using MNL does not necessarily cause biased estimates. For instance, Train (1998) notes that there is probably no general answer whether or not MNL gives correct estimates when heterogeneity is present. Hence, it remains unclear under which circumstances MNL gives biased mean estimates of the WTP's (i.e. the ratio's of marginal utilities). This paper studies some circumstances under which biases in WTP estimates may occur.

The problem of omitted variables occurs when a variable that influences the choices of individuals is unobserved by the analyst and is correlated with one or a number of explanatory variables. In is hence an endogenous explanatory variable problem. The problem of endogenous variables has long since been recognized. However, controlling for endogeneity is more difficult in discrete choice models, than in linear (regression) models (Louviere et al., 2005).

We also study what happens if marginal utilities are heterogeneous, but one ignores this by estimating an MNL model. We analyse three cases of heterogeneity. The first is symmetric random heterogeneity. The second is asymmetric heterogeneity, where the mean, mode and median of the distributions of the marginal utilities are not equal. Thirdly we study the case when two heterogeneous marginal utilities are correlated.

This paper studies some circumstances under which heterogeneous marginal utilities might cause biased WTP estimates from MNL. It also studies the effect of an omitted variable bias, and discusses several situations in which this bias might occur which we think could be circumstances that can occur in real world empirical studies.

In this paper we only look at the effects of potential problems of MNL on the Willingness-To-Pay (WTP) for an attribute. It would also be possible to study the coefficients themselves, or the resulting elasticities. A problem with using elasticities in this setting is that there is the question of how to aggregate them. Furthermore, even if the estimated coefficients are unbiased, it can still be the case that the choice probabilities, on which MNL elasticities depend, are biased. Hence, using elasticities in complicates the task of comparing the design levels with estimates. A problem with studying coefficients is that the coefficients depend on arbitrary scaling, whereas WTP's do not. Consequently, we only compare the WTP's.

We use datasets created by Monte Carlo simulation. Hereby, we can create datasets in controlled environments, enabling us to test the effects of certain issues (such as heterogeneous marginal utilities) in a clean laboratory-type of setting. Conversely, in real world empirical datasets there are a large number of issues at play, making it difficult to analyse separate issues.

Section 2 first discusses some of the literature. Then Section 3 studies the effect of (non-symmetric) heterogeneity in the marginal utilities and discusses the basic setup of the Monte Carlo simulation, which is also used in the following sections. Section 4 discusses omitted

variable biases. Section 5 studies what happens when two heterogeneous marginal utilities of two attributes are correlated, and finally Section 6 concludes.

## 2. Literature discussion

This section discusses some of the literature. First, a number of studies that use both MNL and mixed logit are discussed. Next, we discuss a number of known sources of bias.

Bhat (1998a) finds that the WTP's from mixed logit are on average slightly larger than with MNL for his dataset. He also notes that the absolute value of the elasticity of the choice probability to "costs" is larger with mixed logit than with MNL. Similarly, Bhat (2000) finds that MNL underestimates the WTP's for *out-of-* and *in-vehicle travel time*. Bhat (1998b) notes that the WTP's for *out-of-* and *in-vehicle travel time* are somewhat smaller with his MNL than with his mixed logit.

Train (1998) finds for his data that the WTP's for the attributes are slightly to substantially larger with mixed logit than with MNL. He also finds that the WTP's from his mixed logit with correlated marginal utilities are smaller than those found by MNL and the mixed logit without free correlation. He also concludes that there probably is no general conclusion whether MNL gives good estimates for the WTP's for a given dataset, that the performance of MNL will be different for each dataset. Rizzi and Ortúzar (2006) find that the Values of Risk Reduction (i.e. the WTP's for Risk) in the three surveys they analyse are somewhat lower with MNL. However, the average WTP's from their mixed logits are within the 95% confidence intervals of the WTP's of the respective MNL's. Hensher, Greene and Rose (2008) find for their combined Stated Preference (SP) and Revealed Preference (RP) dataset that the mean price elasticities with nested logit are for some alternatives higher and for others lower than with their mixed logit.

Van den Berg (2008) finds for his dataset that MNL underestimates the WTP for travel time and over-estimates the WTP's of the usage restrictions on the train trip tickets. Interesting is that he finds that not controlling for the *travel time of the train trip on which the SP experiment was based* causes the largest bias in the estimations. This omitted variable bias is because this background variable positively influences the utility of choosing a train ticket alternative and is correlated with the attributes ticket price and travel time.

The empirical papers often find that MNL estimations give different estimates of the WTP's, than mixed logits. However, there is no clear pattern in whether MNL over- or underestimates WTP's. One study finds an overestimation, a second finds an underestimation and a third finds no real difference. These empirical results are however in contradiction with the results of the theoretical analysis of Horowitz (1980) using simulated datasets. He analysed the performance of MNL when there is random heterogeneity in the marginal utilities. He found that the ratio of the two coefficients he estimates on his simulated datasets (i.e. the WTP) is for all amounts of heterogeneity in the marginal utilities almost identical to the design value. He concluded that the ratio of the coefficients is unbiased when one does not control for response heterogeneity. However, he did find that the MNL choice probabilities will be biased. The question is; why would unobserved heterogeneity cause biased WTP estimates from MNL?

A common cause of bias is correlation between unobserved elements of utility and the explanatory variables (i.e. endogenous variables) (Louviere et al., 2005). This issue is discussed in Section 3. If the data has a panel structure, this can also lead to a bias. Examples of panel discrete choice datasets are RP data with multiple observations per individual over time, or

SP data if the respondents face multiple choice situations. The MNL choice probability formula is based on the assumption that the unobserved elements the utility are Independently and Identically Distributed (IID). Train (2003) shows that with panel data the IID assumption of MNL is violated if there are fixed effects in the unobserved elements that are invariant over time for an individual. This will probably make the standard errors of the estimated coefficients incorrect. However, this should have no effect on the estimated coefficients themselves.

Nevertheless, as Carro (2006) discusses, if the  $\alpha_{iq}$  is correlated with an explanatory variable and if previous choices affect current choices, then the estimated coefficients will be biased. Train (2003) notes that if the unobserved elements are correlated over time, then including dynamics in the observed utility (such as a lagged dependent variable) causes the estimation to be inconsistent, if it does not control for the dynamics in the unobserved elements.

Heckman (1981) comes to the same conclusion using Monte Carlo experiments. He first studies the case of strictly endogenous variables and fixed individual effects. He concludes that in this case a probit estimation works well. However, if a lagged dependent variable also affects the choices, then the probit estimation performed badly. Arellano (2003) notes that this is not surprising, as the same issue arises with linear auto regressive models. In the case of dynamics in the unobserved and in the observed elements of utility, a method is needed that can control for this, such as panel mixed logit or panel probit.

This bias is hence caused by the correlation between the (non-simulated) unobserved elements and the explanatory variable, caused by the dynamic structure of the model. However, it is still not clear why unobserved heterogeneity in itself causes a bias. In this paper we research three forms of possible causes of bias. The first is non-symmetric random heterogeneity in the marginal utilities. The second is an endogenous variable caused by an omitted variable. Thirdly, we study the effect of correlation between two heterogeneous marginal utilities.

### 3. Non-symmetric marginal utilities

This section tests whether (non-symmetric) heterogeneity in the marginal utilities could cause biased mean estimates. The idea is that the non-symmetricity in the heterogeneity might not middle out to the mean of the distribution. For example, because of the non-linearity of the logit probability function.

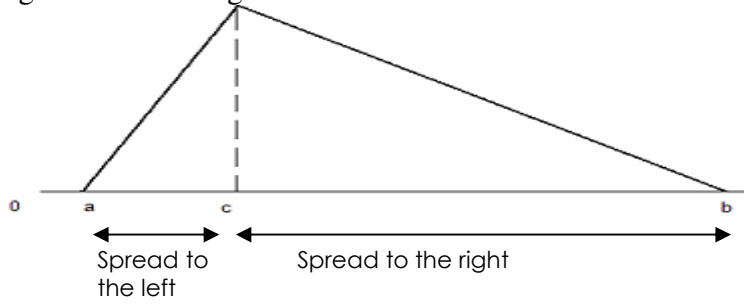
We investigate this through dataset simulations in which we use several different design forms of the random elements of two marginal utilities. Hereby we test if there are differences in the WTP's from MNL as the skewedness of the random elements of the marginal utilities increases. We create 2000 datasets per version of the marginal utilities and then perform MNL estimates on these simulated datasets. The following paragraphs first describes the setup of the created datasets. Thereafter, the results of the simulations of this section are discussed.

We only use triangular distributions for the random elements. Symmetric triangular distributions for random elements were first applied in a discrete choice study by Train (2001). The (symmetric) triangular distribution is increasingly being used in empirical studies using mixed logit (Hensher, Greene and Rose, 2008). Advantages of the triangular distribution are that it is a bounded distribution, thereby preventing excessively large or small marginal utilities, and that it is possible to constrain a marginal utility so that it can never change sign. This last property is useful for attributes for which a positive or negative utility is implausible.

In a more general form, the triangular distribution can also allow for non-symmetric distributions. An example of such a distribution is shown in Figure 1. In this distribution,  $a$  is the

minimum,  $b$  the maximum and  $c$  the mode. Because, it is a non-symmetric distribution the mode is not the same as the mean. The mean of the distribution is given by  $(a+b+c)/3$ . If  $(c-a) < (b-c)$  then the mean is larger than the mode; if  $(c-a) > (b-c)$  it is smaller (Weisstein, 2008).

Figure 1: The triangular distribution



In mixed logit models, the symmetric triangular distribution (i.e.  $(c-a)=(b-c)$ ) is often used. In this case the  $c-a$  is called the *spread*, as it gives the spread of the distribution. To constrain a marginal utility to be positive, the spread is constrained not exceed the mean. A downside of this specification is that it does not allow (the absolute of) the marginal to be larger than twice the absolute of the mean. Thereby, it limits the amount of heterogeneity that can be estimated. Furthermore, the distribution of the random element is constrained to be symmetric.

The symmetric triangular distributed marginal utility of attribute  $k$  for individual  $q$  is given by equation (1).

$$\beta_{kq} = c_k + Spread_k * T_{kq} \quad (1)$$

The  $T_{kq}$  is a random variable with a symmetric triangle distribution, with  $-1 \leq T_{kq} \leq 1$  and a zero mean. It is generated from a standard uniform random variable ( $U_{kq}$ ) by

$$T_{kq} = \begin{cases} \sqrt{2U_{kq}} - 1 & \text{if } U_{kq} \leq 0.5 \\ \sqrt{2(1-U_{kq})} & \text{if } U_{kq} > 0.5. \end{cases}$$

Defining  $d_k = c_k - a_k$  (i.e. the (negative) spread to the left) and  $e_k = b_k - c_k$  (i.e. the spread to the right in Figure 1). Then for the non-symmetric version the marginal utilities are given by (2).

$$\beta_{kq} = \begin{cases} c_k + \sqrt{U_{kq} * (e_k + d_k) * (d_k)} - d_k & \text{if } U_{kq} \leq d_k / (e_k - d_k) \\ c_k + e_k - \sqrt{(1-U_{kq}) * (e_k + d_k) * e_k} & \text{if } U_{kq} > d_k / (e_k - d_k) \end{cases} \quad (2)$$

We now introduce the general layout of the datasets we create. This layout is also used in the later sections. We only use triangular distributions for the marginal utilities. We simulate a choice situation in which there are three alternatives. The utility of alternative  $i$  for individual  $q$  depends on a number of explanatory variables (the vector  $\mathbf{X}_{iq}$ ), the individual parameter vector ( $\beta_q$ ) which is the same for all alternatives and the IID unobserved element  $\varepsilon_{iq}$ . The utility functions for the three alternatives is based on  $U_{iq} = \beta_i^T \mathbf{x}_{iq} + \varepsilon_{iq}$ , and are given by formula (3a-c).

$$U_{1q} = \beta_{1q} * X_{11q} + \beta_{2q} * X_{21q} + \varepsilon_{1q} \quad (3a)$$

$$U_{2q} = \beta_{1q} * X_{12q} + \beta_{2q} * X_{22q} + \beta_{3q} * X_{32q} + \varepsilon_{2q} + ASC_2 \quad (3b)$$

$$U_{3q} = \varepsilon_{3q} + ASC_3 \quad (3c)$$

The first number in subscript behind the X variables ( $k=1,2,3$ ), indicates which X variable is used. The second number indicates the alternative ( $i=1,2,3$ ). Finally, the  $q$  ( $q= 1,2,\dots,N$ ) indicates for which individual the utility is calculated. For easy interpretation we view the  $X_{1iq}$  as the price of alternative  $i$  for individual  $q$ . The (individual) parameters are also called the marginal utilities, because the derivative of the utility to variable  $X_{kiq}$  is  $\beta_{kq}$ . The  $\varepsilon_{iq}$ 's are randomly generated to be Independently and Identically Distributed, with an Extreme Type I distribution form. Discrete choice models use the assumption of utility maximization. Hence, an individual in the simulated datasets chooses the alternative which gives the highest utility. The utility of third alternative is not influenced by any attribute, and can be seen as an "opting out" alternative.

The  $X_{11q}$  and  $X_{12q}$  (i.e. price) variables are created from the variable  $Z_{1q}$  and the  $X_{21q}$  and  $X_{22}$  variables from the variable  $Z_{2q}$ . The  $X_{1iq}$  ( $i= 1$  or  $2$ ) variables are created from  $Z_{1q}$  by a randomly generated difference variable, which can take the values of 10%, 20%, 30% and 40%. Hence,  $Z_{1q}$  times the difference variable gives the difference between the two  $X_{1iq}$  variables. This difference is allocated to an increase relative to  $Z_{1q}$  for  $X_{11q}$  and a decrease for  $X_{12q}$ . This allocation is determined by the standard uniform random variable  $r_{1q}$ . For  $X_{21q}$  and  $X_{22q}$  these variable are difference relative to  $Z_2$  (i.e.  $diff_{2q}$ ) and  $r_{2q}$ . Here  $diff_{2q}$  can take four values, namely 0%, -10%, -20% and -30%. Hence, the  $X_{21q}$  for alternative one is always lower or equal to the  $X_{22q}$  for alternative two. The variables are determined by the formulas (4a-d). The design levels and forms for  $Z_{1q}$  and  $Z_{2q}$  and  $X_{32q}$  are shown in Table 1. Note that  $X_{32q}$  has a uniform distribution, which starts at mean - spread/2 = 5 and ends at mean + spread/2 = 25.

$$X_{11q} = Z_{1q} + Z_{1q} * diff_{1q} * r_{1q}, \quad (4a)$$

$$X_{12q} = Z_{1q} - Z_{1q} * diff_{1q} * (1 - r_{1q}), \quad (4b)$$

$$X_{21q} = Z_{2q} + Z_{2q} * diff_{2q} * r_{2q}, \quad (4c)$$

$$X_{22q} = Z_{2q} - Z_{2q} * diff_{2q} * (1 - r_{2q}). \quad (4d)$$

Table 1: The designs of the variables of the simulated dataset

	Mean	Standard Deviation	Spread	Distribution shape
$Z_{1q}$	10	5	-	Lognormal
$Z_{2q}$	70	40	-	Lognormal
$X_{32q}$	15	-	20	Uniform

The marginal utilities and Alternative Specific Constants (ASC's) are determined by the design values in Table 2. The spreads to the left are determined by the design variable  $Min_k$  and to the right by  $Max_k$ . The spread to the left is given by  $Min_k * mode_k$  and to the right by  $Max_k * mode_k$ . The mode of the distribution is calculated so that given the spreads (as determined by  $Min_k$  and  $Max_k$ ), the mean of the marginal utility is equal to the design mean. Hence, no matter what, the expected value of the marginal utilities is the same for each design version of the datasets, which helps interpreting the results.

Table 2: The designs of the marginal utilities and ASC's

	Mean	Spread to the left	Spread to the right	Distribution shape
$\beta_{1q}$	-0.15	Varies by the design	Varies by the design	Triangular
$\beta_{2q}$	-0.025	Varies by the design	Varies by the design	Triangular
$\beta_3$	-0.02	0	0	Fixed
ASC <sub>2</sub>	-0.4	0	0	Fixed
ASC <sub>2</sub>	-4.5	0	0	Fixed

Table 3 shows that we use 18 different combinations of the *Min\_k* and *Max\_k* design variables. Per version we generate 2000 different datasets, where each dataset has new values for the explanatory variables, marginal utilities and unobserved elements. For each dataset we create 500 individuals, who all face one choice situation. We do on each dataset a MNL estimation in Gauss 6.0 using the maxlik module. The starting values for the coefficients and ASC's were determined randomly within certain bounds, where the bound was different for each coefficient and ASC.

We calculate for each dataset the WTP's for  $X_{2i}$  and  $X_{32}$  (i.e.  $\beta_2/\beta_1$  and  $\beta_3/\beta_1$ ), and then calculate the averages per version, and these are reported in Table 3. We also give the design levels of the WTP's and the average relative sizes of the two. This last figure is defined as mean of the estimated WTP divided by the design WTP.

The design levels of the WTP's are calculated by dividing the dataset mean of  $\beta_{iq}$  by the mean  $\beta_{1q}$ . Given the assumption that the true model is a mixed logit, the coefficients of the MNL estimation should be equal to the means of the randomly heterogeneous marginal utilities. Hence, a WTP from MNL should be equal to the mean of the marginal of some attribute divided by the mean marginal utility of price. This also means that the WTP from an MNL should not be equal to the mean WTP from a mixed logit. The mean WTP of a mixed logit is given by the mean of the marginal utility of some attribute divided by the marginal utility of price. The WTP from an MNL estimated is given by the coefficient of some attribute divided by the coefficient of price. Naturally it is so that  $E(\beta_{kq} / \beta_{price\_q}) \neq E(\beta_{kq}) / E(\beta_{price\_q})$ . This could also explain part of the differences found by the empirical literature between the WTP's of MNL and mixed logit. The absolute of the WTP from a MNL should be smaller than the absolute of the mean WTP from a mixed logit estimated in utility space, if the mean marginal utilities in the mixed logit estimate are equal to the coefficients in the MNL.

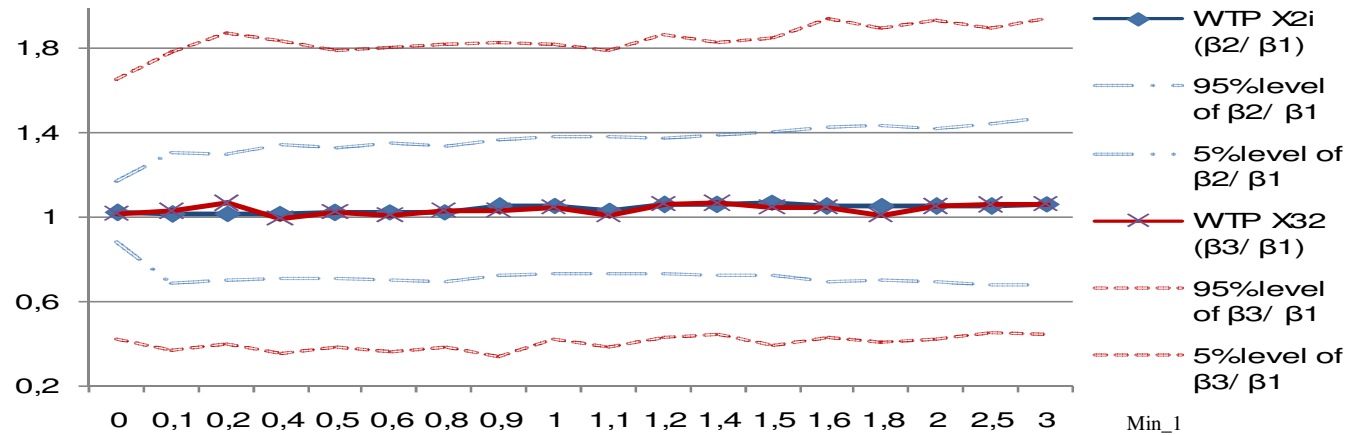
The relative sizes of the estimated WTP's to their design levels are also depicted in Figure 2. This figure also depicts the 90% interval of all the estimates per version of the marginal utilities. Hence, it gives the area within which 90% of the WTP estimates on the simulated datasets lie. This range gives an indication of the accuracy of the simulation and helps in determining whether a mean estimated WTP is really different from its design value.

The first part of Table 3 shows that  $\beta_{2q}$  is given a symmetric triangular distribution. The marginal utility of  $X_{1q}$  (i.e.  $\beta_{1q}$ ) is given an asymmetric distribution. The *spread to the right* is equal to  $0.999 \cdot \text{mode}$  (in all but the first case) and the *spread to the left* varies from 0 to 3. Both marginal utilities remains negative, because the maximum is  $(1-0.999) \cdot \text{mode}_k$ .

Table 3: Results estimations on the datasets with non-symmetric random marginal utilities.

Design variables																		
Min_1	0	0.1	0.2	0.3	0.5	0.6	0.8	0.9	0.999	1	1.1	1.2	1.4	1.5	1.6	2	2.5	3
Min_2	0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Max_1	0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Max_2	0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
WTP variable $X_{2iq}$ ( $\beta_2 / \beta_1$ )																		
Estimated	0.170	0.169	0.169	0.170	0.170	0.170	0.170	0.175	0.176	0.172	0.176	0.176	0.178	0.175	0.176	0.175	0.175	0.177
Design value	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
Relative value	1.018	1.012	1.015	1.017	1.020	1.022	1.023	1.049	1.055	1.032	1.056	1.056	1.067	1.048	1.053	1.050	1.050	1.061
WTP variable $X_{32q}$ ( $\beta_3 / \beta_1$ )																		
Estimated	0.135	0.137	0.142	0.132	0.136	0.134	0.137	0.137	0.140	0.135	0.141	0.142	0.139	0.139	0.134	0.140	0.141	0.141
Design value	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133
Relative value	1.013	1.027	1.065	0.992	1.023	1.008	1.030	1.028	1.047	1.010	1.061	1.064	1.045	1.045	1.006	1.050	1.058	1.060

Figure 2: Relative values estimated WTP's in Table 3 to their design values over the values of  $Min_1$





The table and figure make clear that the skewedness of the distribution of the heterogeneity and the heterogeneity itself have no effect on the mean of the estimated WTP's of the 2000 datasets per version (i.e. per column in Table 3).

For relatively strong asymmetry, the estimated WTP for  $X_2$  ( $\beta_3/\beta_1$ ) is more or less equal to the design value, and for small amounts of asymmetry this is also the case. For the WTP for  $X_{32}$  ( $\beta_3/\beta_1$ ) a similar conclusion holds, with no clear pattern of bias visible due to the skewedness of the distribution. The mean estimated WTP's vary a bit over the table; however there is no clear pattern of effect of (non-symmetric) heterogeneity. Note that the random heterogeneity in the marginal utilities does cause the estimates to have a larger spread and makes the standard errors of the estimates larger.

The conclusion from this exercise is that whether or not the heterogeneity in a marginal utility is symmetric, it need not affect the estimated mean WTP's of an MNL estimation. However, the shape of the random elements of the marginal utilities might affect the shape of the unobserved elements in MNL estimations and thereby cause a violation of the assumption that the unobserved elements all have the same EV1 distribution. In this section we have reproduced the result of Horowitz (1980) that heterogeneity in itself need not cause estimations that ignore the heterogeneity (such as MNL) to give biased WTP estimates.

#### 4. Omitted variables

The previous section found no biases in the WTP's from MNL estimations due to unobserved heterogeneity (both symmetric and not) in marginal utilities. This section studies a second possible source of bias, omitted variable biases. The problem occurs when a variable that influences the choices of individuals is unobserved and is correlated with one or a number of explanatory variables. For linear models the problem can be solved by using an Instrumental Variable (IV) regression. However, for non-linear models, such as the logit and probit models used in discrete choice, the regular IV method can not be used. Then alternative methods such as a "control function" approach or a "fixed effects" method (also referred to as the BLP method (Berry (1994) and Berry, Levinsohn and Pakes (1995))) are needed (Louviere et al., 2005).

Both methods use two-stage estimations to control for the endogeneity. The "fixed effect" method was developed to empirically study (discrete) demand for and supply of differentiated products. Examples of such a structure are automobiles (Berry, Levinsohn & Pakes, 1995), margarine (Petrin & Train, 2004), TV cable / satellite provider (Goolsbee & Petrin, 2004), yogurt and ketchup (Villas-Boas & Winer, 1999). In these cases some explanatory variables are correlated with the unobserved elements. Usually, the studied issue is that the price is correlated with the unobserved quality. Higher quality cars (or ketchups or margarines) are usually more expensive to produce.

The choice of a new auto is a good example to explain the "fixed effect" method by. There are a  $J$  number of different cars and  $N$  individual decision makers. For each different car there is one price and all the other attributes are also the same for all individuals. If one than includes an Alternative Specific Constant for each of the  $J$  alternatives (car types) in a logit or probit, it becomes impossible to include the product attributes, as their effect is captured by the constants. Thereafter, in the second stage the Alternative Specific Constants are regressed on the attributes, to find the effect of the attributes.

The second method is the “control function” approach. This method was developed by Vilas-Boas and Winer (1999) and Blundell and Powell (2004). In the first stage the endogenous explanatory variable is regressed on one or a number of instruments that are uncorrelated with the unobserved element of the second discrete choice stage. Hence, the instruments have the same function as in an IV estimation. Then the errors of this first stage (or a function of them) are added to the second stage, as an approximation of the unobserved element that is correlated with the explanatory variable. The endogenous variable is added without alteration in the second stage (Louviere et al., 2005). If the coefficient of the estimated error from the first stage has a significant effect in the second stage, then there is significant endogeneity (Petrin and Train, 2004). Note that if the error has no effect in the estimation, this does not necessarily mean that there is no endogeneity. It could for instance be the case that the instruments are not strong enough or are correlated with the unobserved element of the second stage.

There is currently a large interest in endogeneity in discrete choice models and it is an active field, see Louviere et al. (2005) for a discussion of the literature. This interest seems logical as endogeneity can cause large biases and can be difficult to control for. The point of this section is not to show that endogeneity causes bias. The goal is to emphasize how common problems with omitted variables can be.

The methods used to control for endogeneity seem to be more difficult to use than the usual IV method used for linear models. Furthermore, there is also the risk that if the methods are not correctly applied (e.g. non-exogenous instrument) that the estimation are still biased, whereas the researcher assumes that it is unbiased. Hence, the best solution seems to be to find data on the omitted variables. Consequently, it seems advisable to put extra effort (and recourses) in acquiring information on (or proxies for) variables that could cause an omitted variable bias. This can prevent the need to use a more difficult and possibly incorrectly applied two stage estimation.

To emphasize how common omitted variable biases can be, we now discuss a number of examples. Our first example of an omitted variable bias comes from Van den Berg (2008), who studies the choice of train tickets using an SP experiment. He finds that not controlling for the background variable *travel time of the trip on which the SP experiment was based* causes positive biases in the coefficients for travel time and price. Thereby a positive bias in the WTP for travel time and a negative bias in the other WTP's result. The probability of choosing a ticket alternative increases with this background variable, as for longer distances there is less competition from other modes. The attribute “travel time” was based on the background variable *travel time of the trip on which the experiment was based* and “ticket price” is a function of travel distance, which is correlated with travel time. Hence, if one does not control for the background variable, the estimated coefficients for “travel time” and “ticket price” also partly measures the effect of the background variable. This type of *distance effect* on the choices and the correlation of this background variable with the attributes can also occur in other situations.

That the strength of competition between modes differs over the length of the trip can occur in many situations. For instance, (in the Netherlands) for short distance the bus often faces little competition from the train, but a lot from biking. Conversely, for middle distance the train is a more serious competitor and the competition from biking is less. Furthermore, it also seems likely that the price and travel time are correlated with trip distance in this and other situations. The distance effect causes an endogeneity problem, just as the endogeneity of price via “quality” in the discussed studies caused an endogeneity problem.

A second example is income. If richer people dislike travelling by public transport and more often buy a first class ticket if they do travel by public transport. Then in this case the preference for travelling by public transport and price are both correlated with income. Hence, if income is unobserved than under the discussed assumptions there can be omitted variable bias.

A third example could be the choice of the road to travel by. Suppose that an individual can either choose to travel by the main road or by a scenic secondary road. Further suppose that the main road has a lower risk of injury and death due to an accident. The final assumption is that there is unobserved heterogeneity in the valuation (i.e. the marginal utility) of scenery. If these assumptions hold than the unobserved element and the attribute risk of the road are correlated, thereby it can cause an omitted variable bias.

A fourth example is the earlier mentioned relation between the price of an attribute and the (unobserved) quality. This endogenous relationship can be very common, as it seems a very natural situation that higher quality products are more desirable and more expensive to make. Furthermore, attribute quality is very difficult to measure and hence often unobserved. Consequently, studies that control for endogeneity most commonly study this relationship.

A last example could be the choice of utility company. In that, companies with better a service, a better helpdesk or who are more flexible in the credit policies have higher costs and hence higher prices. Furthermore, the valuation of these service characteristics might be heterogeneous over the population. Hence, the observed price might be correlated with the unobserved service characteristics of the utility company.

The above examples show that omitted variable biases can have large effects. Furthermore, there can be many causes of such biases. This is why we emphasize the importance of controlling for background variables. However, it can be very difficult to get data on all possible omitted variables. A question on income for example can be difficult to ask in a questionnaire and when asked there is the danger of inaccurate answers. For instance, because people do not want to admit that they have a high or low income. It can also be difficult to find an objective and accurate measure for quality. Furthermore, there are also cases where it is just not possible to directly measure an variable.

The effect of omitted variables is well known. For instance, Wooldridge (2003) discusses the effect of an omitted variable in the case of OLS regressions. Suppose that the true model is

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + v.$$

If than  $x_2$  is omitted in a OLS regression, the estimated coefficient of  $x_1$  is given by

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \text{cov}(x_1, x_2) * \beta_2 / \text{var}(x_1) = \beta_1 + \text{corr}(x_1, x_2) * \beta_2 * \text{var}(x_2).$$

Hence, for a positive covariance between the unobserved element (the omitted variable) and the explanatory variable there is a positive bias in the coefficient, and for negative covariance a negative bias.

Now we present our Monte Carlo study on the effects of two omitted variables on MNL estimations. The datasets have the same basic setup as before. The variables have the same means and standard deviations. The design marginal utilities have the same means, although they are

now all fixed and non-random. Per version of the strength of the relation between the omitted variables and the explanatory variables we created 1000 datasets.

The setup of the utility functions is the same as before. Only now the utility of the third alternative is influenced by the new variables  $X_{43q}$  and  $X_{53q}$ , and determined by equation (3c'). The  $X_{43q}$  is log-normally distributed, with a mean of 80 and a standard deviation of 50. The  $X_{53q}$  is triangularly distributed with a mean of zero and a spread of 15. Their fixed betas ( $\beta_4$  and  $\beta_5$ ) are equal to -0.05 and -0.1. Finally, the design level of the  $ASC\_3$  was -2.5 instead of -4.5. The  $X_{43q}$  variable is correlated with  $Z_{1q}$  (which determines  $X_{1iq}$ ) and  $Z_{2q}$  (which determines  $X_{2iq}$ ). The strength of these relationships are determined by the design scalars  $R_1$  and  $R_2$ . The variable  $Z_{2q}$  is also correlated with  $X_{53q}$ , this relationship is determined by the design scalar  $R_3$ . The  $Z_{1q}$  and  $Z_2$  variables are randomly generated by formulas (5) and (6).

$$U_{3q} = \varepsilon_{3q} + ASC_{3q} + \beta_4 * X_{43q} + \beta_5 * X_{53q} \quad (3c')$$

$$Z_{1q} = \exp(\text{mode}_1 + sd_1 * N_{1q}) + R_1 * (X_{43q} - 80) \quad (5)$$

$$Z_{2q} = \exp(\text{mode}_2 + sd_2 * N_{2q}) + R_2 * (X_{43q} - 80) + R_3 * X_{53q} \quad (6)$$

The  $mode_i$  and  $sd_i$  are set so that the design values for the mean and standard deviation hold when  $R_1$ ,  $R_2$  and  $R_3$  are zero. The  $N_{iq}$  in (5) and (6) is a random variable with a standard normal distribution. The  $X_{1iq}$  and  $X_{2iq}$  are again created from the  $Z_{1q}$  and  $Z_{2q}$  variables by equations (4a-d). In Table 4, the resulting averages of the estimated WTP's for each version are tabulated, when we do not control for  $X_{43q}$  and  $X_{53q}$  in a MNL estimation. Figure 3 shows for Table 4 the relative values of the estimated WTP's of  $X_{2iq}$  to their design levels. Figure 4 does this for  $X_{32q}$ . In the figures the solid lines with markers are the resulting relative values. The dashed lines give the 90% intervals of the relative values, thus the area between which the 90% the estimates lie. This area gives an impression on how certain we can be that the estimated results are different than the design WTP's. The first part of the tables gives the design values of the relation strength, the scalars  $R_1$ ,  $R_2$ , and  $R_3$ . The starting values for the coefficients and ASC's in the estimations were determined randomly within certain bounds.

Table 4 and Figure 3 show that as the relation between  $Z_{2q}$  and the omitted variables (i.e.  $R_2$  and  $R_3$ ) becomes stronger, the WTP of  $X_{2iq}$  becomes smaller. Even for relatively weak relations between  $Z_{2q}$  and the omitted variables (e.g.  $R_2 = 0.08$  and  $R_3 = 0.05$ ) the mean of the estimated WTP's for  $X_{2iq}$  becomes negative. Furthermore, for somewhat stronger relations between  $X_{2iq}$  and the omitted variables ( $R_2 \geq 0.25$  and  $R_3 \geq 0.15$ ) the 90% interval of the estimates is also below zero, indicating that it is rather certain that the estimated WTP's is negative. In these cases the omitted variable bias is larger than the design size of the coefficient of the second attribute.

It is interesting that a negative WTP for the attribute travel time is exactly what Van den Berg (2008) found when he did not control for the effect of the background variable *travel time of the trip the experiment was based on*. Hence, the strength of the endogenous (omitted variable) relation we simulate can certainly occur in applications.

When the  $X_{43q}$  variable is not added to the estimation, the WTP for the  $X_{32q}$  variable increases with the strength of the relationship between  $Z_{1q}$  and  $X_{43q}$  (as given by  $R_1$ ), this is causes an positive omitted variable bias in the coefficient of  $X_{1iq}$ . However, the 90% of the estimations interval is rather large. Hence, we can not be certain that this WTP is larger than it's design value. This result is primarily caused by the large range in the estimates of the coefficient of  $X_{32q}$ .

Table 4: MNL estimation with no control  $X_{43q}$  and  $X_{53q}$

Design variables (strength relation between $Z_{iq}$ and $X_{jq}$ )										
Between $Z_1$ and $X_{43}$ ( $R_1$ )	0.00	0.01	0.02	0.03	0.04	0.06	0.07	0.08	0.09	0.10
Between $Z_2$ and $X_{43}$ ( $R_2$ )	0.00	0.08	0.17	0.25	0.33	0.42	0.50	0.58	0.66	0.75
Between $Z_2$ and $X_{53}$ ( $R_3$ )	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
WTP variable $X_{2iq}$ ( $\beta_2/\beta_1$ )										
Estimated	0.170	-0.033	-0.127	-0.214	-0.349	-0.250	-0.266	-0.229	-0.196	-0.143
Design value	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
Relative value	1.018	-0.199	-0.762	-1.281	-2.096	-1.499	-1.598	-1.377	-1.176	-0.856
WTP variable $X_{32q}$ ( $\beta_3/\beta_1$ )										
Estimated	0.132	0.202	0.253	0.210	0.379	0.278	0.292	0.356	0.276	0.249
Design value	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133
Relative value	0.992	1.518	1.898	1.578	2.846	2.086	2.190	2.673	2.069	1.867

Figure 3: Relative values estimated WTP's for  $X_{2i}$  in Table 4 to their design values

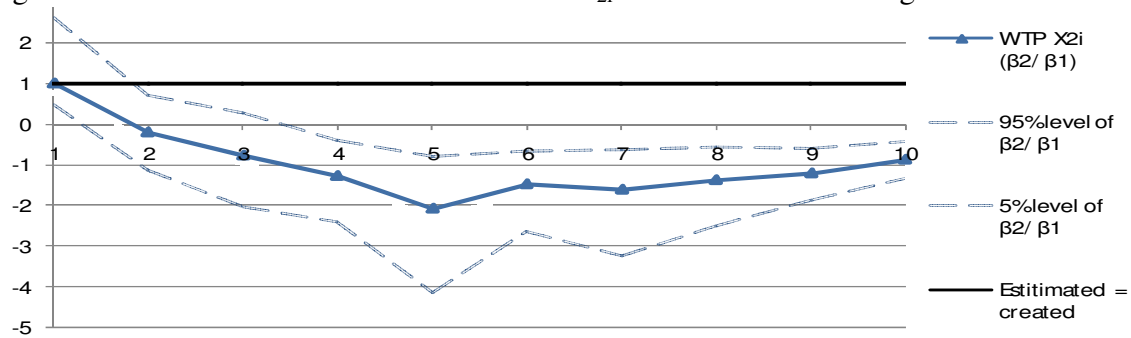
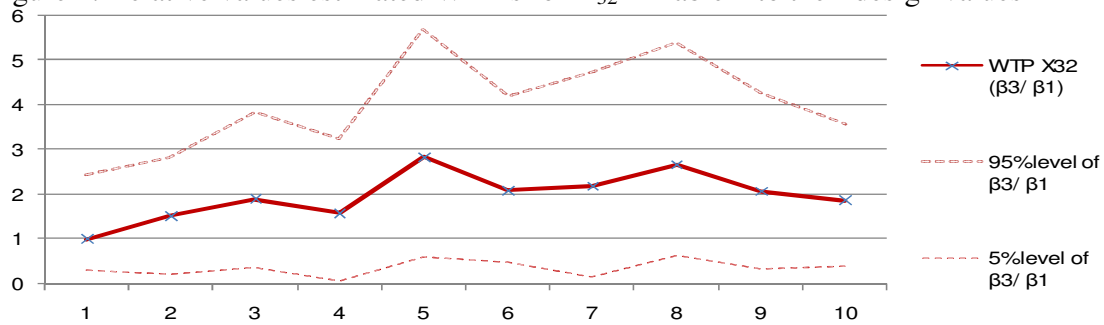


Figure 4: Relative values estimated WTP's for  $X_{32}$  in Table 4 to their design values



This section showed that omitted variables can cause very large biases in the WTP estimates from MNL. We argue that endogeneity problem from omitted variables can be quite common in discrete choice studies.

## 5. Correlation between two heterogeneous marginal utilities

The previous section showed the large biases that omitted variables can cause, and that the problem of omitted variables can be quite common. This section studies the effect of correlated heterogeneous marginal utilities, that the unobserved heterogeneity in the marginal utilities of attributes is correlated.

For instance, consider the value of statistical life (or for risk reduction). Suppose that for richer people the marginal utility of money is lower, because of a decreasing marginal utility of income. Furthermore, suppose that the marginal disutility of a risk of an accident is increasing with income, because richer people perceive that they have more to lose. Then, of course, these two marginal utilities are correlated.

Similarly, the marginal utility of travel time could be correlated with the marginal utility of price. In the study of purchase of appliances (see Revelt and Train (1998) for a study on refrigerator purchase) the marginal utility of the monetary saving due to the efficiency of the appliance might be correlated with the marginal utility of the efficiency level of the appliance, or with the marginal utility of the (possible) rebate given when purchasing a more efficient appliance. This rebate is presumably given to promote the purchase of high efficiency appliances.

Train (2007) studies the choice of electricity supplier and finds for example significant covariance between the marginal utility of whether the rates are differentiated over the day and the marginal utilities of price, whether the electricity supplier is a known company and if the electricity price differs over the year.

This section shows that correlated marginal utilities can lead to biased mean WTP's in MNL estimations. Train (1998) found that the WTP's from the mixed logit with correlated marginal utilities were lower than the WTP's from the regular mixed logit and MNL. Hence, not controlling for this type of correlation can have an effect. Random coefficient models, such as mixed logit and probit, can allow for full covariance between all coefficients (i.e. correlation between the random elements of the marginal utilities).

However, this full generality is rarely used, due to numerical difficulties in maximizing the simulated log-likelihood for such a general model, with the large amount of parameters that are needed. In most cases no covariance between the random marginal utilities is assumed, and in some studies "partly free covariance" is used (Train, 2007). A possible solution to this problem could be using different algorithms that do not directly maximize the log-likelihood to find the optimal values of the coefficients. Examples of these are the *hierarchical Bayes* method as supported by chapter 12 of Train (2003) or the *EM algorithm* of Train (2007; 2008).

Now we consider using a regular MNL when there is unobserved heterogeneity in the marginal utilities and there is correlation between the marginal utilities. For this we use the same basic set up as in Section 3 and the utility functions of (3a-c). The random elements of the parameters are again determined by Table 1. The means of the marginal utilities are also the same as in Section 3. However, now we only use symmetric distributions for the marginal utilities, with the design spreads equal to the mean. The main difference is that now the heterogeneous marginal utilities of  $X_{1iq}$  and  $X_{2iq}$  are related. In particular they are given by formulas (7) and (8).

$$\beta_{1q} = a_1 + Spread_1 * T_{1q} \tag{7}$$

$$\beta_{2q} = \left( a_2 * \frac{\rho}{(\beta_{1q} * \text{mean}(1/\beta_{1q}))} \right) + (1-\rho) * (a_2 + \text{Spread}_2 * T_{2q}) \quad (8)$$

Here  $a_2$  is the design mean of the marginal utility of  $X_{2iq}$  (i.e.  $\beta_{2q}$ ) and  $\text{spread}_{2q}$  gives the spread of the random element of this marginal utility. The strength of the relation between  $\beta_{2q}$  and  $\beta_{1q}$  is given by  $\rho$ , which is between zero and one. The  $T_{kq}$  is a random variable with a triangle distribution, with  $-1 \leq T_{kq} \leq 1$  and a zero mean. Note that the spread of the marginal utility of  $X_{2iq}$  is somewhat different for each level of  $\rho$ . Formula (8) shows that the larger  $\rho$ , the stronger the negative correlation between the marginal utility of  $X_{1iq}$  and  $X_{2iq}$ .

To facilitate interpretation, the  $X_{1iq}$  is again seen as the price of alternative  $i$ . For the analysing the effect of correlated heterogeneous marginal utilities on MNL estimates, we compare the estimated WTP's with the design WTP's. The design levels of the WTP's are found by dividing the dataset mean of  $\beta_{iq}$  by the mean  $\beta_{1q}$  Table 5 shows the results for the estimated mean relative values of coefficients (i.e. WTP's) of the 4000 different estimations and datasets per level of  $\rho$ . It has eight different values for  $\rho$ . Figure 5 shows the relative values from Table 5 for the estimated WTP's for  $X_{2iq}$  to the design values of the WTP for this attribute. The dashed lines in the figure give the area in which 90% of the estimations lie. Figure 6 does this for the WTP of  $X_{32q}$ .

The table also gives the average correlation between  $\beta_{1q}$  and  $\beta_{2q}$  per value of  $\rho$ . Note that even for  $\rho=1$  the linear correlation coefficient is not -1, as  $\beta_{2q}$  is related to the inverse of  $\beta_{1q}$ . The strongest correlation between the marginal utilities we simulate is -0.6. Revelt and Train (1998) find correlations between the random elements of the marginal utilities as large as +0.59 and -0.68. Though they also find some pairs of marginal utilities between which the correlation is basically zero. Train (2007) finds correlations between 0.1 and 0.94, with three correlation coefficients being larger than 0.9. Hence, these authors find some very strong relationships between (the random elements of) the marginal utilities. This shows that the strength of the correlation patterns we simulate can certainly occur in empirical applications.

Both Table 5 and Figure 5 clearly show that the stronger the relation between the two random marginal utilities the larger the difference between the design value and the estimated value of the WTP of  $X_{2iq}$ . Figure 5 shows that if the relation between the two heterogeneous marginal utilities is not that strong (say  $\rho < 0.4$ ) then the MNL estimates are not that affected, as then the estimated WTP's for  $X_{2iq}$  is still close to the design value. Conversely, for larger value of  $\rho$  there is a clear underestimation of the WTP for  $X_{2iq}$ .

However, the relation between the two marginal utilities has to be rather strong ( $\rho \geq 0.8$ ) for the entire 90% interval of relative values of estimated WTP's to be below the "estimated=design" line (i.e. where the relative value is 1). This indicates that correlated heterogeneous marginal utilities can indeed cause biased WTP estimated from MNL, where the problem is most likely to occur if the relation between the two marginal utilities is strong.

Note that this negative effect of relation between the heterogeneous marginal utilities on the accuracy of MNL estimates is not limited to the non-linear (inverse) relation we simulate in this section. When we reran the simulation using a linear decreasing relation between  $\beta_{1q}$  and  $\beta_{2q}$  we found that this type of relation can also cause the WTP's from MNL to be biased.

The estimated WTP's for the third attribute  $X_{32q}$  in Figure 6 and Table 5 seems to be somewhat above the design levels of this WTP. However, for this WTP there is no clear pattern of effect of the correlation between the two marginal utilities and the "estimated=design" line is

well inside the 90% interval. Hence, we cannot conclude that this WTP is affected. Note that in the simulation of the datasets the fixed marginal utility of  $X_{32q}$  is uncorrelated with the marginal utility of price ( $X_{1iq}$ ). Again the 90% interval for the WTP's of  $X_{32q}$  is much larger than the 90% interval for the WTP's of  $X_{2iq}$ . This seems to be caused by the larger variance in the estimates of the coefficient of  $X_{32q}$ .

Table 5: MNL estimations when the heterogeneous marginal utilities are negatively correlated

Design variables								
Strength relation between $\beta_{1q}$ and $\beta_{2q}$ ( $\rho$ )	0	0.2	0.333	0.4	0.5	0.666	0.8	1
Resulting correlation between $\beta_{1q}$ and $\beta_{2q}$	-0.00	-0.32	-0.46	-0.51	-0.55	-0.57	-0.57	-0.60
WTP variable ( $X_{2iq}$ ) ( $\beta_2/\beta_1$ )								
Estimated	0.168	0.168	0.164	0.165	0.160	0.155	0.151	0.143
Design value	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
Relative value	1.007	1.008	0.983	0.991	0.959	0.930	0.903	0.860
WTP variable $X_{32q}$ ( $\beta_3/\beta_1$ )								
Estimated	0.152	0.143	0.143	0.142	0.137	0.137	0.139	0.150
Design value	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133
Relative value	1.138	1.070	1.070	1.068	1.029	1.031	1.045	1.126

Figure 5: Relative sizes estimated WTP's for  $X_{2iq}$  to their design values for Table 5

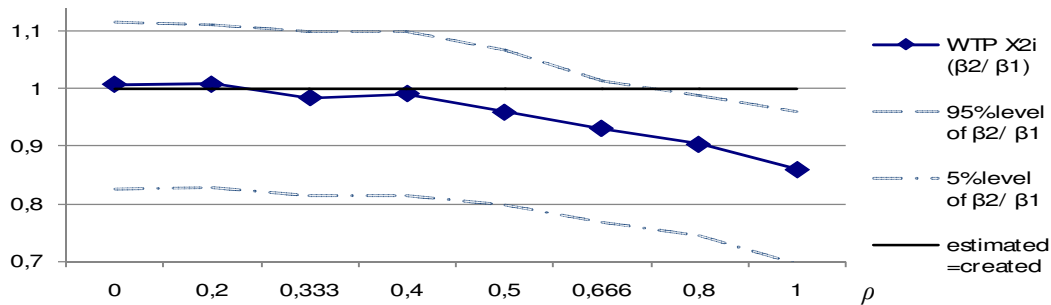


Figure 6: Relative sizes estimated WTP's for  $X_{32q}$  to their design values for Table 5

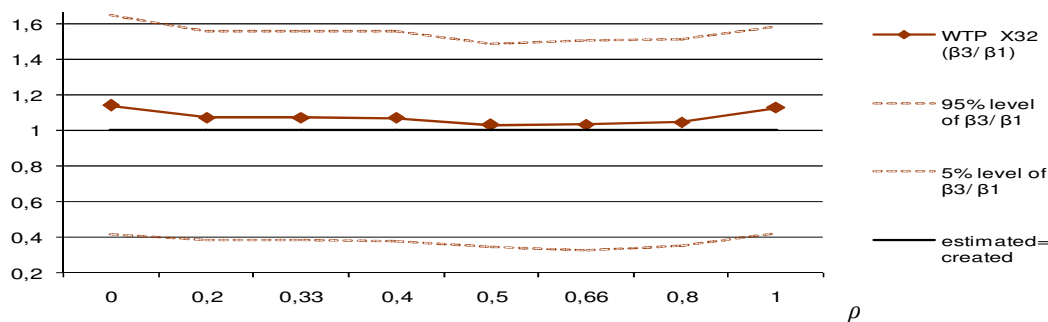




Table 5 shows that if the marginal utility of  $X_{2iq}$  is, following formula (8), a decreasing function of the marginal utility of  $X_{1iq}$ , a MNL on this choice situation can result in an under-estimation of the relative size of the (mean) marginal utility of  $X_{2iq}$  to  $X_{1iq}$  (i.e.  $\beta_2/\beta_1$ ). Conversely, if the  $\beta_{2q}$  is an increasing function of  $\beta_{1q}$ , following equation (9), this can result in an over-estimation by an MNL of the WTP  $X_{2iq}$  (i.e.  $\beta_2/\beta_1$ ).

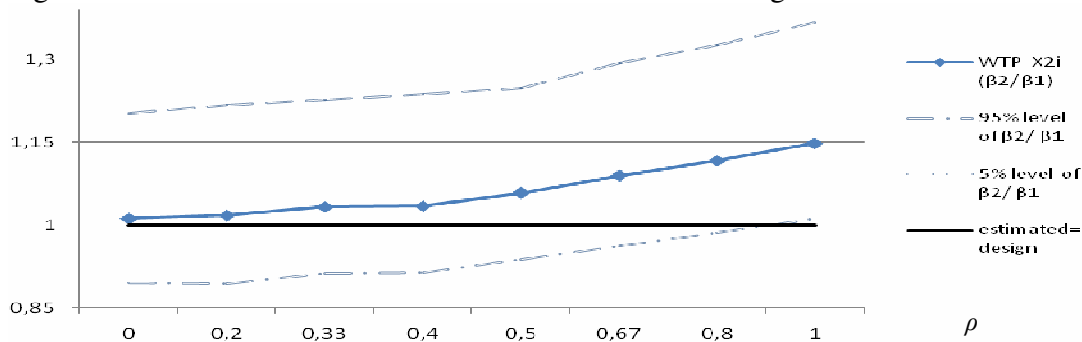
$$\beta_{2q} = \left( a_2 * \rho * \left( 2 - \frac{1}{(\beta_{1q} * \text{mean}(1/\beta_{1q}))} \right) \right) + (1 - \rho) * (a_2 + \text{Spread}_2 * T_{2q}) \quad (9)$$

The results of this simulation, with 4000 different datasets created by Monte Carlo simulation per level of  $\rho$ , are tabulated in Table 6 and shown in Figure 7. They show that a stronger (increasing) relationship of  $\beta_{1q}$  on  $\beta_{2q}$ , boosts the estimated WTP of  $X_{2iq}$ , even though the design value of the WTP is always the same. However, different from the earlier simulation with a negative relation between the marginal utilities, now the 90% interval is only above the estimated is the design value line (i.e. one) for very strong relations between the two marginal utilities ( $\rho=1$ ).

Table 6: MNL estimations when the heterogeneous marginal utilities are positively correlated

<b>Design variables</b>								
Strength relation between $\beta_{1q}$ and $\beta_{2q}$ ( $\rho$ )	0	0.2	0.333	0.4	0.5	0.666	0.8	1
Resulting correlation between $\beta_{1q}$ and $\beta_{2q}$	-0.00	0.32	0.47	0.51	0.54	0.59	0.59	0.60
<b>WTP variable <math>X_{2iq}</math> (<math>\beta_2/\beta_1</math>)</b>								
Estimated	0.169	0.170	0.172	0.172	0.176	0.181	0.186	0.191
Design value	0.167	0.167	0.167	0.167	0.167	0.167	0.166	0.167
Relative value	1.012	1.017	1.033	1.034	1.058	1.089	1.116	1.147
<b>WTP variable <math>X_{32q}</math> (<math>\beta_3/\beta_1</math>)</b>								
Estimated	0.152	0.157	0.158	0.153	0.161	0.156	0.161	0.158
Design value	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133
Relative value	1.139	1.178	1.184	1.148	1.211	1.167	1.208	1.186

Figure 7: Relative sizes estimated WTP's for  $X_{2i}$  to their design values for Table 6



A solution to the problem caused by correlated marginal utilities could be to estimate a mixed logit or probit model that allows the marginal utilities to be correlated (i.e. with free covariance). For future research it is interesting to study how mixed logit without free covariance performs when the heterogeneous marginal utilities are correlated.

This section found that if two heterogeneous marginal utilities are correlated with each other, this can cause the WTP's from MNL estimations to be biased, and that the absolute value of the bias increases with the strength of the correlation. If the marginal utility of the attribute to be valued (e.g. travel time) is an increasing function of the marginal utility of the monetary variable, this seems to cause a positive bias in the WTP estimated by MNL. Conversely, if there is a decreasing relation between the two marginal utilities, this can result in a negative bias in the WTP from MNL. Furthermore, this seems to be the case for both linear and non-linear relations between the two heterogeneous marginal utilities. The strengths of the correlations between the marginal utilities that we simulated, with maximum correlation coefficients of  $\pm 0.60$ , have also been observed in empirical studies using free-covariance mixed logits.

## 6. Conclusion

This paper studied three situations in which using Multinomial Logit (MNL) models might result in biased Willingness-To-Pay (WTP) estimates. A substantial number of empirical studies found that with mixed logit the estimated (mean) WTP's are different than with MNL. This is surprising as in theory the (random) unobserved heterogeneity, for which mixed logit controls, need not affect the relative values of the coefficients from MNL.

A well known cause of bias is endogeneity of the explanatory variables, in that some of the explanatory variables are correlated with the unobserved elements of the utility. We show that omitted variables (a form of endogeneity where an omitted variable influences the dependent variable directly and is correlated with an explanatory variable) can cause large biases. We discussed that the omitted variable problem can be quite common in discrete choice studies. For instance, in a study on the choice of transportation mode, the attributes price and travel time can be correlated with the unobserved elements, because these attributes and the relative preferences for the modes are partly determined by the *length in kilometres of the trip* about which the individual is interviewed. It could also be the case that the price attribute and unobserved elements are correlated because they are both influenced by the income of the decision maker or the *quality* of an alternative (e.g. *better* tomato juice is more expensive to make).

There are estimation methods that can control for endogeneity. However, these methods are more difficult and time consuming to use. In the case of omitted variable bias the preferred solution seems to be to simply include the omitted variable in the estimation. Hence, it might be advisable to put extra effort in acquiring information on variables that could cause an omitted variable bias.

To study the effect of heterogeneous marginal utilities on the WTP estimates from MNL, we compare the MNL estimates of the WTP's with the design WTP's. We find that if two heterogeneous marginal utilities are related, this can result in biased WTP estimates by MNL, which ignore the heterogeneity. In contrast, uncorrelated heterogeneity in the two marginal

utilities seems to have no detrimental effect on the MNL estimates. In our Monte Carlo simulations we find that the stronger the relation between two heterogeneous marginal utilities, the larger the bias in the WTP from MNL. This is the case for both linear and non-linear relations between the marginal utilities. If the relation between the two marginal utilities is increasing (i.e. positively correlated) this bias is positive, and if the relation is decreasing the bias is negative.

This could explain the different results in the empirical literature on whether MNL estimates give different WTP estimates than mixed logit, and if so in which direction. Our results suggest that if the MNL estimates are biased, the sign and size of this bias could be determined by the correlation pattern of the heterogeneous marginal utilities in the population. Thus if the correlation pattern is different one finds a different bias and if there is no correlation between the heterogeneous marginal utilities the MNL estimates of the WTP's should be unbiased.

### **Acknowledgements**

This paper is part of the project *Betrouwbaarheid van transportketens*, which is made possible with support of Transumo. Transumo (TRANSition SUSTainable MOBility) is a Dutch platform for companies, governments and knowledge institutes that cooperate in the development of knowledge with regard to sustainable mobility. We thank Paul Koster for his comments on an earlier version of this paper. The opinions and any remaining errors in this paper are purely our own responsibility.

### **Literature**

- Arellano, M. (2003) "Discrete Choices with Panel Data." *Investigaciones Económicas* 27, 423-458.
- Berg, van den, V. (2008) "Choice of Train Ticket: a Study on Dutch Travellers." ERSA conference 2008.
- Berry, S. (1994) "Estimating Discrete Choice Models with Product Differentiation." *RAND Journal of Economics* 25, 242-262.
- Berry, S., Levinsohn, J. and Pakes, A. (1995) "Automobile Prices in Market Equilibrium." *Econometrica* 63(4), 841-890.
- Bhat, C. R. (1998a) "Accommodating variations in responsiveness to level-of-service variables in travel mode choice models." *Transportation Research A* 32, 455-507.
- Bhat, C. R. (1998b) "Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice." *Transportation Research Part B* 32(7), 455-466.
- Bhat, C. R. (2000) "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling." *Transportation Science* 34(2), 228-239.
- Blundell, R. W. and Powell, J. L. (2004). "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies* 71(3): 655-679.
- Carro, J.M. (2007) "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects." *Journal of Econometrics* 140, 503-528.
- Goolsbee, A. and Petrin, A. (2004) "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica* 72(2): 351-381.
- Heckman, J. (1981), The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process, in C.Manski and D.McFadden, eds, 'Structural Analysis of Discrete Data with Econometric Applications', MIT Press, Cambridge, Massachusetts, pp. 179-85.

- Hensher, D. A., Rose, J. M., Greene, W. H. (2008) "Combining RP and SP data: Biases in Using the Nested Logit Trick - Contrasts with Flexible Mixed Logit Incorporating Panel and Scale Effects." *Journal of Transport Geography*, 16 (2), 126-133.
- Horowitz, J. (1980) "The Accuracy of the Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand." *Transportation Research B* 14:331-342.
- Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J.R., Fiebig, D., Greene, W., Hensher, D. and Waldman, D. (2005) "Recent Progress on Endogeneity in Choice Modeling." *Marketing Letters* 16(3-4), 255-265.
- Petrin, A. and Train K. (2004) "Omitted product Attributes in Differentiated product Models." University of California workingpaper. Retrieved at 15-4-2008, from <http://elsa.berkeley.edu/~train/pt42504.pdf>
- Rizzi, L. I. and Ortúzar, J de D. (2006) "Road safety valuation under a stated choice framework." *Journal of Transport Economics and Policy* 40(1), 69-94.
- Revelt, D. and Train, K.E. (1998) "Models with Taste Differences over People." *Land Economics* 74(2): 230-239.
- Train, K.E. (2001) "A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. Working Paper, Department of Economics, University of California, Berkeley. Retrieved at 7-4-2008, from <http://elsa.berkeley.edu/~train/compare.pdf>
- Train, K. E. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.
- Train, K.E. (2007) "A Recursive Estimator for Random Coefficient Models." University of California Working paper, Berkeley. Retrieved at 15 April 2008, from <http://elsa.berkeley.edu/~train/re.pdf>
- Train, K.E. (2008). "EM algorithms for nonparametric estimation of mixing distributions." *Journal of Choice Modelling* 1, 40-69.
- Weisstein, E. W. (2008) "Triangular Distribution." From *MathWorld*--A Wolfram Web Resource. Retrieved at 9-4-2008, from <http://mathworld.wolfram.com/TriangularDistribution.html>
- Villas-Boas, J. M. and Winer, R. S. (1999) "Endogeneity in Brand Choice Models." *Management Science* 45(10): 1324-1338.
- Wooldridge, J.M. (2003) *Introductory Econometrics: A Modern Approach. 2nd edition*. Mason: South-Western College Publishing.