Testing and Correcting for Sample Selection Bias in Discrete Choice Contingent Valuation Studies

Jan A. Eklöf[*] and Sune Karlsson[**]

Dept. of Economic Statistics
Stockholm School of Economics
P.O. Box 6501
SE-113 83 Stockholm

SSE/EFI Working Paper Series in Economics and Finance, No. 171
May 1997
Revised June 1999

Abstract: The discrete choice or "referendum" contingent valuation technique has become a popular tool for assessing the value of non-market goods. Surveys used in these studies frequently suffer from large non-response which can lead to significant bias in parameter estimates and in the estimate of mean Willingness to Pay. We investigate the properties of tests for sample selection bias and the losses made by applying estimators assuming no sample selection. The effects of sample selection bias can be sizable but bivariate probit estimation give unbiased estimates. A computationally straightforward test for sample selection bias is found to perform well.

Keywords: bivariate probit, discrete choice, contingent valuation, non-response, bias correction, willingness to pay, omitted variables test.

JEL Classification: C25, C34, C35, Q20.

---

[*] E-mail: jan.eklof@hhs.se

[**] E-mail: sune.karlsson@hhs.se

1. Introduction

The discrete choice or "referendum" technique has become increasingly popular in the contingent valuation literature following the pioneering work of Bishop and Heberlein [1979]. Typically, a random sample of the population is asked a (hypothetical) yes/no question if they are willing to contribute a specific amount towards the preservation of some environmental resource or the provision of a common good. These responses are then used to estimate the distribution of willingness to pay (*WTP*) conditional on various background variables. From the conditional distribution and knowledge about the population characteristics of the background variables it is then possible to obtain estimates of, say, the population mean or median *WTP*. See Hanemann [1984], Cameron [1988] and McConnell [1990] for a discussion of the utility-theoretic underpinnings of the discrete choice contingent valuation methodology (*DC-CVM*).

Non-response is common and often large[1] in contingent valuation studies and can bias the results from *DC-CVM* in two different ways. The first case may occur when the population characteristics of some or all of the background variables are unknown. The population characteristics (most commonly the mean) must then be estimated from the sample at hand. These estimates are biased if the non-response is systematically related to the background variables, we refer to this as **non-response bias**. Secondly, the non-response can be systematically related to an individuals actual *WTP*. This is commonly referred to as **sample selection bias** and leads to inconsistent parameter estimates. The issue of non-response bias and sample selection bias in contingent valuation has received considerable interest. Edwards and Anderson [1987], Loomis [1987], Whitehead et al. [1993] and Álvarez-Farizo, Hanley and Wright [1996] discuss non-response bias and sample selection bias, mainly in the context of open-ended[2] *WTP* questions where estimation is done by OLS. Dalecki et al. [1993] considers tests for non-response bias and weighting schemes which can be used to compensate for the bias. Whitehead et al. [1994] employ the censored probit

[1] "Response rates for the mail shot were good by CVM standards, being greater than 50% in almost all sub-samples." (Álvarez-Farizo, Hanley and Wright [1996]), "40-60% [response rate] seems average for general population CVM surveys" (Loomis [1987]), "…minimizing both sample non-response and item non-response are important. The former is unlikely to be below 20% even in very high quality surveys…" (National Oceanic and Atmospheric Administration [1993]).

model discussed below to test for sample selection bias in a *DC-CVM* setting. See Vella [1998] for a survey of the sample selection literature when the equation of interest is linear, i.e. the case of open-ended WTP questions.

This paper adds to the existing literature by giving a systematic account of the possibility to test for, and the effects of, sample selection bias in *DC-CVM*. To put the issue of sample selection bias in perspective we also provide evidence of the relative effects of non-response and sample selection bias.

The issue of sample selection bias is of course not only relevant in the *DC-CVM* setting. Indeed, much of the analysis in this paper applies to discrete choice modeling in general and sheds light on how sample selection bias can affect parameter estimates.

The remainder of the paper is organized as follows. Section 2 sets up the discrete choice model with sample selection and discusses how sample selection and non-response biases can affect parameter and mean population *WTP* estimates. In section 3 we introduce the appropriate bivariate probit estimator and several tests for sample selection bias. Section 4 is the heart of the paper and presents the results of our Monte Carlo study. In section 5 the conclusions are summarized.

2. Sample Selection and Non-response Bias

Given an initial random sample of N individuals, the data actually available to the investigator is generated by the bivariate model

$$y_{i1}^* = \mathbf{x}_{i1}\beta_1 + \varepsilon_{i1} \tag{1a}$$

$$y_{i2}^* = \mathbf{x}_{i2}\beta_2 + \varepsilon_{i2}. \tag{1b}$$

The unobservable variables y_{ij}^* are related to the binary (observed) outcomes y_{ij} by the rule

$$y_{ij} = \begin{cases} 1, & y_{ij}^* > 0 \\ 0 & y_{ij}^* \leq 0 \end{cases}$$

and $(\varepsilon_{i1}, \varepsilon_{i2})$ is iid standard bivariate normal with correlation ρ. In the context of *DC-CVM* y_{i2} is 1 if the individual is willing to contribute the prescribed amount towards the investigated cause and y_{i1} is the selection or response variable. That is, y_{i2} is only observed if y_{i1} = 1 or with (marginal) probability $\Phi(\mathbf{x}_{i1}\beta_1)$, leaving $N_2 < N$ observations on (1b). We will re-

[2] That is, the respondents are simply asked to state the maximum amount they are willing to pay rather than, as in the *DC-CVM* framework, asked if they are willing to pay a specific amount.

fer to (1a) as the selection equation and (1b) as the *WTP* equation. To be more specific we write (1b) as

$$y_{i2}^{*} = \alpha + \gamma bid_i + \mathbf{x}_{i2}^{*}\delta + \varepsilon_{i2}$$

where bid_i is the amount individual i is asked to contribute and $\mathbf{x}_{i2}^{*}$ are the remaining variables in $\mathbf{x}_{i2}$. Mean population *WTP* is then obtained as

$$E(WTP) = -\frac{\alpha + \mu\delta}{\gamma} \tag{2}$$

where μ is the population mean of $\mathbf{x}_{i2}^{*}$. If μ is known, the only issue is the quality of the estimate of $\beta_2 = (\alpha, \gamma, \delta)$. Sample selection bias occurs when $\rho \neq 0$ and bivariate estimation is required for consistency. When $\rho = 0$, the univariate estimates based on $N_2 < N$ observations are consistent but inefficient compared to the case when there is no non-response.

When μ is unknown it has to be estimated as well, typically using the N_2 available observations on (1b). If the non-response is systematically related to $\mathbf{x}_{i2}^{*}$, for example by $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}^{*}$ having variables in common – or more generally if they contain correlated variables, the estimate of μ will be inconsistent, leading to non-response bias.

It should be made clear that the randomness or representativity of the initial sample is not an issue when estimating β_2. The univariate probit estimates are consistent and asymptotically normal when $\varepsilon_{i2} \sim N(0,1)$ conditionally on the data being available for observation i, no matter how the sample is obtained. β_2 can, in other words, just as well be estimated using an appropriate convenience sample as a (potentially costlier) random sample (see Harrison and Lesley [1996] for a discussion of the use of convenience samples in *DC-CVM*). On the other hand, convenience samples are just as susceptible to sample selection as random sampling schemes. The importance of testing and correcting for sample selection does not depend on the nature of the sampling scheme.

The consistency of the estimate of μ does, of course, depend crucially on the randomness of the sample or the use of an appropriate weighting scheme. While sample selection and non-response biases may occur simultaneously, they affect the quality of the estimates through distinctly different mechanisms and can be addressed separately.

3. Testing for Sample Selection Bias

With the censored probit model (1) estimation is by Maximum Likelihood and the likelihood is given (van de Ven and van Praag [1981], Meng and Schmidt [1985]) by

$$\ln L = \sum_{i=1}^{N}\{y_{i1}y_{i2}\ln F(z_{i1}, z_{i2}, \rho) + y_{i1}(1-y_{i2})\ln[\Phi(z_{i1}) - F(z_{i1}, z_{i2}, \rho)] + (1-y_{i1})\ln[1-\Phi(z_{i1})]\} \quad (3)$$

where F is the standard bivariate normal distribution function, Φ the standard normal distribution function and $z_{ij} = x_i\beta_j$. There are several tests for sample selection bias available for the censored probit model. Dubin and Rivers [1989] discuss LR, Wald and LM-tests and Vella [1992] proposes a conditional moment test and an omitted variables test similar to Heckman's [1979] test for sample selection in a linear regression.

3.1 The LM-test

Under the null-hypothesis, $\rho = 0$, the gradient of the censored probit likelihood (3) simplifies to

$$\frac{\partial \ln L_i}{\partial \beta_1} = x_{i1}e_{i1} = g_{i1} \quad (4a)$$

$$\frac{\partial \ln L_i}{\partial \beta_2} = y_{i1}x_{i2}e_{i2} = g_{i2} \quad (4b)$$

$$\frac{\partial \ln L_i}{\partial \rho} = y_{i1}e_{i1}e_{i2} = \tau_i \quad (4c)$$

where e_{ij} are the generalized residuals (Gourieroux et al. [1987]) from the univariate models

$$e_{ij} = \frac{\phi(z_{ij})}{\Phi(z_{ij})[1-\Phi(z_{ij})]}[y_{ij}-\Phi(z_{ij})]. \quad (5)$$

The LM-statistic can thus be obtained as $\lambda_{\mathrm{LM}} = NR^2$ (uncentered R^2) from a regression of g_{i1}, g_{i2} and τ_i (evaluated at the univariate probit estimates) on a vector of ones.

3.2 The Conditional Moment Test

The conditional moment (CM) test of Vella [1992] is based on the condition $E(\varepsilon_{i1}\varepsilon_{i2}) = E[E(\varepsilon_{i1}|y_{i1})E(\varepsilon_{i2}|y_{i2})] = E(e_{i1}e_{i2}) = 0$. The sample equivalent is $\hat{\tau} = N^{-1}\sum_{i=1}^{N} e_{i1}(y_{i1}e_{i2})$, where the unconditional expectation, $E(\varepsilon_{i2}) = 0$, is used when $y_{i1} = 0$ and y_{i2} is unobserved. Let $g_i = (g_{i1}, g_{i2})$ and assume that the standard asymptotic result

$$N^{1/2}\begin{pmatrix} N^{-1}\sum_{i=1}^{N}\tau_i \\ N^{-1}\sum_{i=1}^{N}g_i \end{pmatrix} \xrightarrow{d} N(0, \mathbf{V})$$

holds. Taylor expanding $\hat{\tau}$ around $(\beta_1, \beta_2, 0)$ the asymptotic distribution is obtained as $N^{1/2}\hat{\tau} \xrightarrow{d} N(0, \mathbf{aVa}')$ where $\mathbf{a} = (1, -[plim\ N^{-1}\sum_{i=1}^{N}\partial\hat{\tau}/\partial(\beta_1, \beta_2)]I^{-1})$ and I is the information matrix for β_1 and β_2. Noting that all the quantities in $\mathbf{aVa}'$ are consistently estimated by forming the appropriate cross-products of τ_i and g_i. Pagan and Vella [1989] obtain the asymptotically $\chi^2(1)$ test statistic as

$$\lambda_{CM} = \frac{N^2\hat{\tau}^2}{\mathbf{t}'\mathbf{t} - \mathbf{t}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{t}}$$

where $\mathbf{t}$ is the $N\times1$ vector with elements τ_i and $\mathbf{G}$ is the matrix with rows g_i. A convenient way to calculate λ_{CM} is to run an auxiliary regression of τ_i on a constant term and g_i. λ_{CM} is then obtained as the square of the t-statistic for testing that the constant is zero.

3.3 The Omitted Variables Test

As an alternative to the conditional moment test, Vella [1992], suggested a test similar to Heckman's [1979] omitted variables (OV) test for sample selection in linear regression models. Using the bivariate normality of $(\varepsilon_1, \varepsilon_2)$ we can rewrite (1b) as

$$y_{i2}^{*} = \mathbf{x}_{i2}\beta_2 + \rho\varepsilon_{i1} + \eta_i, \qquad (1b')$$

conditioning on the observed y_{i1} we obtain

$$y_{i2}^{*} = \mathbf{x}_{i2}\beta_2 + \rho e_{i1} + \eta_i. \qquad (1b'')$$

In the case of censoring, $y_{i1} = 1$ when the *WTP*-equation is observed and the generalized residual, e_{i1}, simplifies to $\phi(z_{i1})/\Phi(z_{i1})$, the inverse of Mill's ratio, for these observations. The test for absence of sample selection is then implemented by first estimating (1a) by probit using the full sample, use these estimates to calculate e_{i1}, estimate (1b″) by probit using the censored sample and testing the hypothesis $\rho = 0$. That is, $\lambda_{OV} = \hat{\rho}^2/\hat{s}^2(\hat{\rho})$.

This also suggests an alternative estimation strategy – estimate (1b″) using univariate probit – when a censored probit routine is unavailable. The probit estimator of (1b″) is, however, inconsistent under the alternative. The inconsistency arises because η_i is heteroskedastic, $Var(\eta_i) = 1 - \rho^2 e_{i1}(z_{i1} + e_{i1})$. Van de Ven and van Praag [1981] suggest a three-step estimation procedure that is consistent under the alternative.

4. The Monte Carlo Study

We use a Monte Carlo study to evaluate the properties of available tests for sample selection bias. In addition we study how sample selection affects parameter estimates and – of particular interest for *DC-CVM* studies – estimates of mean population *WTP*. For the latter we behave as if truncation ($\mathbf{x}^*_{i2}$ is unobservable) of the *WTP* equation occurs simultaneously with censoring in order to study the effect of non-response bias on *WTP* estimates.

4.1 Data Generating Process

The data generating process used in the Monte Carlo experiments is a simple extension of the univariate equation used by Cooper [1994] to study the properties of different methods for constructing confidence intervals for population *WTP* measures. We add the selection equation (6a) to Cooper's *WTP*-equation (6b),

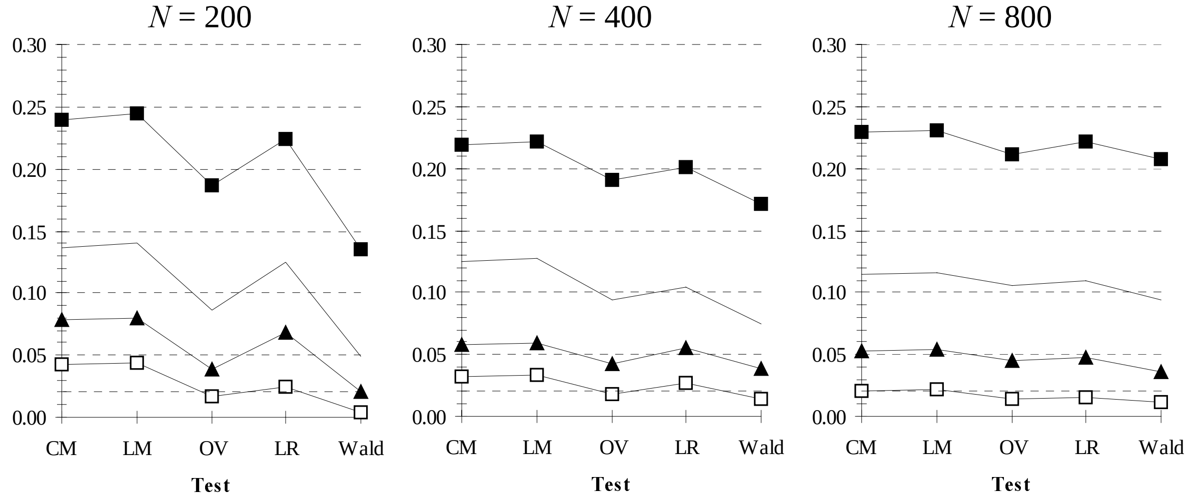
$$y^*_{i1} = a + 0.5x_{i1} + 0.5x_{i2} + \varepsilon_{i1} \tag{6a}$$

$$y^*_{i2} = 1.4136 - 0.008561bid_i + 0.00372x_{i3} + \varepsilon_{i2}, \tag{6b}$$

where x_{i1}, x_{i2} and x_{i3} are normally distributed. x_{i1} and x_{i2} with mean zero and standard deviation 1 and x_{i3} with mean 95.78 and standard deviation 119.226. The variable *bid* consists of repetitions of the sequence {2.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 250, 300, 400, 500, 700}, the so called bid-vector here consisting of 20 alternatives. With these parameter values true mean *WTP* is 206.74. The parameter *a* take the values {–0.31, 0.31, 1.03, 1.57} corresponding to average selection or response probability of 0.4, 0.6, 0.8 and 0.9. We let ρ, the correlation between ε_{i1} and ε_{i2}, take the values {–0.8, –0.4, 0, 0.4, 0.8}. To study the effect of non-response bias we let x_{i2} and x_{i3} be correlated with correlations δ = {–0.5, 0, 0.5}. x_{i1} is independent of x_{i2} and x_{i3} in all cases.

The sample sizes used are N = {200, 400, 800}, corresponding to expected actual sample sizes of N_2 = {120, 240, 480} for (6b). For each case and value of ρ, we generate 2,500 data sets with 800 observations, test the null hypothesis $\rho = 0$ and estimate mean *WTP* using the first N observations. That is, the test statistics are calculated using N observations on the selection equation and N_2 observations on the *WTP* equation. Estimates of mean population *WTP* are calculated using the probit estimates and the mean of the explanatory variables for the censored sample of N_2 observations. As a basis of comparison we also estimate mean *WTP* using the full sample and a univariate probit estimator.

Figure 1. Empirical Size of Tests for Sample Selection Bias, δ = 0



Lines correspond to the nominal size of the tests 0.02 0.05 0.1 0.2

Due to convergence problems[3] some of the data sets were discarded and the actual Monte Carlo sample size varies between 2,126 and 2,495. For the sake of brevity we only present a subset of the results. Unless explicitly noted, the results are for the simulations with a 60% response rate (a = 0.31). A full set of results is available on request from the authors.

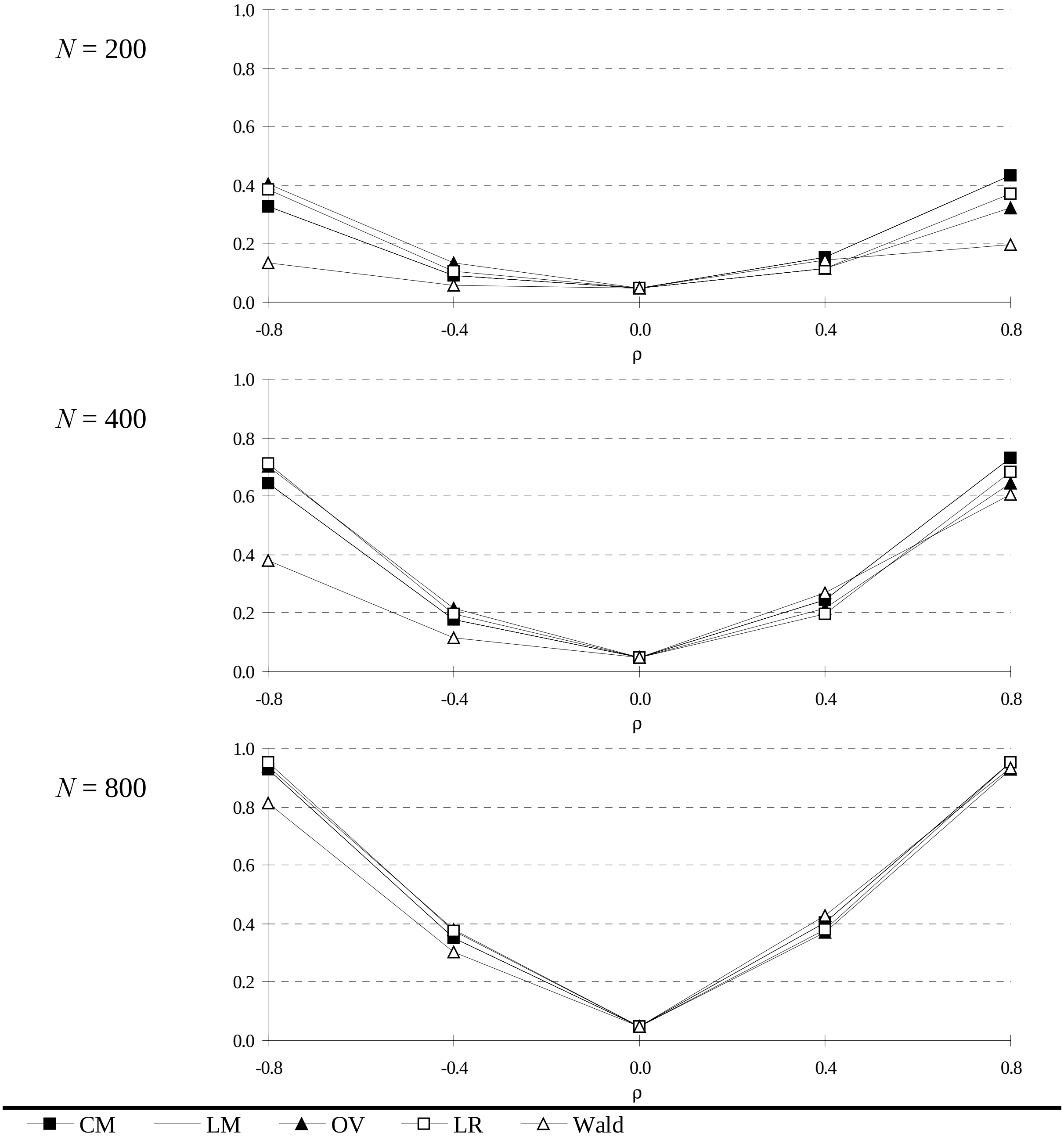
4.2 Tests for Sample Selection Bias

The empirical size of the tests is reported in Figure 1 for case where $\delta = 0$. It is observed that the Wald test is severely undersized in case 1 for $N = 200$ but has reasonable size properties for $N = 800$. The CM and LM tests are very similar and oversized for all sample sizes except for $N = 800$ and nominal significance levels of 0.05 and less. The LR test is also oversized for $N = 200$ but is very close to the nominal size for $N = 400$ and 800. Overall, the OV test has the best size properties although it is slightly undersized for $N = 200$.

While the size properties of the tests depend on the actual data generating process, the empirical size of the tests does not vary much (less than two percentage points for tests with 5% nominal significance level) over the cases considered here. The empirical size tends to be slightly higher for $\delta \neq 0$ than for $\delta = 0$.

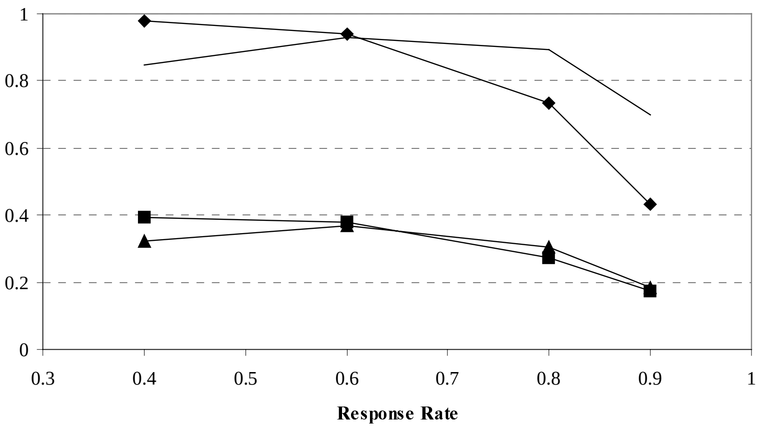
[3] A convergence failure was declared if the number of iterations exceeded 100. In addition there were a few cases where the estimated variance-covariance matrix (based on the Hessian matrix) for the censored probit failed to be positive definite or the censored probit converged to a local maximum with a smaller log-likelihood value than under the null hypothesis. All calculations were performed using Gauss version 3.2.15 and version 4.0.7 of the Gauss application "Maximum Likelihood". The Gauss code is available on request from the authors.

Figure 2. Size Corrected Power of Tests for Sample Selection Bias, δ = 0, 5% Size



When discussing the power properties of the tests we concentrate on tests at the 5% level. The size corrected power of the tests for the 5% significance level and $\delta = 0$ is plotted in Figure 2. The power of the Wald test is very poor for $N = 200$ and still lower than the other tests for $N = 800$ and $\rho < 0$. The power properties of the CM and LM tests are identical. For $N = 200$ and 400 they have the highest power of the tests when $\rho > 0$, but lower power than the OV and LR tests for $\rho < 0$. For $N = 800$ the differences in power between the CM, LM, OV and LR tests are very small.

Figure 3. Power of the OV Test, $N = 800$, $\delta = 0$



True value of ρ -0.8 -0.4 0.4 0.8

The power of the tests is quite good for $N = 800$. With smaller sample sizes the power is low for small correlations and we can only expect to detect large values of ρ.

One striking feature of Figure 2 is that the tests – with exception for the LR and OV tests – have better power against positive ρ than negative ρ. One reason for this is that the estimates of ρ in the censored probit model are biased towards zero, but more so for negative ρ than positive ρ. The Wald test is based on the estimates under the alternative and it is quite clear how the differences in bias could affect this test. It is less clear why the CM and LM tests have higher power for positive ρ.

The tests have lower size corrected power when $\delta \neq 0$ and there are noticeable differences in power for all tests even when $N = 800$. The loss of power is most pronounced when $\delta = –0.5$. Testing at the 5% level and for $N = 800$, the power decreases by as much as 8 percentage points for the Wald test and 5 percentage points for the other tests when $\delta = –0.5$. With $\delta = 0.5$ the power losses are less than 4 percentage points for the Wald test and less than 6 percentage points for the other tests.

The power of the tests decline with increasing response rate and more so for extreme values of ρ. The size corrected power of the OV test as a function of the response rate is displayed in Figure 3; the results for the other tests are similar. It is interesting to note that the pattern is different for negative and positive ρ. It thus appears that the symmetry of the power function for the OV and LR tests noted above is an artifact of the 60% response rate in Figure 2.

Summing up the results on testing for sample selection bias it is clear that the tests, with the exception of the Wald test, perform well for larger ($N = 800$) sample sizes and large fractions of non-response. It is, however, quite disturbing that the tests lose power when the explanatory variables in the *WTP* and selection equations are correlated. The common, and from a specification point reasonable, practice of having explanatory variables common to the two equations may thus lead to a substantial reduction in the ability to detect sample selection bias.

4.3 The Effects of Non-response and Sample Selection Bias on Estimates

4.3.1 Parameter Estimates

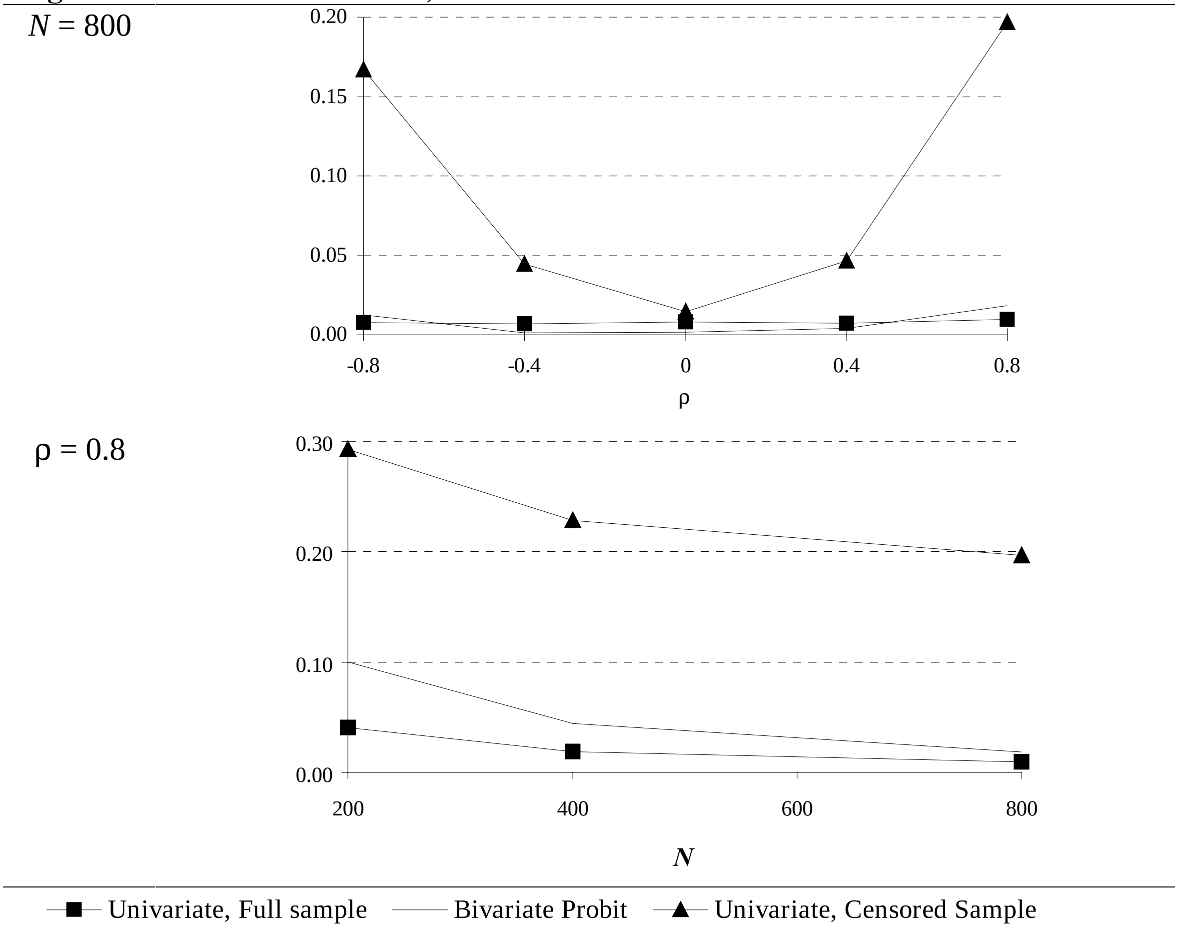
We estimate the parameters in the *WTP*-equation (6b) using three different estimators. Univariate probit on the censored sample, the censored bivariate probit and – as a benchmark – univariate probit on the full sample. The censored sample univariate probit is biased and inconsistent unless $\rho = 0$ and the results from this estimator allow us to assess the costs of falsely ignoring the sample selection issue.

The relative bias and relative Root Mean Square Error (RMSE) for the variable *bid* is displayed in Figures 4 and 5 for a subset of the simulations. A full set of results for $\delta = 0$ are given in Tables 1 to 4 of the appendix.

For $\rho = 0$ all estimators are consistent and the bias and RMSE are decreasing with the sample size. The RMSE of the censored sample univariate probit is about 1.4 times as large as the RMSE of the full sample estimator, corresponding to the 40% reduction in sample size. It is interesting to note that the bias of the censored bivariate probit is smaller than the bias of both the univariate probit estimators and the RMSE is – with the exception of the constant term – smaller than the RMSE of the censored sample univariate probit. There is, provided that the selection equation is correctly specified, little to lose by using the censored bivariate probit estimator even when sample selection bias is not a problem.

With $\rho \neq 0$ the bias and RMSE of the censored bivariate probit are increasing with the magnitude of ρ but shows clear sign of going to zero as the sample size increases. The bias and RMSE of the censored sample univariate probit are also decreasing but are considerably larger, and decrease at a slower rate. For $N = 800$ the RMSE of the censored bivariate probit estimates of *bid* and x_3 is between 1.1 and 1.7 times the RMSE of the full sample univariate probit. For the censored sample univariate probit the RMSE is between 1.3 and 3.1 times the RMSE of the full sample estimator. The loss in precision for the

Figure 4. Relative Bias for *bid*, δ = 0

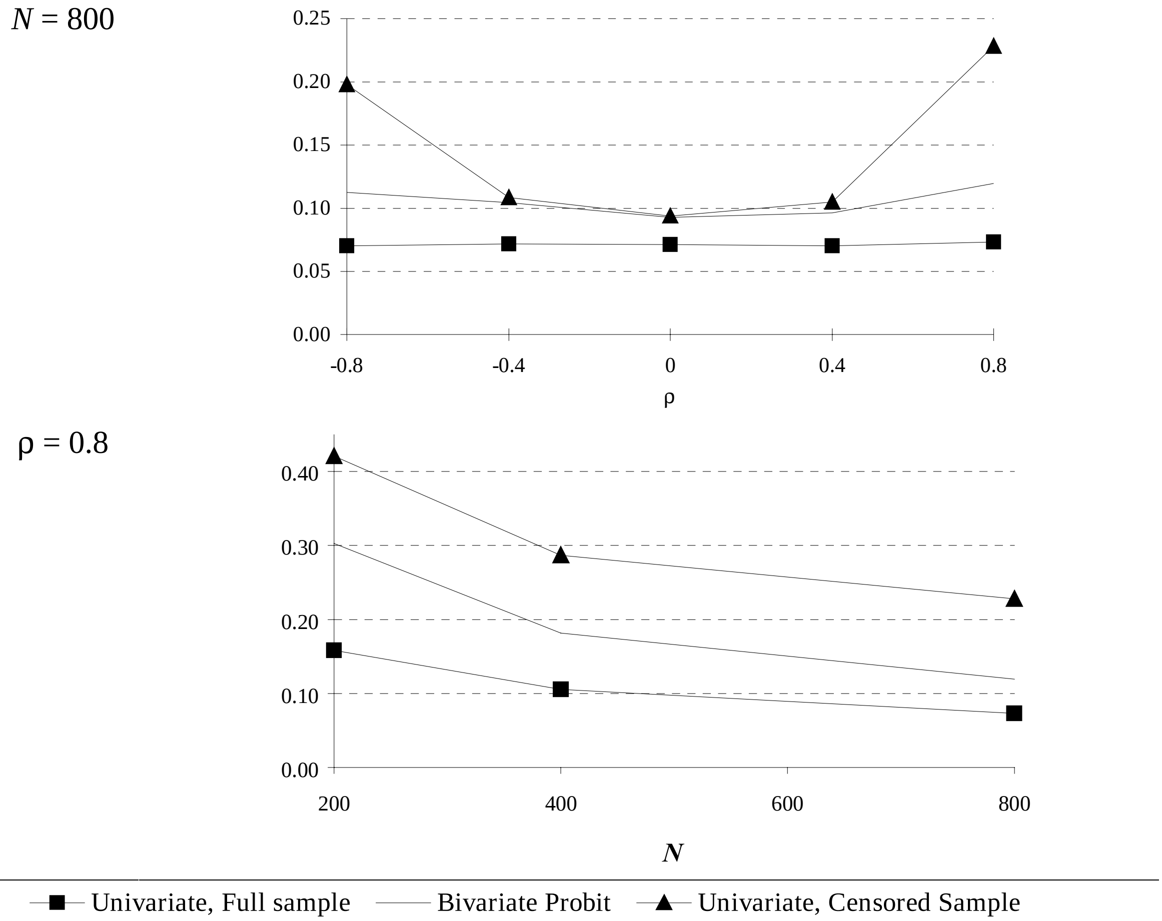


censored bivariate probit is thus roughly in agreement with the reduction in sample size for the *WTP*-equation. The bias of the bivariate censored probit is comparable to that of the full sample univariate probit whereas the bias of the censored sample univariate probit is more then ten times larger than the full sample univariate probit.

Varying the response rate has virtually no effect on the bias of the bivariate probit estimates. The RMSE on the other hand changes with the response rate. For the smaller values of ρ and the larger sample sizes the change in RMSE is roughly in agreement with the change in effective sample size for the *WTP* equation. For ρ = –0.8, 0.8 or N = 200 the RMSE increase faster with decreasing response rate. Changing the response rate from 90% to 40% increases the RMSE for *bid* by as much as a factor of 2.5 rather than 1.5 as can be expected from the change in effective sample size.

The censored sample univariate probit is severely affected by decreasing response rates. For N = 800 and ρ = 0.8 the relative bias of *bid* increase from 0.09 to 0.25 and the relative RMSE increase from 0.12 to 0.30 when the response rate is changed from 90% to 40%. For

Figure 5. Relative RMSE for *bid*, δ = 0



high response rates the sample selection bias may be relatively benign but with response rates common in *DC-CVM* studies the effect of sample selection bias could be disastrous.

Both the bivariate probit and the censored sample univariate probit are affected by the correlation, δ, between x_2 and x_3. With δ = –0.5 and *N* = 800 the RMSE of *bid* increase by as much as 8% for the bivariate probit with negative ρ and changes very little for ρ ≥ 0. The RMSE for x_3 decreases with 8% for ρ = –0.8 and increases by as much as 10% for larger values of ρ. With δ = 0.5, the RMSE for *bid* increases by less than 4% and decease slightly for ρ = –0.4 and 0. For x_3 the RMSE increases by more than 10% for ρ ≤ 0 and decreases by 4% for ρ = 0.8. The effect of the correlation is even stronger for the censored sample univariate probit but the pattern is similar to the bivariate probit.

To summarize: when δ = –0.5 (0.5), the estimates of *bid* deteriorates for negative (positive) ρ and the estimates of x_3 improves for negative (positive) ρ and deteriorates for positive (negative) values of ρ. This particular pattern of the effect of the signs of ρ and δ appear to be due to the design of the Monte Carlo study. Both x_2 and x_3 have positive coefficients. A positive correlation between them will thus lead to an increase (decrease) in the response probability when the probability of a yes answer to the *WTP* question is high (low) given the other variables in the *WTP*-equation. The converse obtains for a negative correlation between x_2 and x_3. A positive δ will thus enhance the sample selection effect of a positive ρ and partially offset the effect of a negative ρ. This explains the effect on the estimate for the uncorrelated variable *bid*. For the correlated variable x_3 it is not clear why the estimates improve when δ and ρ have the same sign and deteriorate with opposite signs.

4.3.2 Estimating Population Willingness to Pay

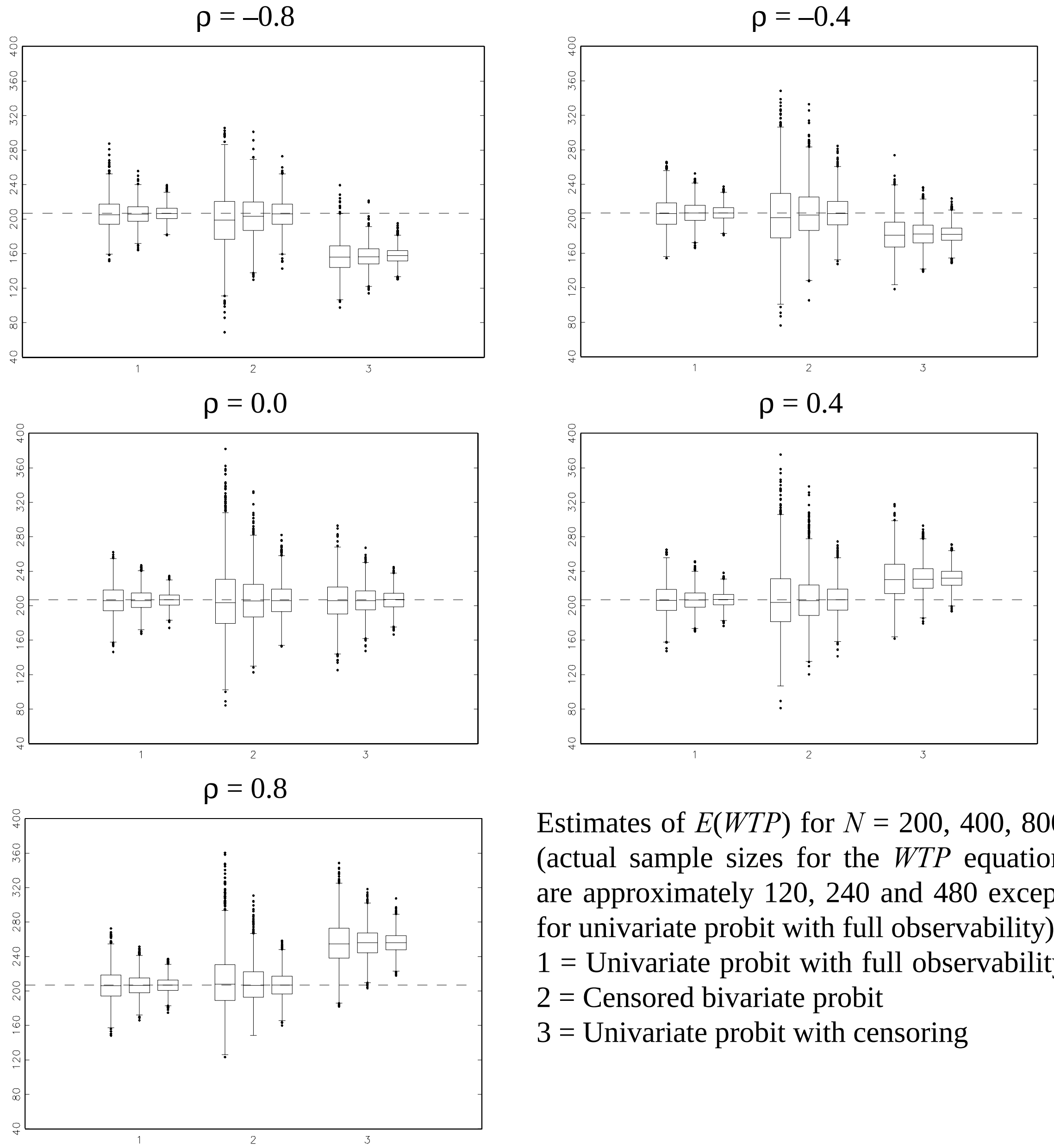
The estimate of mean population *WTP*, $E(WTP)$, obtained from (2) depends on the quality of the parameter estimates in the *WTP* equation as well as the quality of the estimates of the population mean of the explanatory variables. That is, the estimate of $E(WTP)$ is susceptible to both sample selection bias and non-response bias, the latter occurring when the response rate is related to one or more of the explanatory variables in the *WTP* equation with biased estimates of the variable means as a result.

The results for $\delta = 0$, where the only issue is sample selection, are displayed in Figure 6. It is clear from Figure 6 that the univariate censored sample probit performs poorly with a bias of 25% of true $E(WTP)$ for $\rho = \pm 0.8$ and about 12% of true $E(WTP)$ for $\rho = \pm 0.4$. As can be expected, true $E(WTP)$ is overestimated with positive correlations and underestimated with negative correlations. For $\rho = 0$ the censored sample univariate probit is consistent and there is no bias in the estimates of $E(WTP)$. The censored bivariate probit, on the other hand, gives estimates of $E(WTP)$ which are essentially unbiased for the larger sample sizes. The price paid for this unbiasedness is a higher variability in the estimates. This is, however, more than offset by the smaller bias and the censored bivariate probit has a considerably smaller RMSE than the censored sample univariate probit when $\rho \neq 0$.

As a comparison, the results for the (infeasible) full sample univariate probit estimator are also given. We note that this estimator gives unbiased estimates of $E(WTP)$ with considerably smaller variance than the bivariate probit. For $\rho = 0$ the censored sample univariate also gives unbiased estimates of $E(WTP)$, but the variance is almost twice as high due to the 40% loss in effective sample size.

It is clear that the losses from using the wrong estimator are substantial. Using the bivariate probit when $\rho = 0$ doubles the variance compared to the censored sample univariate probit and using the censored sample univariate probit when $\rho \neq 0$ leads to a substantial bias and a tenfold increase in MSE compared to the bivariate probit estimator. Substantial

Figure 6. *E*(*WTP*) Estimates, Case 1



Estimates of $E(WTP)$ for N = 200, 400, 800 (actual sample sizes for the WTP equation are approximately 120, 240 and 480 except for univariate probit with full observability)
1 = Univariate probit with full observability
2 = Censored bivariate probit
3 = Univariate probit with censoring

Boxes show the first quartile, median and third quartile, the length of the whiskers is 1.5 times the interquartile range and each dot represent one outlier. The dashed line represents true $E(WTP)$.

gains can, of course, also be made by reducing the non-response by follow-up or other measures.

With correlation between the explanatory variables in the WTP and selection equations, the estimate of the population mean of the explanatory variables is inconsistent and we have non-response bias in the estimates of $E(WTP)$. The non-response bias is present both in the censored sample univariate probit and the bivariate censored probit estimates of $E(WTP)$. The bias is, however, relatively small at about 3% of true $E(WTP)$. The size of the bias is, of course, sensitive to the strength of the correlation between the explanatory variables and their importance for determining *WTP* and the selection probability. While the non-response bias is relatively small compared to the sample selection bias for the censored sample univariate probit in this case it is quite conceivable that the non-response bias can be the dominating factor in some situations.

The sign of the non-response bias depends on the sign of the correlation between the variables and the sign of their coefficients in the selection and *WTP* equations. If the product of the coefficients and the correlation is negative, the bias is negative ($E(WTP)$ is underestimated) and the bias is positive when the sign of the product is positive. A particular case of interest is when the are variables common to the selection and *WTP* equations. Frequently they will enter with the same sign in the two equations, leading to a positive non-response bias. It is also reasonable to expect the correlation, ρ, between the disturbances in the two equations to be positive, that is that the sample selection bias of the censored sample univariate probit is positive as well. Non-response bias will thus add to the sample selection bias rather than offset it in this common case.

5. Summary

Sample selection bias can have a large effect on both parameter estimates and estimates of population willingness to pay based on univariate probit estimation. It is thus important to test for sample selection as an integral part of the analysis. The Monte Carlo study shows that the computationally straightforward OV test, which does not require maximization of the full bivariate probit likelihood performs well. It has good size properties and the size-adjusted power is comparable to the power of the other tests.

The censored bivariate probit estimator performs well when sample selection is present, giving essentially unbiased estimates of the parameters in the model and of population willingness to pay. The increased variance compared to univariate probit estimation is more than offset by the absence of bias and the MSE is several orders of magnitude smaller. The reduction in MSE is especially noticeable for the estimates of willingness to pay where the MSE of the censored bivariate probit can be as little as 10% of the MSE for the univariate probit.

Non-response bias – that is, inconsistent estimates of the population mean of the explanatory variables due to a systematic relation between the non-response and the explanatory variables – is also an issue of concern when estimating population willingness to pay. For the data generating process considered in the Monte Carlo study the non-response bias is found to be considerably smaller than the sample selection bias and thus to be of secondary importance.

References

Álvarez-Farizo, B., N. Hanley and R. Wright, (1996), 'Sample Selection Bias and the Estimation of Bid Curves in Open-Ended Contingent Valuation Studies,' Discussion Paper in Ecological Economics, No 96/5, Department of Economics, University of Stirling.

Bishop, R.C., and T.A. Heberlein, (1979), 'Measuring Values of Extra-Market Goods: Are Indirect Methods Biased?' *American Journal of Agricultural Economics,* **61,** 926-930.

Cameron, T.A., (1988), 'A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression', *Journal of Environmental Economics and Management,* **15,** 355-379.

Cooper, J.C., (1994), 'A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys,' *Land Economics*, **70**, 111-122.

Dalecki, M.G., J.C. Whitehead and G.C. Blomquist, (1993), 'Sample Non-response Bias and Aggregate Benefits in Contingent Valuation: An Examination of Early, Late and Non-respondents,' *Journal of Environmental Management*, **38**, 133-143.

Dubin, J.A., and D. Rivers, (1989), 'Selection Bias in Linear Regression, Logit and Probit Models,' *Sociological Methods and Research*, **18**, 360-390.

Edwards, F.E., and G.D. Anderson, (1987), 'Overlooked Biases in Contingent Valuation Surveys: Some Considerations,' *Land Economics*, **63**,168-178.

Gourieroux, C., A. Monfort, E. Renault and A Trognon, (1987), 'Generalized Residuals,' *Journal of Econometrics*, **34**, 5-32.

Hanemann, W.M., (1984), 'Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses,' *American Journal of Agricultural Economics,* **66,** 332-341.

Harrison, G.W., and J.C. Lesley, (1996), 'Must Contingent Valuation Surveys Cost So Much?', *Journal of Environmental Economics and Management* **31**, 79-95.

Heckman, J., (1979), 'Sample Selection Bias as a Specification Error,' *Econometrica*, **47**, 153-161.

Loomis, J.B., (1987), 'Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practice and Proposed Solutions,' *Land Economics*, **63**, 396-402.

McConnell, K.E., (1990), 'Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation,' *Journal of Environmental Economics and Management*, **18**, 19-34.

Meng, C.-L., and P. Schmidt, (1985), 'On the Cost of Partial Observability in the Bivariate Probit Model,' *International Economic Review*, **26**, 71-85.

National Oceanic and Atmospheric Administration, (1993), 'Report of the NOAA Panel on Contingent Valuation,' *Federal Register*, **58** No 10, 4601-4614.

Pagan, A., and F. Vella, (1989), 'Diagnostic Tests for Models Based on Individual Data: A Survey,' *Journal of Applied Econometrics*, **4**, S29-S59.

van de Ven, W.P.M.M., and B.M.S. van Praag, (1981), 'The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection,' *Journal of Econometrics*, **17**, 229-252.

Vella, F., (1992), 'Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models,' *Journal of Applied Econometrics*, **7**, 413-421.

Vella, F., (1998), 'Estimating Models with Sample Selection Bias: A Survey,' *The Journal of Human Resources*, **33**, 127-168.

Whitehead, J.C., P.A. Groothuis and G.C. Blomquist, (1993), 'Testing for Non-response and Sample Selection Bias in Contingent Valuation: Analysis of a Combination Phone/Mail Survey,' *Economics Letters*, **41**, 215-220.

Whitehead, J.C., P.A. Groothuis, T.J. Hoban and W.B. Clifford, (1994), 'Sample Bias in Contingent Valuation: A Comparison of the Correction Methods,' *Leisure Science*, **16**, 249-258.

Appendix 1

Table A1. Parameter Estimates in the *WTP* Equation, *constant*, δ = 0, 60% Response Rate

		ρ									
N	Estimator	–0.8		–0.4		0.0		0.4		0.8	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
200	Univ. probit, full sample	0.033	0.042	0.041	0.045	0.042	0.045	0.035	0.044	0.041	0.044
	Censored biv. probit	–0.007	0.064	–0.024	0.100	–0.004	0.139	0.043	0.216	0.200	0.375
	Univ. probit, censored	–0.216	0.115	–0.110	0.086	0.068	0.089	0.354	0.239	0.912	1.058
400	Univ. probit, full sample	0.014	0.020	0.024	0.021	0.022	0.019	0.015	0.020	0.020	0.020
	Censored biv. probit	–0.004	0.027	–0.010	0.044	0.005	0.066	0.015	0.102	0.089	0.149
	Univ. probit, censored	–0.244	0.091	–0.138	0.052	0.036	0.036	0.303	0.137	0.823	0.750
800	Univ. probit, full sample	0.007	0.009	0.009	0.010	0.010	0.010	0.010	0.010	0.012	0.010
	Censored biv. probit	–0.001	0.012	–0.005	0.020	–0.001	0.033	0.012	0.048	0.035	0.063
	Univ. probit, censored	–0.254	0.079	–0.157	0.040	0.017	0.016	0.284	0.101	0.776	0.636

True parameter value is 1.4136.

Table A2. Parameter Estimates in the *WTP* Equation, *bid*, δ = 0, 60% Response Rate

		ρ									
N	Estimator	–0.8		–0.4		0.0		0.4		0.8	
		Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6
200	Univ. probit, full sample	–0.31	1.72	–0.34	1.91	–0.33	1.81	–0.30	1.74	–0.35	1.84
	Censored biv. probit	–0.67	4.50	–0.25	3.84	–0.11	3.34	–0.23	3.58	–0.86	6.71
	Univ. probit, censored	–1.93	8.00	–0.86	4.54	–0.55	3.83	–0.91	4.48	–2.51	13.00
400	Univ. probit, full sample	–0.15	0.76	–0.17	0.83	–0.17	0.79	–0.13	0.79	–0.16	0.82
	Censored biv. probit	–0.30	1.97	–0.08	1.69	–0.07	1.39	–0.06	1.48	–0.38	2.43
	Univ. probit, censored	–1.60	4.34	–0.54	1.84	–0.29	1.48	–0.53	1.73	–1.96	6.05
800	Univ. probit, full sample	–0.07	0.36	–0.06	0.38	–0.07	0.37	–0.06	0.36	–0.08	0.39
	Censored biv. probit	–0.11	0.93	–0.01	0.80	–0.02	0.64	–0.04	0.68	–0.16	1.05
	Univ. probit, censored	–1.43	2.88	–0.38	0.87	–0.12	0.65	–0.40	0.81	–1.68	3.82

True parameter value is –0.008561.

Table A3. Parameter Estimates in the *WTP* Equation, x_3, δ = 0, 60% Response Rate

		ρ									
N	Estimator	–0.8		–0.4		0.0		0.4		0.8	
		Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6	Bias×10^3	MSE×10^6
200	Univ. probit, full sample	0.15	1.39	0.15	1.38	0.11	1.39	0.14	1.36	0.18	1.43
	Censored biv. probit	0.27	2.03	0.09	2.19	0.03	2.44	0.13	2.79	0.44	3.99
	Univ. probit, censored	0.77	3.02	0.33	2.49	0.21	2.65	0.44	3.31	1.21	6.30
400	Univ. probit, full sample	0.06	0.66	0.06	0.62	0.05	0.63	0.06	0.65	0.07	0.65
	Censored biv. probit	0.13	0.94	0.00	0.97	–0.01	1.01	0.03	1.22	0.17	1.56
	Univ. probit, censored	0.63	1.51	0.18	1.04	0.09	1.05	0.24	1.39	0.90	2.78
800	Univ. probit, full sample	0.04	0.32	0.02	0.30	0.02	0.31	0.02	0.30	0.02	0.31
	Censored biv. probit	0.06	0.45	–0.02	0.47	–0.01	0.48	0.02	0.59	0.06	0.69
	Univ. probit, censored	0.57	0.87	0.13	0.50	0.03	0.50	0.18	0.65	0.75	1.44

True parameter value is 0.00372.

Table A4. Estimates of ρ, Bivariate Probit, δ = 0, 60% Response Rate

	ρ									
	–0.8		–0.4		0.0		0.4		0.8	
N	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
200	0.097	0.133	0.065	0.221	–0.001	0.238	–0.046	0.212	–0.082	0.138
400	0.041	0.055	0.032	0.116	–0.009	0.126	–0.023	0.105	–0.033	0.051
800	0.011	0.024	0.010	0.056	0.001	0.064	–0.012	0.045	–0.008	0.017