Rendtel, Ulrich

On the choice of a selection-model when estimating regression models with selectivity

DIW Discussion Papers, No. 53

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Rendtel, Ulrich (1992) : On the choice of a selection-model when estimating regression models with selectivity, DIW Discussion Papers, No. 53, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/95758

DIW Diskussionspapiere
Discussion Papers

Discussion Paper No. 53

On the Choice of a Selection-Model When Estimating Regression Models With Selectivity

by
Ulrich Rendtel*
(Deutsches Institut für Wirtschaftsforschung)

Deutsches Institut für Wirtschaftsforschung, Berlin
German Institute for Economic Research, Berlin

Die in diesem Papier vertretenen Auffassungen liegen ausschließlich in der Verantwortung des Verfassers und nicht in der des Instituts.

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the Institute.

Deutsches Institut für Wirtschaftsforschung

Discussion Paper No. 53

On the Choice of a Selection-Model When Estimating Regression Models With Selectivity

by
Ulrich Rendtel*
(Deutsches Institut für Wirtschaftsforschung)

*) Helpful comments and suggestions by Ulrich Poetter and Mario Verbeek are gratefully acknowledge.

Berlin, June 1992

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, 1000 Berlin 33
Telefon: 49-30 - 82 991-0
Telefax: 49-30 - 82 991-200

1 Indroduction

It is well known that regression analysis may be affected by selectivity. Most commonly the case is treated, where the dependent variable Y is not observed for a substantial proportion in the sample while the information for the covariate vector X is available for all sample units. In this situation OLS-estimates based on the complete observations may be an inconsistent estimate for $E(Y \mid X)$ in the population.

Since situations where the dependent variable is missing in a substantial part of the data set occur rather frequently, there has been a great interest in procedures that promise to give consistent estimates of $E(Y \mid X) = X\beta$ in the presence of selectivity. The most popular procedure in this context is a two- step procedure, where first a probit model for the response variable R is estimated and secondly from the probit model $E(\varepsilon \mid X, R = 1)$ the expected value of the error term ε conditional that Y and X is observed, is estimated, which is added as a new variable to the regression equation. This enlarged regression equation is solved by OLS; see for example Heckman (1979), Lee (1982) and Olsen (1980). In order to estimate $E(\varepsilon \mid X, R = 1)$ one has to make some assumptions about the joint distribution of the error terms of the regression equation and the probit equation. Most commonly a bivariate normal distribution is assumed. In this case it is possible to calculate also ML estimates for the joint regression/probit model, see for example Amemiya (1984) and Nelson (1984). The use of the two-stage procedure has become very popular, since OLS and probit estimation routines are readily available.

This article deals with a topic relevant for both, the two-step procedure and MLE: The overlap of covariate vectors that explain Y and R. It seems that this topic is widely ignored by empirical researchers. While for the regression-equation in most cases exist careful considerations about the choice of covariates, only few ideas exist to explain the observability of Y. The only hint one gets from econometric literature is to avoid a perfect overlap of covariate sets in order to stabilize the two-step estimates. So mostly the choice of the R-covariates is some kind of ad-hoc solution and no special attention is given to the relationship of Y- and R-covariates.

As will be shown an improper choice of Y- *and* R-covariates yields inconsistent estimates of $E(\varepsilon \mid X, R = 1)$. Some simulation results reveal in such cases that in the resulting estimates for β are seriously biased.

What is mostly refered in literature is the fact the two-step procedure

yields consistent but very unstable estimates of β. So Nelson (1984) gives a comparison of asymptotic estimation errors from the two-step procedure and MLE. He states that the two-step procedure is especially imprecise in situations where the bias from OLS-estimates is biggest. With respect to asymptotic variance Nelson states that MLE is much more efficient than the two-step procedure. In a recent simulation study Stolzenberg and Relles (1990) compare the performance of OLS and the two-step procedure in situations where 90% of observations are incomplete. They resume that in 50% of the experiments the absolute deviation of estimates from the true value was increased by the two-step procedure in comparison with OLS. Similar results comparing the MSE of OLS and the two-step procedure were obtained by Zuehlke and Zeman(1991). Little and Rubin (1987,p.229) report the empirical results of Lillard et. al. (1982,1986) where the two-step procedure produced unrealistic high estimates for incomes.

But also ML estimation has been subject to critisism. Little and Rubin (1987,p.225ff.) stress the high sensibility of the ML estimates with concern to distributional assumptions about Y.

Only a small number of comments is found concerning the choice of the selection model. What is mostly refered, is the fact that the parameters of the two-step procedure are only identified by the nonlinearity of $E(\varepsilon \mid X, R = 1)$ in X if the covariate sets of both models coincide. From this reason researchers are advised to chose the set of R-covariates different from the Y- covariates, see for example Little and Rubin (1987,p.230).

It is the purpose of this paper to display the effect of different overlaps of the Y- and the R-model for the estimates of the regression model. By Monte-Carlo-simulation the distributions of three estimates are compared: OLS of complete data, the two- step procedure and MLE. In order to reduce the number of possible simulation parameters only some prototyps of overlap-combinations were regarded. These typical situations indicate that different sets of covariates may produce seriously biased estimates and standard errors for both the two-step procedure and MLE. Contrary to standard recommendations the advise is given that the covariate- sets for the selection- and the regression-model should coincide.

The article is organized as follows:

Section 2 refers the definition of nonignorable selection. In section 3 the two-step procedure and ML-estimation are introduced. Here also some theoretical results were formulated, that raise doubts about the statistical

foundation of these estimates. In section 4 the simulation study displays the serious numerical impact of the theoretical arguments. Section 5 concludes.

2 Ignorable and non-ignorable Selection

Let:

$$Y = X\beta + \varepsilon \tag{1}$$

with $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2$, where X is a vector of covariates. The unit index i is omitted for notational convenience. The interpretation $E(Y \mid X) = X\beta$ assures that the error term is uncorrelated with each covariate.

Some of the Y-values are missing, while the covariate vectors X are observed for all units. The fact that Y is observed is indicated by a dichotomous variable R, where $R = 0$ means that Y is missing and $R = 1$ stands for Y is observed.

The selection process R is said to be "*missing at random*"(MAR) or ignorable if the conditional distribution of $Y \mid X, R$ does not depend on R, cf. David 1979 a,b and Rubin 1976. In the language of error terms this is equivalent to the fact that the distribution of ε coincides for $R = 0$ and $R = 1$. The MAR- condition is satisfied if for each value of the covariate -vector the units were chosen (i.e. $R = 1$) by simple random sampling from the universe of all units with the same value of covariate vector. Thus the MAR-condition depends on the population we want to make inferences for. The MAR-condition is violated if Y depends on some other variable , say V, not included in the covariate set and R depends on V.

Since we are dealing with the selection of covariates it is worth to remember that:

1. Omitting covariates from the Y-model may destroy the MAR- property
2. But also adding covariates to the Y-model may introduce a dependence of $Y \mid X, R$ on R.

Let $\hat{\beta}$ be some standard estimator of β, for example the OLS-estimator. Denote by $\hat{\beta}_{ObS}$ the estimate of β where $\hat{\beta}$ is applied to all units where Y is observed. If $\hat{\beta}$ is a consistent estimate for β if no selection occurs then under

the MAR-condition $\hat{\beta}_{OLS}$ is consistent for β too. Thus under the MAR-condition $\hat{\beta}_{OLS}$ converges to β if number of units where Y is observed tends to infinity.[1] Sometimes much weaker conditions, depending on $\hat{\beta}$, guarantee the consistency of $\hat{\beta}_{OLS}$, cf. Manski 1989 and Verbeek/Nijman 1980.

Although the MAR-condition is easily stated, there is little chance to check this condition unless one has a-priori information about the distribution of $Y \mid X, R = 0$.

3 The two-Step Procedure and the MLE in the presence of different covariates for the Y- and the R-model

In order to get consistent estimates for β in the case of nonignorable selection one usually assumes a joint model for Y and R. The standard model is a threshold model for R

$$R^* = Z\gamma + \delta \tag{2}$$

$$R = \begin{cases} 1 & \text{if } R^* > 0 \\ 0 & \text{if } R^* \leq 0 \end{cases} \tag{3}$$

where Z is a vector of covariates to explain missing Y-values and δ is an error-term. δ and ε may be correlated.

The starting point of the two-step procedure is that only $Y \mid X, R = 1$ is observed and hence OLS estimates only $E(Y \mid X, R = 1)$. By (1) we get:

$$E[Y \mid X, R = 1] = X\beta + E(\varepsilon \mid X, R = 1) \tag{4}$$

[1] One should not forget that the variance of $\hat{\beta}_{OLS}$ may be grossly inflated if the selection-process produces a bad conditioned moment matrix. In extreme cases it may happen that the effect of certain covariates cannot be estimated, since in the selected sample there is no variance with respect to that variable.

If we use the decomposition[2] $\varepsilon = \sigma_{\varepsilon\delta}\delta + \eta$, where η is independent from δ and $\sigma_{\varepsilon\delta}$ is the covariance between ε and δ, we have:

$$E[\varepsilon \mid X, R = 1] = \sigma_{\varepsilon\delta} E[\delta \mid X, Z\gamma + \delta > 0] + E[\eta \mid X, Z\gamma + \delta > 0] \quad (5)$$

Most textbooks proceed by equating:

$$E(\eta \mid X, Z\gamma + \delta > 0) = E(\eta) = 0 \quad (6)$$

and

$$\begin{aligned} E(\delta \mid X, Z\gamma + \delta > 0) &= \phi(Z\gamma)/\Phi(Z\gamma) \\ &=: H(Z\gamma) \end{aligned} \quad (7)$$

where ϕ is the density function of the standard normal distribution and Φ is its cumulative. Inserting (7) into (4) one gets:

$$E(Y \mid X, R = 1) = X\beta + \sigma_{\varepsilon\delta} H(Z\gamma) \quad (8)$$

On the basis of (8) among others Heckman (1979) suggested the following two-step procedure:

Step 1: Estimate γ from all units by probit- estimation and calculate from $\hat{\gamma}$ variable $H(Z\hat{\gamma})$ for all units with $R = 1$.

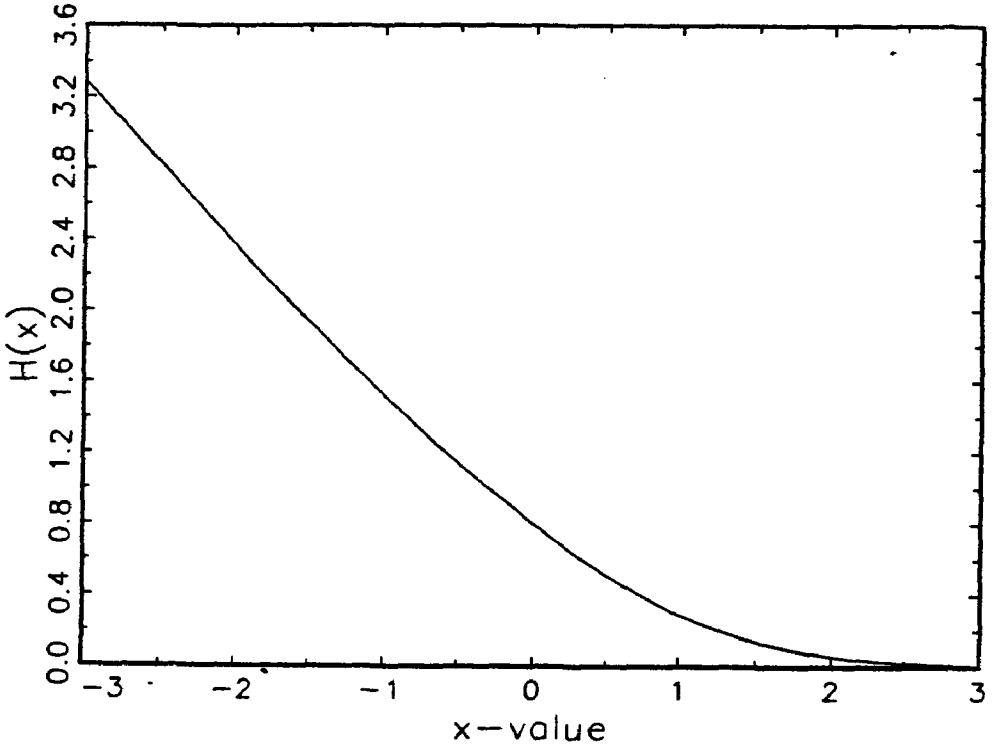
Step 2: Estimate β and $\sigma_{\varepsilon\delta}$ by OLS where Y is regressed upon X and H.

If the X and Z variable coincide, the identification of the parameters β and $\sigma_{\varepsilon\delta}$ depends solely on the non-linearity of $H(\cdot)$. But as shown in Figure 1 $H(X)$ behaves in broad ranges like an almost linear function, so the estimates of β may become very unstable.

In order to remove this instability many textbooks (for example Rubin/Little 1987, p.222) recommend to add a least one variable.

[2]If ε and δ are bivariate - normal with zero- means all required distributional assumptions are fullfilled. Indeed this decomposition may be derived under weaker distributional assumptions. If suffices that the conditional distribution of $\varepsilon \mid \delta$ is linear in δ and δ is normal with zero mean, see Olsen (1980).

Figure 1: The Quasi-Linearity of $H(X) = \phi(X)/\Phi(X)$



It has to be stressed that equations (6) and (7) may **not** hold and hence the results of the two-step procedure may be misleading.

Lets first consider the case $X = Z$, i.e. the set of covariates for Y and R are the same. Regarding $E(\eta \mid X, X\gamma + \delta > 0)$ we know that by construction δ is independent from η. But we don't know whether X and η are independent or at least orthogonal. All we know is that X and $\varepsilon = \sigma_{\varepsilon\delta}\delta + \eta$ are orthogonal. But from this we may *not* conclude that X is orthogonal to each of the components δ and η. Hence one may construct examples were eq.(6) and (7) do not hold.

Next, suppose that an additional variable, say V, is added to the set of R-covariates. In this case one has to compute $E(\delta \mid X, X\gamma_x + V\gamma_v + \delta > 0)$ and $E(\eta \mid X, X\gamma_x + V\gamma_v + \delta > 0)$. To do this one needs information about the joint distribution of V,X and δ and V,X and η which is not available, since δ and η are not observed. All that is known about δ and η is that $\varepsilon = \sigma_{\varepsilon\delta}\delta + \eta$ is orthogonal to the X-vector. The relation of δ and η to other variables is not

determined. Even if we assume that V is independent from δ, η and the X-vector, there is still the problem to evaluate the two conditional expectations, since V is not fixed by the conditioning.

Now suppose some of the X-variables, say W, is not included in the computation of the selection correction variable H. If the corresponding γ-coefficient γ_w in eq. (2) is different from zero, then $\gamma_w W$ becomes part of a new error- term $\tilde{\delta} = \delta + \gamma_w W$. Since ε is orthogonal to W it is not possible to find a decomposition $\varepsilon = \sigma_{\varepsilon\tilde{\delta}}\tilde{\delta} + \tilde{\eta}$, where $\tilde{\eta}$ is independent from $\tilde{\delta}$. Hence also the exclusion of X-variables from the Z-variables may destroy the justification for the use of the two-step procedure.

Now we treat the case of MLE. If we assume a bivariate normal distribution

$$\begin{pmatrix}\varepsilon\\ \delta\end{pmatrix} \sim N\left[\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\sigma_\varepsilon^2 & \rho_{\delta\varepsilon}\sigma_\varepsilon\\ \rho_{\delta\varepsilon}\sigma_\varepsilon & 1\end{pmatrix}\right] \tag{9}$$

for the error terms of the joint Y-and R-model, then one may compute the ML-solution for the joint model. Again we assume that the R-model contains an additional variable V that is not part of the Y-model. In order to compute the likelihood for the unknown parameter $\vartheta = (\beta, \gamma, \sigma_\varepsilon, \rho_{\varepsilon\delta})$ one has to distinguish the cases $R = 1$ and $R = 0$.
For $R = 0$ we observe only X and V. Hence:

$$\begin{aligned} L(\vartheta \mid X, V, R = 0) &= P(R = 0 \mid X, V) \\ &= P(X\gamma_x + V\gamma_v + \delta > 0 \mid X, V) \\ &= \Phi(X\gamma_x + V\gamma_v) \end{aligned} \tag{10}$$

For $R = 1$ we observe Y,X and V. Hence:

$$L(\vartheta \mid Y, X, V, R = 1) = P(R = 1 \mid Y, X, V) f(Y \mid X, V) \tag{11}$$

The problem to compute the likelihood arises from the second factor, which is the density of Y given X and V. Since $Y \mid X \sim X\beta + \varepsilon$ one has to know the distribution of $\varepsilon \mid V$. Only if ε and V are independent we have $f(Y \mid X, V) = f(Y \mid X)$ what is used in Amemiya (1984, p.32). But ε is unobserved, so there is no chance to check a condition like $E(\varepsilon V) = 0$. If $f(Y \mid X) \neq f(Y \mid X)$ the use of $f(Y \mid X)$ yields not a proper

likelihood function. Under the model- assumptions $P(R = 1 \mid Y, X, V)$ may be computed.

It follows from (11) that:

$$\delta \mid \varepsilon \sim N(\rho_{\delta\varepsilon}\varepsilon/\sigma_\varepsilon, 1 - \rho^2_{\delta\varepsilon}) \tag{12}$$

Conditional on Y,X (which together fix ε) and V one may treat $(X\gamma_x + V\gamma_v$ as a constant and the distribution of δ is given by (13).

$$P(R = 1 \mid Y, X, V) = \Phi\left[1/\sqrt{1 - \rho^2_{\delta\varepsilon}}\left(\frac{\rho_{\delta\varepsilon}}{\sigma_\varepsilon}(Y - Y\beta) + X\gamma_x + V\gamma_v\right)\right] \tag{13}$$

Once one has reasonable starting values for $\beta, \gamma, \sigma_\varepsilon$ and $\rho_{\delta\varepsilon}$ the ML- estimate may be obtained by using standard maximation techniques. Starting values for β and σ_ε may be obtained by OLS of the Y-model, while the probit estimate of the R-model provides a starting value for γ. The only parameter at free guess is $\rho_{\delta\varepsilon}$.

In Econometric literature it is often stated that a selection rule is ignorable if and only if $\rho_{\delta\varepsilon} \neq 0$, since in these cases the two-step procedure is equivalent with OLS and $P(R = 1 \mid Y, X, V)$ becomes independent from β in the ML-case. It is worth to remember that this is only true in the case where $X = Z$. As will be shown by an example in the next paragraph one may immediately construct situations where by switching unobserved variables to observed ones $\rho_{\delta\varepsilon} = 0$ changes to $\rho_{\delta_\epsilon} \neq 0$ although the selection process is still the same.

4 The stability of estimated Y-model parameters when the R- model changes

In this section we want to see by means of a simulation study the numerical effects of a different overlap of the R-covariate set and the Y-covariate set. For all different overlap-combinations the distribution of the estimates for the Y-model are inspected for OLS, the two-step procedure and the MLE.

4.1 Parameters of Simulation run

The simulation results are generated from the following equation:

$$Y = 0.0 + 2X_1 - 2X_2 + D + v + e \tag{14}$$

$$R^* = 0.2 + 0.5X_1 + 0.5X_2 + D + v + d \tag{15}$$

Y is observed, if $R^* > 0$. X_1, X_2, D, v, e, and d were generated independently although for different estimates the same values for X_1, X_2, D, V, e and d were used. With the exception of D which is 0-1-variable with $P(D = 1) = 0.5$ all variables were generated from a standard normal distribution. The variables X_1, X_2 and D were treated as Y-covariates. Thus $\varepsilon = v + e$ may be regarded as error terms.[3]

The motivation for this simulation design is as follows: Dummy variables rather frequantly occur in empirical work but rather seldom in simulation studies. One may expect that for $D = 1$ there is a high probability to observe Y while for $D = 0$ the probability to observe Y is much smaller. This reduces the variance of D in eq. (14) a great deal. So one expects that the selection- effects for the estimate of β_D are most severe namely one expects a substantial trade-off with the estimate for the constant β_0. The variance of the continuous contributions $0.5X_1$ and $0.5X_2$ in eq. (15) are the same as the variance of D. Because of their continuous nature there remains a lot of variance for $R = 1$ for X_1 and X_2 in eq. (14). This variance is enlarged by the coefficients $\beta_{X_1} = 2$ and $\beta_{X_2} = -2$ So one expects that selection-effects for the estimate of β_{X_1} and β_{X_2} are much smaller than for β_D. The different signs of β_{X_1} and β_{X_2} compensate for the fact that Y is observed mostly for high values of X_1 and X_2. The covariates were chosen to be independent in order to disentangle the effects of different covariates on the estimation of eq. (14).

Each simulation generates a sample of $n = 400$ units, which seems a reasonable size in empirical work. The value for the constant in the selection equation was chosen to meet the requirement that 1/3 of the Y- values are missing.[4] 100 replications for each overlap combination were run.

[3]The ratio of explained to total variance is $8.25/10.25 \approx 0.80$.

[4]The standard deviation of that drop -out rate was 2.2%.

4.2 Same covariates in Y-and R-model

First we investigate the case were the covariates of the Y- and the R-model are the same.

In this case one would regard $\delta = d + v$ as error terms in the threshold-model. Because of the joint component ε and δ are correlated. Hence the selection is non-ignorable and the estimation of the Y-model by OLS is inconsistent.

Table 1 compares the estimates from OLS, the two-step procedure and MLE. The first column in table 1 displays the average of the estimates over the replications, while the second column displays the standard deviation of the estimates. In the third column the mean of the estimated standard errors is displayed.

OLS		d(beta)	std(beta)	d(sig)
	consty	0.4958	0.1191	0.1309
	X1	1.8785	0.0827	0.0844
	X2	-2.1233	0.0928	0.0839
	D	0.7807	0.1518	0.1667
two-step		d(beta)	std(beta)	d(sig)
	consty	0.0831	1.0689	1.0559
	X1	1.9669	0.2616	0.2442
	X2	-2.0362	0.2611	0.2399
	D	0.9859	0.5415	0.5342
	H	0.5716	1.4852	1.4625
ml		d(beta)	std (beta)	d(sig)
	consty	0.2166	0.5390	0.3738
Y-	X1	1.9361	0.1437	0.3738
model	X2	-2.0624	0.1468	0.1178
	D	0.9144	0.2840	0.2482
	constr	0.1390	0.0896	0.0937
R-	X1	0.3621	0.0688	0.0735
model	X2	0.3564	0.0715	0.0734
	D	0.7554	0.1220	0.1417
	$\rho_{\varepsilon\delta}$	0.2605	0.4856	0.3080
	σ_ε	1.4293	0.1252	0.1147

Table 1: Same covariates in Y- and R-model

d(beta) = Mean of estimates
Std(beta) = Std. Derivation of estimates
d(sig) = Mean of estimated standard errors

Table 1 reveals that the OLS-estimates of the normally distributed covariates X_1 and X_2 are somewhat downbiased. But the relative error is only of small size (about 6%). On the other hand the coefficient for the dummy - variable D is seriously downbiased (about 22 %), while the constant of Y-model is grossly over-estimated. If we measure the bias in units of the standard deviation for the estimates, we see, that the effect of all covariates is downbiased about 1.5 of their standard deviation. Comparing column 2 and 3 we see, that the estimates of the standard estimation are not effected by the selection rule.

Now we switch to the two-step procedure. The means of the estimates for the Y-model fit the theoretical values nearly perfectly.

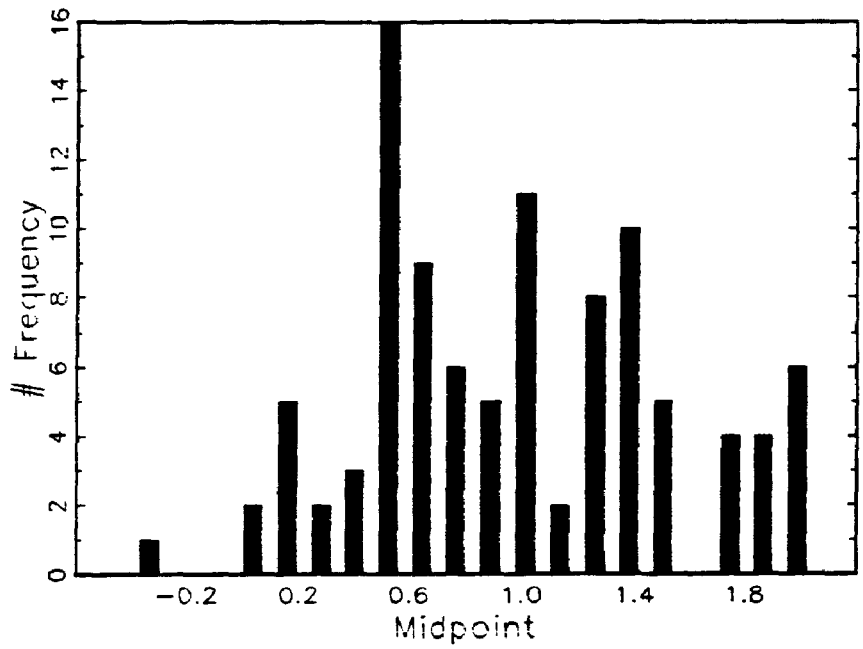


Figure 2: The distribution of $\hat{\beta}_D$ from the two-step procedure. Histogram of 100 estimates. R-covariates equal Y-covariates.

But this appealing result is misleading, since the standard deviation of all estimates are increased by a factor of about 3.0 with respect to the OLS-value (for the constant this factor amounts 8.0). Moreover the distribution of the estimates does not look unimodal with the modus given by the average. For example look at the histogram of the estimates of β_D in Figure 2. This histogram seems to belong to a rectangular distribution rather than to a normal distribution. From the standard deviation of estimates one would conclude that the effect of the discrete covariate is insignificant and that there is no significant covariance $\sigma_{\varepsilon\delta}$. Hence one would conclude that the selection process is ignorable.

The results for the ML-estimation are some what in between the OLS- and the two-step estimates. The bias of the OLS-estimates is reduced by factor 0.5 and the standard deviation of the estimates is only half of the corresponding value for the two-step estimate (but still about 1.5 times bigger than the OLS- values). Also the ML-estimate of $\rho_{\delta\varepsilon}$ is insignificant.

4.3 Additional variables in the R-model

In order to reduce the variance of the two step-procedure (but also in case of the ML-estimate) an additional variable is incorporated into the R-model.

There are 4 possible cases: The added variable is:

(i) correlated with Y and R
(ii) correlated with Y but not with R
(iii) correlated with R but not with Y
(iv) uncorrelated with R and Y

In the simulation these cases were achieved by adding to the R- model: v(case i), e(case ii), d(case iii). For case (iv) an independent standard normal variable z was generated.

Before we switch to the corresponding estimation results, a remark concerning the error terms is to be made. In case (i) the error term of the R-model is given by $\delta = d$. The error term of the Y-model is $\varepsilon = v + e$, hence the error terms of the Y- and the R-model are independent. In case (iii) we have for the error terms $\delta = v$ and $\varepsilon = v + e$. Hence the error terms are correlated. But the generated selection process is in both cases the same. If we would judge the MAR- condition on the basis of $\rho_{\delta\varepsilon}$ we would conclude that in case (i) the selection is ignorable and in case (iii) it is not ignorable.

Table 2 displays the results for the two-step procedure and the ML-estimate for the case (i) to (iv). Since the OLS estimate is independent from the R-model, the OLS results coincide with table 1. We start with case (iv), where the added variable z is independent from Y and R. As one expects, this does not influence the estimation and therefore the corresponding results in table 1 and 2 are almost equal. If the added variable is correlated with R but uncorrelated with Y (case iii) then the two step-procedure and the ML- estimate perform very well. The high standard errors of estimate are reduced to the level of the OLS- estimates.[5] Also the distribution of the estimates - which is not displayed here - turns out to be approximate normal. The coefficients of the Y-model are estimated with high accuracy. Also the estimated value for $\rho_{\delta\varepsilon}$ is in perfect accordance with the theoretical value from the simulated error terms (which is $1/\sqrt{2} = 0.70$)

In case (i) - were the added variable is correlated with Y and R a strong negative correlation between ε and δ is estimated $[\hat{\rho}_{\delta\varepsilon} = -0.9]$, which is far away from the true value. The resulting estimates for the Y-model are downbiased. Compared with the OLS-results from table 1 the effect of selectivity is enlarged. For example the effect of the discrete variable D is estimated by OLS to be 0.78, while it is 0.45 by ML and 0.17 by the two-stage procedure.

[5]Comparing column 2 and 3 in case 3 (iii) we see that the estimated error of the two-step procedure are a little (about 10%) downbiased. This is due to the fact, that the variation of results depending on the estimate of the R-model is ignored. See Lee et. al. (1980) for an un biased estimate of the standard errors.

Table 2

	case (i)			case (ii)		
two-step	d(beta)	std(beta)	d(sig)	d(beta	std(beta)	d(sig)
consty	1.7225	0.2034	0.1590	-0.1675	4.5139	0.8707
X1	1.6114	0.0918	0.0751	2.0626	0.9948	0.2076
X2	- 2.3805	0.1107	0.0745	-19730	0.9938	0.2099
D	0.1796	0.1967	0.1509	1.0685	2.1012	0.4338
H	-2.4041	0.3057	0.2254	0.8966	6.2340	1.1922
ml	d(beta)	std(beta)	d(sig)	d(beta)	std(beta)	d(sig)
consty	1.1121	0.1283	0.1362	-0.2339	0.8700	0.1540
X1	1.7322	0.0766	0.0848	2.0823	0.2244	0.0963
X2	-2.2613	0.0865	0.0845	-1.9284	0.2369	0.0961
D	0.4526	0.1709	0.1678	1.1128	0.4586	0.1897
constr	0.1100	0.1100	0.1055	0.0437	0.0983	0.0904
X1	0.4325	0.0904	0.0838	0.3038	0.0654	0.0671
X2	0.4286	0.0895	0.0841	0.3082	0.0684	0.0666
D	0.8908	0.1766	0.1634	0.6057	0.1355	0.1311
v/e/d/z	1.0750	0.1026	0.0966	-0.3083	0.3683	0.0529
$\rho_{\varepsilon\delta}$	-0.9022	0.0358	0.0340	0.6354	0.7155	0.0182
σ_{ε}	1.4164	0.0640	0.0656	1.6618	0.0785	0.0869

continuation table 2

	case(iii)			case(iv)		
two-step	d(beta)	std(beta)	d(sig)	d(beta)	std(beta)	d(sig)
consty	0.0019	0.1980	0.1874	0.2247	1.0192	0.9181
X1	1.9914	0.0961	0.0878	1.9641	.2481	0.2194
X2	-2.0086	0.0970	0.0871	-2.0351	0.2453	0.2226
D	1.0190	0.1835	0.1753	0.8955	0.5139	0.4589
H	0.9770	0.2676	0.2629	0.4080	1.3582	1.2587
ml	d(beta)	std(beta)	d(sig)	d(beta)	std(sig)	d(sig)
consty	-0.0031	0.1730	0.1640	0.2818	0.5312	0.3570
X1	1.9906	0.0926	0.0885	1.9524	0.1491	0.1164
X2	-2.0079	0.0927	0.0881	-2.0498	0.1613	0.1176
D	1.0229	0.1771	0.1757	0.8611	0.2987	0.2360
constr	0.1937	0.1129	0.10182	0.1344	0.0956	0.0934
X1	0.5145	0.0826	0.0879	0.3664	0.0714	0.0735
X2	0.4953	0.0779	0.0871	0.3692	0.0657	0.0731
D	1.0585	0.1744	0.1699	0.7304	0.1435	0.1417
v/e/d/z	1.0232	0.1094	0.1080	0.0025	0.0806	0.0679
$\rho_{\varepsilon\delta}$	0.6963	0.1180	0.1050	0.2309	0.4899	0.2942
σ_ε	1.4025	0.0705	0.0712	1.4049	0.1034	0.1112

continuation table 2

	case (v)			case (vi)		
ols	d(beta)	std(beta)	d(sig)			
consty	0.5267	0.0809	0.0868			
X1	1.8925	0.0609	0.0564			
X2	-2.1137	0.0574	0.0565			
D	0.7414	0.1085	0.1113			
e/m	0.9884	0.0509	0.0554			
two-step	d(beta)	std(beta)	d(sig)	d(beta)	std(beta)	d(sig)
consty	-0.0192	0.7076	0.6882	1.7121	0.1860	0.1559
X1	2.0132	0.1677	0.1620	1.6307	0.0981	0.0748
X2	-1.9885	0.1640	0.1630	-2.3738	0.1004	0.0749
D	0.9997	0.3739	0.3379	0.1786	0.1800	0.1490
e/m	0.9856	0.0694	0.0630	-	-	-
H	0.7560	0.9381	0.9445	-2.3275	0.2604	0.2204
ml	d(beta)	std(beta)	d(sig)	d(beta)	std(beta)	d(sig)
consty	0.1273	0.3092	0.1985	1.1383	0.1273	0.1351
X1	1.9810	0.0966	0.0742	1.7451	0.0809	0.0845
X2	-2.0222	0.0894	0.0737	-2.2565	0.0866	0.0844
D	0.9277	0.1950	0.1465	0.4234	0.1573	0.1668
e/m	0.9875	0.0586	0.0599	-	-	-
constr	0.1320	0.0937	0.0930	0.0880	0.0914	0.1056
X1	0.3645	0.0676	0.0728	0.4289	0.0802	0.0841
X2	0.3663	0.0688	0.0723	0.4285	0.0836	0.0829
DIS	0.7316	0.1381	0.1404	0.8785	0.1627	0.1639
e/m	-0.0022	0.0763	0.0695	0.2681	0.0347	0.0239
$\rho_{\varepsilon\delta}$	0.5466	0.3860	0.2046	-0.8991	0.0351	0.0351
σ_ε	0.9845	0.0761	0.0787	1.4011	0.0543	0.0654

continuation table 2

case (vii)

ols	d(beta)	std(beta)	d(sig)
consty	0.0142	0.1076	0.1041
X1	2.0061	0.0647	0.0638
X2	-1.9974	0.0644	0.0639
D	o.9777	0.1251	0.1262
e/m	0.2281	0.0174	0.0176
two-step	d(beta)	std(beta)	d(sig)
consty	-0.0133	0.2751	0.3627
X1	2.0081	0.0878	0.0835
X2	-1.9991	0.0904	0.0839
D	0.9781	0.1688	0.1704
e/m	0.2479	0.0328	0.0323
H	0.0003	0.3601	0.3629
ml	d(beta)	std(beta)	d(sig)
consty	0.0041	0.2480	0.2307
X1	2.0098	0.0810	0.0792
X2	-1.9964	0.0858	0.0796
D	0.9817	0.1630	0.1608
e/m	0.2490	0.0292	0.0293
constr	0.1806	0.1049	0.1096
X1	0.5155	0.0828	0.0910
X2	0.5104	0.0871	0.0900
D	1.0466	0.1758	0.1763
e/m	0.2577	0.0294	0.0279
$\rho_{\varepsilon\delta}$	0.0131	0.3059	0.3003
σ_{ε}	0.9964	0.0454	0.0509

Table 2: Adding a variable to the R-Model.
The added variable is:
(i) v added, correlated with R and Y
(ii) e added, correlated with Y but not with R
(iii) d added, correlated with R but not with Y
(iv) z added, uncorrelated with R and Y
(v) e added to the R- and Y-model
(vi) $m = 1/\sqrt{4}(v + d + e + z)$ added to the R-model
(vii) m added to the R- and Y-model

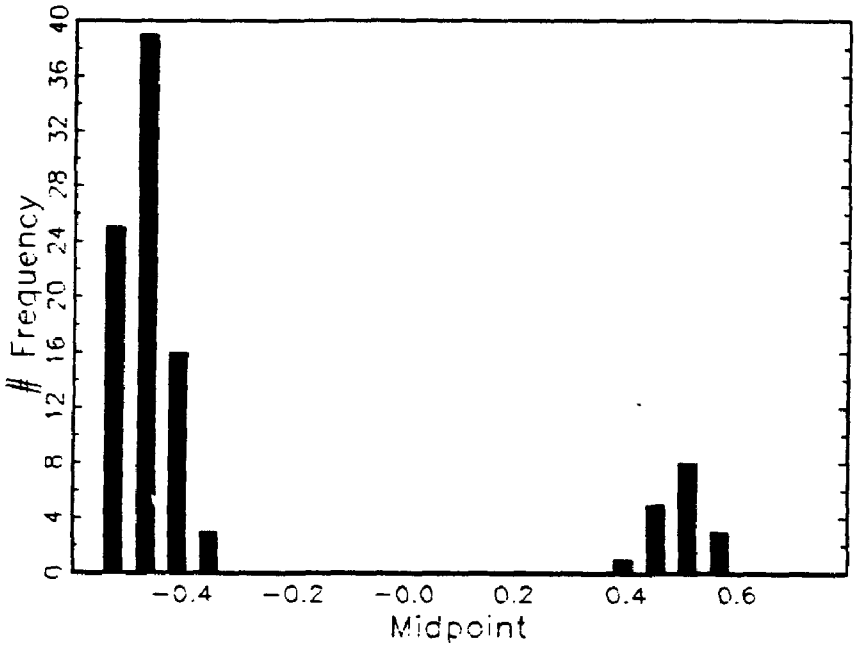


Figure 3a: Simulated distribution of ML-estimate of γ_e, where e was added to R-model. e uncorrelated with R but correlated with Y.

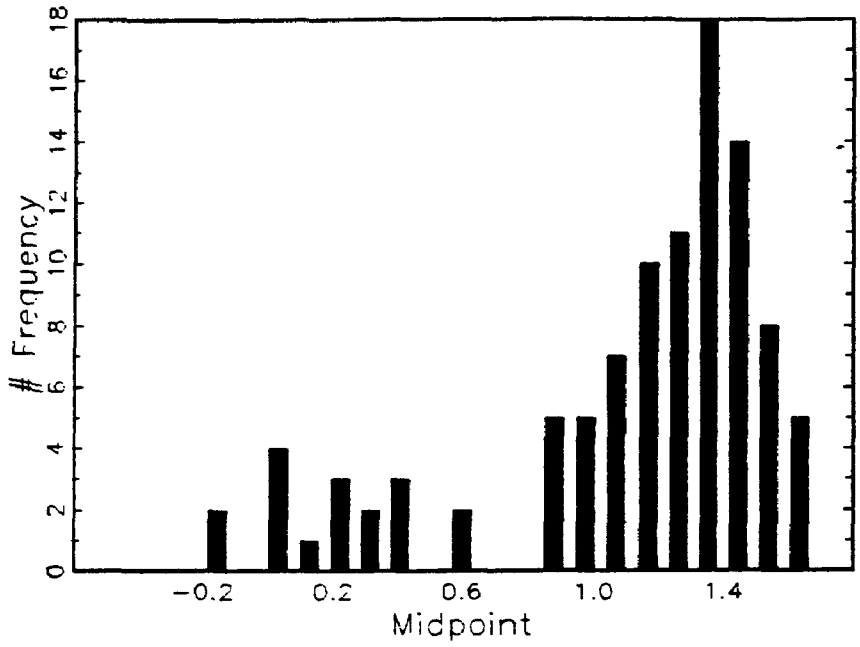


Figure 3b: Simulated distribution of ML-estimate of β_D. Added to the R-model: e uncorrelated with R but correlated with Y).

In case (ii) the estimated correlation between ε and δ was on the average 0.63. Hence the resulting estimates for the Y-model are corrected into the right direction but on the average the results are positively biased. If we take a look at the estimated and the simulated standard errors of estimates, we see that estimated standard errors are grossly downbiased. The simulated standard errors are even much bigger than in the case, where the Y- and R-variables coincide. If one takes a look at the shape of the distribution of the estimates, they are far from being unimodal or normal. The ML-estimates exhibit a clear bimodal pattern, which is shown in Figure 2a and 2b. Figure 2a shows the distribution of the ML-estimate for γ_e, the effect of e on R^*. This is clearly different from the theoretical value 0. Figure 2b demonstrates the corresponding bimodality of the estimate of β_D. So we have to realize that the most frequent estimate for β_D is aproximately 1.4, which is much higher than the mean value (1.1). Hence most frequently the bias due to high

estimated correlations between the Y- and the R- error terms is much bigger than indicated by the mean values in Table 2 case (iii). The bimodality of the ML-estimates results from a large proportion of $\rho_{\varepsilon\delta}$-estimates about 0.95 and a minor proportion, where $\hat{\rho}_{\varepsilon\delta}$ is about -0.95. Thus the addition of e into the R-model makes the estimation of the joint model nearly collapse.

In empirical work one is rather seldom in a position to know a- priori the relationship of the added variable to R and Y. Hence - because of the severe consequences, if the added variable is correlated with Y- one should check for this. This brings one back to the situation, where the Y- and the R- covariates coincide.

If the added variable turns out to be insignificant in the Y-model but significant in the R-model, it may be dropped from the Y-model in order to stabilize the estimates. If the added variable is significant in the Y-model this may considerably improve the Y- model and thus reduce the standard errors. For example, if e is included in the Y-model, case (v) in table 2, then the standard errors of MLE are about the same size as in case (iii), which proved to be optimal for cases (i) to (iv). This is not true for the two-step procedure. Also the means of the estimates now are very near to the true values.

The cases (i) to (v) are extreme cases to display the possible effect of adding a variable to the R-model. In practice one would expect intermediate cases, where the added variable is some combination of v, d, e and z. This situation is displayed in cases (vi) and (vii) in table 2 (continued), where the additional variable is $m = 1/\sqrt{4}(v + d + e + z)$. In case (vi) m was added to the R-covariates, while in case (vii) m was added also to the Y-covariates. As table 2 reveals the two-step procedure and MLE do not remove the selection bias if m is included in R-model but not the Y-model, case (vi). The selection bias is even bigger than in the OLS- estimate.

If m is included in the Y-model too, one realizes that the OLS- selection-bias has vanished. This due to the fact that m is correlated with R. Here the situation is different from case (v), where the added variable was not correlated with R and consequently no reduction of OLS-selection-bias was achieved. Of course also the two-step procedure and MLE provide accurate estimates but the variance of their estimates is bigger (in case of the constant much bigger) than the corresponding OLS- variances.

4.4 Omitting a variable from the R-model

Now we treat the case that one variable of the Y-covariates is not included in the R-model. This may be due to fact that this variable is not known for all units. It may happen for example that the omitted variable is missing whenever Y is missing. Alternatively one may be motivated to omit a variable in order to reduce the instability of the two-step- or the ML- estimate. Since the results are very similar for all covariates only the case, where D is omitted from the R-model is displayed in table 3.

Table 3

two-step		d(beta)	std(beta)	d(sig)
	consty	0.1577	1.5126	1.3082
	X1	1.9711	0.3978	0.3785
	X2	-2.0190	0.4005	0.3712
	D	0.7835	0.1623	0.1671
	H	0.6179	2.7808	2.3901
ml		d(beta)	std(beta)	d(sig)
	consty	0.2213	0.4125	0.3255
Y-	X1	1.9585	0.1328	0.1290
model	X2	-2.0362	0.1480	0.1262
	D	0.7761	0.1611	0.1656
	constr	0.4778	0.0691	0.0686
R-	X1	0.3370	0.0680	0.0707
model	X2	0.3281	0.0651	0.0705
	$\rho_{\varepsilon\delta}$	0.3330	0.4546	0.3407
	σ_ε	1.4446	0.1171	0.1260

Table 3: D omitted in the R-model

Comparing table 1 and table 3 one realizes that the two-step procedure and the ML-estimate have lost their ability to correct for selection when β_D is estimated. Also the standard errors for the other β-values estimates remain at high levels-like in table 1. Only the standard error of $\hat{\beta_D}$ has decreased. Thus omitting variables to stabilize the two-step estimates or MLE is a pretty bad idea. It should be mentioned that the independence of the covariates in the simulation experiments comes into play here. If some covariates are correlated, the omission of one covariate in the R-model may be compensated by another covariate that remains in the R-model.

4.5 The mixed case

Now we deal with the mixed case, where some covariate of the Y- model is omitted, while some other variable is added to the R- model. This case appears to be the most frequent in empirical applications.

Table 4

	case (i)			case (ii)		
two-step	d(beta)	std(beta)	d(sig)	d(beta	std(beta)	d(sig)
consty	1.4337	0.1677	0.1308	-0.4753	6.0432	0.9239
X1	1.5884	0.1036	0.0733	2.2052	1.6679	0.2731
X2	-2.4108	0.1052	0.0733	-1.8515	1.6880	0.2749
D	0.9839	0.1262	0.1364	0.7338	0.1505	0.1512
H	-2.5424	0.2916	0.2184	1.7091	10.7944	1.6425
ml	d(beta)	std(beta)	d(sig)	d(beta)	std(beta)	d(sig)
consty	0.9249	0.1154	0.1247	-0.0975	0.5937	0.1339
X1	1.7331	0.0822	0.0854	2.1018	0.2137	0.0978
X2	-2.2662	0.0871	0.0854	-1.9015	0.2133	0.0975
D	0.9012	0.1340	0.1452	0.6948	0.1380	0.1386
constr	0.4809	0.0709	0.0793	0.3250	0.0677	0.0673
X1	0.377	0.0646	0.0774	0.2867	0.0622	0.0655
X2	0.3876	0.0696	0.0776	0.2904	0.0644	0.0650
v/e/d	0.9726	0.0862	0.0856	-0.3391	0.3322	0.0506
$\rho_{\varepsilon\delta}$	-0.9093	0.0328	0.0306	0.7117	0.6440	0.0167
σ_ε	1.4327	0.0628	0.0681	1.6998	0.0800	0.0898

continuation table 4

	case (iii)		
heck	d(beta)	std(beta)	d(sig)
consty	0.1735	0.1607	0.1592
X1	2.0100	0.0978	0.0880
X2	-1.9878	0.0966	0.0882
D	0.6583	0.1604	0.1634
H	0.9738	0.2760	0.2652
ml	d(beta)	std(beta)	d(sig)
consty	0.1952	0.1344	0.1441
X1	2.0089	0.0902	0.0893
X2	-1.9888	0.0969	0.0894
DIS	0.6233	0.1618	0.1629
constr	0.6199	0.0723	0.0818
X1	0.4585	0.0821	0.0817
X2	0.4517	0.0696	0.0806
v/e/d	0.9004	0.0826	0.0956
$\rho_{\varepsilon\delta}$	0.6893	0.1060	0.1074
σ_ε	1.4066	0.0715	0.0759

Table 4: deleted from the R-model: D
Added to the R-model:
case (i): v
case (ii): e
case (iii): d

In table 4 D was deleted from the R-model and v (case i), e (case ii), or d (case iii) was added. Note that the added and omitted variable are independent. So one would expect that the effects of adding and omitting these variables are independent too, like in OLS-estimation. For example, deleting D from R-model as well as adding v effect a serious underestimation of β_D. Thus one would expect that in the mixed case β_D is underestimated too. But, as table 4 reveals, this is not true: $\hat{\beta}_D$ is with high precision near the true value.[6] Still the estimates of β_{x1} and β_{x2} are downbiased in the same

[6]Analogues results were obtained when X_1 and X_2 were omitted from the R-model.

way as in table 2. In cases (ii) and (iii) the effects of omitting and including independent variables are disentangled. This appears to be reasonable, since in these cases the included variable effects only R or Y. In case (i) v and D both effect R and Y, and hence unexpected interaction-effects may arise. Note that this case seems to be typical one is applied work.

5 Conclusions

As demonstrated by the simulation results the choice of the R- model has a great impact on the estimation of the Y-model.

Omitting Y-variables in the R-model may yield inconsistent estimates for the Y-model. Thus one is advised to use al least the Y-covariates in the R-model. This may be impossible if some of the Y-covariates are subject to selectivity too. If the Y-variables and R-variables are the same, the two-step procedure becomes very unstable and the distribution of estimates is far from being normaly or even unimodal. But also the MLE standard errors are markedby bigger than the corresponding OLS-values. If one calculates the MSE the simulation results, the two-step procedure yields about 4 times the OLS-MSE, and the MLE-amounts to 1.3 times the OLS-MSE. [7] The simulation results strongly suggest that the inclusion of additional covariates into the R-model but not into the Y-model is a hazardous - strategy. If one is able to find a variable that is correlated with R but not with Y, the two-step procedure and MLE perform quite well: The true values for β are estimated without selection bias and the variances of $\hat{\beta}$ are only slightly bigger then the corresponding OLS-values. In terms of MSE one gets less than half of the OLS-MSE if β_D is regarded.[8] In this case the two-step procedure operates nearly as efficient as MLE.

As the simulation results indicate, the estimation of the Y-model may become very poor if the added variable is correlated with Y. The bias of estimates may become even bigger than selection-bias of the OLS-estimate. In terms of MSE the two-step-procedure yields about 10 times and MLE 5 times the OLS -value for β_D. [9] In order to avoid wrong conclusions from two-step or ML-estimation one is strictly advised to check the correlation of

[7] Values computed from table 1, columns 1 and 2.

[8] Values computed from table 1 and table 2 case (iii).

[9] Values computed from table 1 and table 2 case (i).

the added variable and Y. This brings one back to the situation where the two covariate set are equal.

It should be mentioned that the addition of a new variable to the Y-model also effects the OLS-estimate. If the added variable is correlated with R the selectivity bias may disappear. If the additional variable is correlated with Y the standard errors of estimates are reduced.

Comparing the two-step procedure and MLE, the latter turns out to be more stable and more efficient. But it is rather striking that - although the distributional assumptions are satisfied - MLE may produce rather misleading estimates.

If one uses the MSE-criterion to bring bias and variance into account one realizes that OLS-estimation performs quite well in relation to both the two-step procedure and the ML- estimate. The superiority of OLS over MLE was not stated before, while the bad performance of the two step procedure with respect to MSE was mentioned earlier by Stolzenberg/Relles (1990) and Zuehlke/Zeman (1991). Thus - as far as estimation is concerned - risk averse researchers would be better off with simple and easy to compute OLS-estimation.

References

[1] Amemiya, T. (1984): *Tobit Models. A Survey.* Journal of Econometrics, 24, 3-62.

[2] David, A. (1979a): *Conditional Independence in Statistical Theory.* J. Roy. Statist. Soc., B, 41, 1-31.

[3] David, A. (1979b): *Some misleading Arguments involving Conditional Independence.* J. Roy. Statist. Soc., B, 41, 249-252.

[4] Heckman, J. (1979): *Sample Selection Bias as a Specification Error.* Econometrica, 45, 153-161.

[5] Lee, L.-F. (1982): *Some Approaches to the Correction of Selectivity Bias.* Review of Economic Studies, 49, 352 -372.

[6] Lee, L.-F., Maddala, G. and Trost, R. (1980): *Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Models for Simultaneous Equations Models with Selectivity.* Econometrica, 48, 977-996.

[7] Lillard, L., Smith, J.P., and Welch, F. (1982): *What do we really know about wages: The importance of nonreporting and census imputation.* The Rand Corporation, Santa Monica, CA.

[8] Lillard, L., Smith, J.P., and Welch, F. (1986): *What do we really know about wages? The importance of nonreporting and Census imputation.* Journal of Political Economy, 94, 489-506.

[9] Little, R. (1985): *A Note About Models For Selectivity Bias.* Econometrica, 53, 1469-1474.

[10] Little, R. and Rubin, D. (1987): *Statistical Analysis With Missing Data.* Wiley, New York.

[11] Manski, C.F. (1989): *Anatomy of the Selection Problem.* Journal of Human Resources, 24, 343-360.

[12] Nelson, F. (1984): *Efficiency of the Two-Step Estimator for Models with Endogenous Sample Selection.* Journal of Econometrics, 24, 181-196.

[13] Olsen,R.J. (1980): *A Least Squares Correction for Selectivity Bias.* Econometrica, 48, 1815-1820.

[14] Rubin, D. (1976): *Inference and Missing Data.* Biometrica, 63, 581-592.

[15] Stolzenberg, R. and Relles, D. (1990): *Theory Testing in a World of Constrained Research Design. The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research.* Sociological Methods and Research, 18, 395-415.

[16] Verbeck, M. and Nijman, T. (1990): *Testing for Selectivity Bias in Panel Data Models.* CentER Discussion Paper 9018, Tilburg University, forthcoming in International Economic Review.

[17] Zuehlke, T. and Zeman, A. (1991): *A Comparison of Two-Stage Estimators of Censored Regression Models.* The Review of Economics and Statistics, 73, 185-187.