THE HECKMAN CORRECTION FOR SAMPLE SELECTION AND ITS CRITIQUE

Patrick A. Puhani

*SIAW, University of St. Gallen**

Abstract. This paper gives a short overview of Monte Carlo studies on the usefulness of Heckman's (1976, 1979) two-step estimator for estimating selection models. Such models occur frequently in empirical work, especially in microeconometrics when estimating wage equations or consumer expenditures.

It is shown that exploratory work to check for collinearity problems is strongly recommended before deciding on which estimator to apply. In the absence of collinearity problems, the full-information maximum likelihood estimator is preferable to the limited-information two-step method of Heckman, although the latter also gives reasonable results. If, however, collinearity problems prevail, subsample OLS (or the Two-Part Model) is the most robust amongst the simple-to-calculate estimators.

Keywords. Estimator performance; Sample selection model; Two-Part model; OLS

1. Introduction

Selection problems occur in a wide range of applications in econometrics. Prominent examples are the estimation of wage equations and consumer expenditures. For example, when trying to estimate the returns to schooling on the wage rate, one has the problem that some individuals who have received schooling do not work. Those people will have an offered wage below their reservation wage. If schooling has a positive influence on wages, people with little schooling will on average have a lower offered wage and therefore a lower employment rate than those with many years of schooling. As a consequence, one will only observe the wages of those people with few years of schooling who receive comparatively high wage offers. This is the problem of sample selection. The result of this problem is that simple OLS regression of wages on years of schooling will lead to downward-biased estimates, because the sample (working people) is unrepresentative of the population one is interested in (all people who have received schooling). Heckman (1976, 1979) has proposed a simple practical solution for such situations, which treats the selection problem as an omitted variable problem. This easy-to-implement method, which is known as the two-

*At the time of writing, the author was employed at ZEW, Mannheim and a research affiliate at SELAPO, University of Munich

0950-0804/00/01 0053–16 JOURNAL OF ECONOMIC SURVEYS Vol. 14, No. 1

step or the limited information maximum likelihood (LIML) method and has become the standard estimation procedure for empirical wage equations, has been criticised recently. The debate around the Heckman procedure is the topic of this short survey. The survey makes no claim on completeness, and the author apologises for possible omission of some contributions.

The paper is structured as follows. Section 2 outlines Heckman's LIML as well as the FIML estimator. Section 3 summarises the main points of criticism, whereas Section 4 reviews Monte Carlo studies. Section 5 concludes.

2. Heckman's proposal

Suppose we want to estimate the outcome (wage) equation [1a] of the following model:

$$y_{1i}^* = \mathbf{x}_{1i}'\beta_1 + u_{1i} \tag{1a}$$

$$y_{2i}^* = \mathbf{x}_{2i}'\beta_2 + u_{2i} \tag{1b}$$

$$\begin{aligned} y_{1i} &= y_{1i}^* \qquad \text{if } y_{2i}^* > 0 \\ y_{1i} &= 0 \qquad\quad \text{if } y_{2i}^* \leqslant 0. \end{aligned} \tag{1c}$$

Model (1b) could be a probit-type selection equation that describes the propensity to work or to have an observed wage. In principle, the variables y_1^* and y_2^* are unobserved, whereas y_1 is observed. One of the $\mathbf{x}_1$-variables may be years of schooling. As economists we will be interested in the wage difference an extra year of schooling pays in the labour market. Yet we will not observe a wage for people who do not work. This is expressed in (1b) and (lc).

Economic theory suggests that exactly those people who are only able to achieve a comparatively low wage given their level of schooling will decide not to work, as for them, the probability that their offered wage is below their reservation wage is highest. In other words, u_1 and u_2 can be expected to be positively correlated. It is commonly assumed that u_1 and u_2 have a bivariate normal distribution:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim BN\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right] \tag{2}$$

Given this assumption, the likelihood function of model (1) can be written (Amemiya, 1985, p. 386):

$$L = \prod_{y_1 = 0} 1 - \Phi\left(\frac{\mathbf{x}_2'\beta_2}{\sigma_2}\right) \prod_{y_1 > 0} \Phi\left\{ \left(\mathbf{x}_2'\beta_2 + \frac{\sigma_{12}}{\sigma_1^2}(y_1 - \mathbf{x}_1'\beta_1) \right) \sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}} \right\}$$
$$\times \frac{1}{\sigma_1}\phi\left(\frac{(y_1 - \mathbf{x}_1'\beta_1)}{\sigma_1}\right) \tag{3}$$

As the maximisation of this likelihood (full-information maximum likelihood, FIML) took a lot of computing time until very recently, Heckman (1979) proposed to estimate likelihood [3] by way of a two-step method (limited-information maximum likelihood, LIML).

For the subsample with a positive y_1^* the the conditional expectation of y_1^* is given by:

$$E(y_{1i}^* \mid \mathbf{x}_{1i}, y_{2i}^* > 0) = \mathbf{x}_{1i}'\beta_1 + E(u_{1i} \mid u_{2i} > -\mathbf{x}_{2i}'\beta_2). \tag{4}$$

It can be shown that, given assumption (2), the conditional expectation of the error term is:

$$E(u_{1i} \mid u_{2i} > -\mathbf{x}_{2i}'\beta_2) = \frac{\sigma_{12}}{\sigma_2} \frac{\phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}{1 - \Phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}, \tag{5}$$

where $\phi(.)$ and $\Phi(.)$ denote the density and cumulative density functions of the standard normal distribution, respectively. Hence we can rewrite the conditional expectation of y_1^* as

$$E(y_{1i}^* \mid \mathbf{x}_{1i}', y_{2i}^* > 0) = \mathbf{x}_{1i}\beta_1 + \frac{\sigma_{12}}{\sigma_2} \frac{\phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}{1 - \Phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}. \tag{6}$$

Heckman's (1979) two-step proposal is to estimate the so-called inverse Mills ratio

$$\lambda(\mathbf{x}_{2i}'\beta_2/\sigma_2) = \frac{\phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}{1 - \Phi(-(\mathbf{x}_{2i}'\beta_2/\sigma_2))}$$

by way of a Probit model and then estimate equation (7):

$$y_{1i} = \mathbf{x}_{1i}'\beta_1 + \frac{\sigma_{12}}{\sigma_2}\lambda(\mathbf{x}_{2i}'\beta_2\hat{/}\sigma_2) + \varepsilon_1 \tag{7}$$

in the second step. Hence, Heckman (1979) characterised the sample selection problem as a special case of the omitted variable problem with λ being the omitted variable if OLS were used on the subsample for which $y_1^* > 0$. As long as u_2 has a normal distribution and ε_1 is independent of λ, Heckman's two step estimator is consistent.[1] However, it is not efficient as ε_1 is heteroscedastic. To see this, note that the variance of ε_1 is given by

$$V(\varepsilon_{1i}) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\left[\frac{\mathbf{x}_{2i}'\beta}{\sigma_2}\lambda\left(\frac{\mathbf{x}_{2i}'\beta}{\sigma_2}\right) + \lambda\left(\frac{\mathbf{x}_{2i}'\beta}{\sigma_2}\right)^2\right]. \tag{8}$$

Clearly, $V(\varepsilon_{1i})$ is not constant, but varies over i, as it varies with $\mathbf{x}_{2i}$. In order to obtain a simple and consistent estimator of the asymptotic variance-covariance matrix, Lee (1982, p. 364f.) suggests to use White's (1980) method. Under the null hypothesis of no selectivity bias, Heckman (1979, p. 158) proposes to test for

selectivity bias by way of a t-test on the coefficient on λ. Melino (1982) shows that the t-statistic is the Lagrange multiplier statistic and therefore has the corresponding optimality properties.[2]

Although there are other LIML estimators of model (l) than the one proposed by Heckman (e.g. Olsen, 1980; Lee, 1983), this paper will use LIML to denote Heckman's estimator unless stated otherwise. Heckman (1979) considered his estimator to be useful for 'provid(ing) good starting values for maximum likelihood estimation'. Further, he stated that '(g)iven its simplicity and flexibility, the procedure outlined in this paper is recommended for exploratory empirical work.' (Heckman, 1979, p. 160). However, Heckman's estimator has become a standard way to obtain final estimation results for models of type (1). As we will see in the following section, this habit has been strongly criticised for various reasons.

3. The critique of Heckman's estimator

Although LIML has the desirable large-sample property of consistency, various papers have investigated and criticised its small-sample properties. The most important points of criticism can be summarised as follows:

1) It has been claimed that the predictive power of subsample OLS or the Two-Part Model (TPM) is at least as good as the one of the LIML or FIML estimators. The debate of sample-selection versus two-part (or multi-part) models was sparked off by Duan *et al.* (1983, 1984, 1985). The Two-Part Model (see also Goldberger, 1964, pp. 251ff.; and Cragg, 1971, p. 832) is given by

$$y_{1i}^* \mid y_{2i}^* > 0 = \mathbf{x}_{1i}'\beta_1 + u_{1i} \tag{9a}$$

$$y_{2i}^* = \mathbf{x}_{2i}'\beta_2 + u_{2i}, \text{ and} \tag{9b}$$

$$\begin{aligned} y_{1i} = y_{1i}^* \quad &\text{if} \quad y_{2i}^* > 0, \text{ and} \\ y_{1i} = 0 \quad &\text{if} \quad y_{2i}^* \leqslant 0. \end{aligned} \tag{9c}$$

The point is that (9a) models y_1^* *conditional* on y_1^* being positive. The expected value of y_1^* is then

$$E(y_{1i}^*) = \Phi(\mathbf{x}_{2i}'\beta_2/\sigma_2) \times (\mathbf{x}_{1i}'\beta_1) \tag{10}$$

Marginal effects on the expected value of y_1^* of a change in an $\mathbf{x}_1$-variable would thus have to be calculated by differentiating (10) with respect to the variable of interest.

There are three main ways in the literature to interpret the TPM. The first is to claim that it is not the unconditional, but rather the conditional expectation of y_1^* that is of interest to us. This approach is taken by Duan *et al.* (1983, 1984, 1985). The other approach is to stress the behavioural structure of the model (Maddala, 1985a, 1985b), to which the selection process is central. In this case, LIML and TPM estimate the same behavioural relation. The TPM, however, then makes an implicit distributional assumption for the unconditional distribution, which will

be a mixing distribution also depending on the distribution driving the selection mechanism. This approach, however, seems unsatisfactory from a theoretical point of view (see Hay and Olsen, 1984; and Maddala, 1985a, p. 14). A third and very crude interpretation, which is not explicitly stated in the literature on the TPM, is to interpret the coefficients of (9a) also as the ones of the unconditional equation (la). This is tantamount to estimating (la) by subsample OLS. As we will see below, the justification for the latter two interpretations will be given on statistical rather than theoretical grounds.

2) In practical problems, $\mathbf{x}_1$ and $\mathbf{x}_2$ often have a large set of variables in common. In some cases, they are even identical. One says that there are no exclusion restrictions if no variables that are in $\mathbf{x}_2$ are excluded from $\mathbf{x}_1$. In these cases, equation (7) is only identified through the nonlinearity of the inverse Mills ratio λ. However, collinearity problems are likely to prevail as $\lambda(.)$ is an approximately linear function over a wide range of its argument. This is illustrated in Figure 1.

Note that the probability to work for a person with characteristics $\mathbf{x}_2$ is given by $\Phi(\mathbf{x}_2\beta_2/\sigma_2)$. Only if this probability is higher than about 97.5% will $\mathbf{x}_2'\beta_2/\sigma_2$ be higher than 2. If most cases in a particular sample are not such extreme examples, most observations will lie within the quasi-linear range of the inverse Mills ratio, as demonstrated in Figure 1. It follows that regression (7) is likely to yield rather unrobust results due to collinearity problems. Therefore, Little and Rubin (1987, p. 230) state that 'for the (Heckman) method to work in practice, variables are needed in ($\mathbf{x}_2$) that are good predictors of (y_2^*) and do not appear in ($\mathbf{x}_1$), that is, are not associated with (y_1) when other covariates are controlled'. Unfortunately,

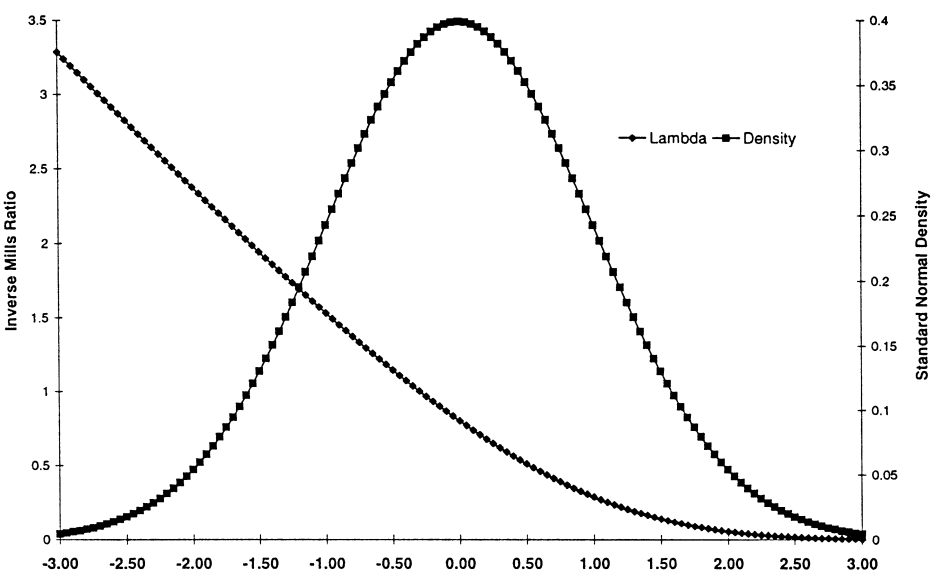

Figure 1. The quasi-linearity of the inverse mills ratio.

it is often very difficult to find such variables in practice. In our wage example, theory would suggest that household variables like children and the income of the spouse are likely to influence the reservation wage, but unlikely to influence the gross offered wage and hence should only be included in $\mathbf{x}_2$. However, these household data are not always available, and even if they are, it is not guaranteed that these variables are good predictors of the propensity to work y_2^*. But even if they are, the household variables may well be also associated with the offered wage y_1^*, especially if the after-tax wage is being observed, as children and the income of other family members have an impact on the tax rate in many tax regimes.

3) Another line of criticism stresses the sensitivity of the estimated coefficients with respect to the distributional assumptions placed on the error terms in (la) and especially in (lb) (Little and Rubin, 1987, pp. 225ff.). Instead of making strong distributional assumptions, some authors suggest semi-parametric or non-parametric procedures (see, for example, Chamberlain, 1986; Duncan, 1986; Powell, 1986; Robinson, 1988; Newey, Powell, and Walker, 1990; Cosslett, 1991; Ichimura and Lee, 1991; Ahn and Powell, 1993; Lee, 1996; Stern, 1996). These studies will not be surveyed here.

In the following, we give a summary of important Monte Carlo studies on the performance of the LIML estimator.

4. Monte Carlo studies

The existing Monte Carlo studies are summarised in Table 1. The table displays the key design features of the Monte Carlo experiments as well as the main conclusions drawn.

All of the analyses compare the LIML (or FIML) estimator with either subsample OLS or the Two-Part Model (TPM). The relative performance of the estimators is studied in relation to the joint distributions of the error terms u_1 and u_2, the correlations between the error terms, the degree of censoring, and the degree of collinearity between the regressors $\mathbf{x}_1$ and $\mathbf{x}_2$ or between $\mathbf{x}_1$ and the inverse Mills ratio λ.

As to the joint distribution of u_1 and u_2, no clear result emerges when the distributional assumption of joint normality is violated. For the extreme case with Cauchy errors, Paarsch (1984) and Hay, Leu, and Rohrer (1987) do not identify an estimator which behaves superior to the others. Also Zuehlke and Zeman (1991), who model bivariate t_5 and χ_5^2 errors, do not reach firm conclusions on this issue.

The correlation between the error terms, $\text{corr}(u_1, u_2)$, seems to have an effect on the performance of the LIML estimator. Although, Hay, Leu, and Rohrer (1987), Manning, Duan, and Rogers (1987), and Zuehlke and Zeman (1991) do not reach any strong results, Nelson (1984), Stolzenberg and Relles (1990), and Nawata (1993; 1994) provide evidence that the higher the correlation between u_1 and u_2, the greater the superiority of the FIML (and maybe OLS) estimator over LIML in terms of efficiency.

Table 1. Summary of Monte Carlo Studies

Study	Models analysed	Estimators used	Sample size	Repetitions	Distributions of u_1, u_2	Variables changed	Judgement criteria for estimators	Main results
Nelson (1984)	sample selection model with and without exclusion restrictions	LIML FIML OLS	2,000	n.a.	biv. normal	$R^2(\lambda, \mathbf{x}_1) = 0, 0.35, 0.641, 0.953, 0.999$ $\text{Corr}(u_1, u_2) = -0.5, 0, 0.25, 0.5, 0.75, 0.95$	bias and variance of parameter estimates	relative efficiency of FIML over LIML rises with higher $R^2(\lambda, \mathbf{x}_1)$ and $\text{corr}(u_1, u_2)$ OLS dominates LIML only when $R^2(\lambda, \mathbf{x}_1)$ is very high and/or $\text{corr}(u_1, u_2)$ is small
Paarsch (1984)	sample selection model without exclusion restrictions and identical errors (Tobit)	LIML FIML (Tobit) OLS Powell's LAD*	50 100 200	100	normal Laplace Cauchy	degree of censoring 25 and 50%	bias, variance, median, lower and upper quartile of parameter estimates	LIML much less efficient than FIMIL (Tobit) when errors are normal (or Laplace), FIMIL (Tobit) performs poorly when errors are Cauchy OLS worst estimator in all cases use of Powell's LAD limited
Hay, Leu, and Rohrer (1987)	sample selection model without exclusion restrictions	LIML FIML TPM	300 1,500 3,000		biv. Normal logistic/ normal Cauchy/ Cauchy	$\text{corr}(u_1, u_2) = 0, 0.33, 0.66, 0.90, 1.00$	mean squared error of fit mean bias of fit mean squared error of parameter estimates	TPM most robust when error distributions are normal or logistic, in the Cauchy case, none of the models can establish a superiority over the others no firm results concerning $\text{corr}(u_1, u_2)$

(*continued*)

Table 1. *Continued*

Study	Models analysed	Estimators used	Sample size	Repetitions	Distributions of u_1, u_2	Variables changed	Judgement criteria for estimators	Main results
Manning, Duan and Rogers (1987)	sample selection model with and without exclusion restrictions	LIML FIML TPM Data-Analytic TPM*	1,000	100	biv. normal	$\text{corr}(u_1, u_2) = 0.5, 0.9$ degree of censoring 25, 50, 75%	mean squared error of fit mean bias of fit	LIML worst when no exclusion restrictions (Data-Analytic TPM and best then) FIML and LIML perform badly when censoring is high no firm results concerning $\text{corr}(u_1, u_2)$
Stolzenberg and Relles (1990)	sample selection model with exclusion restrictions	LIML OLS	500	100	normal/ normal	$\text{corr}(\mathbf{x}_1, \mathbf{x}_2)^2 = 0, 0.25, 0.5, 0.75$ $\text{Corr}(u_1, u_2)^2 = 0, 0.25, 0.5, 0.75$ $\text{Var}(u_1) = 1/9, 1, 9$ $\text{Var}(u_2) = 0.25, 1, 4$	bias and mean absolute error of parameter estimates	no clear relationship between the variances of u_1 and u_2 and the performance of the two estimators high $\text{corr}(\mathbf{x}_1, \mathbf{x}_2)^2$ and high $\text{corr}(u_1, u_2)^2$ render LIML superior to OLS in terms of bias, than in OLS in over a third of cases

Zuehlke and Zeman (1991)	sample selection model without exclusion restrictions	LIML OLS Lee's robust estimator*	100	1,000	biv. normal biv. t_5 biv. χ^2_5	$\text{corr}(u_1, u_2) = 0, 0.5, 1$ Degree of censoring 25, 50, 75%	bias and mean squared error of parameter estimates	LIML reduces bias, but has very large standard error compared to OLS due to the collinearity of $\mathbf{x}_1$ and λ OLS preferable to LIML, especially when the degree of censoring is high Lee's robust estimator worst of all no firm results concerning $\text{corr}(u_1, u_2)$
Rendtel (1992)	sample selection model with and without exclusion restrictions	LIML FIML OLS	400	100	normal/ normal	additional variable in selection model (i) correlated with y_1 and y_2, (ii) correlated only with y_1, (iii) correlated only with y_2, (iv) correlated with neither y_1 nor y_2,	bias and variance of parameter estimates	without exclusion restrictions OLS is slightly preferable to FIML and clearly preferable to LIML with exclusion restrictions LIML and especially FIML dominate OLS only if the additional variable in the selection model is only correlated with y_2 (case (iii)); otherwise (cases (i), (ii), and (iv)) exclusion restrictions do not improve the FIML estimator

(*continued*)

Table 1. *Continued*

Study	Models analysed	Estimators used	Sample size	Repetitions	Distributions of u_1, u_2	Variables changed	Judgement criteria for estimators	Main results
Nawata (1993)	sample selection model with and without exclusion restrictions	LIML OLS	200	500	biv. normal	$\text{corr}(u_1, u_2) = 0, 0.2, 0.4, 0.6, 0.8, 1$ $\text{Corr}(\mathbf{x}_1, \mathbf{x}_2) = 0, 0.5, 0.8, 0.9, 0.95, 1$	bias, variance, median, lower and upper quartile of parameter estimates	LIML less efficient the higher ($\text{corr}(\mathbf{x}_1, \mathbf{x}_2)$ $\text{Corr}(u_1, u_2) > 0.9$ renders the LIML estimator very unstable OLS preferable for high $\text{corr}(\mathbf{x}_1, \mathbf{x}_2)$ and high $\text{corr}(u_1, u_2)$
Nawata (1994)	sample selection model with and without exclusion restrictions	LIML FIML	200	200	biv. normal	$\text{corr}(u_1, u_2) = 0, 0.4, 0.8$ $\text{corr}(\mathbf{x}_1, \mathbf{x}_2) = 0, 0.5, 0.8, 0.9, 0.95, 1$	bias, variance, median, lower and upper quartile of parameter estimates	FIML dominates LIML especially for high $\text{corr}(\mathbf{x}_1, \mathbf{x}_2) > 0.9$ renders the LIML estimator very unstable FIML generally preferable

Leung and Yu (1996)	sample selection model with and without exclusion restrictions TPM model	LIML FIML TPM Data-Analytic TPM*	1,000	100	biv. normal	degree of censoring 25, 50, 75% $\text{corr}(\mathbf{x}_1, \mathbf{x}_2) = 0, 0.5$	mean squared error of fit mean bias of fit bias and mean squared error of parameter estimates bias and mean squared error of elasticity estimates	degree of collinearity between $\mathbf{x}_1$ and λ crucial for performance of LIML and FIML as well as the t-test for sample selectivity superiority of FIML over LIML could not clearly be established if TPM is the true model, TPM dominates the other models, but the Data-Analytic TPM performs worse than LIML and FIML in terms of mean squared error of parameter estimates for high censoring

Notes: Powell's LAD and Lee's robust estimator are hardly used in empirical research and the reader is referred to the respective studies and the references cited therein for a description of these estimators. The Data-Analytic TPM is a TPM which also includes higher-order terms of $\mathbf{x}_1$ amongst the regressors.

The most important difference for the performance of the alternative estimators arises from the existence of exclusion restrictions, i.e. whether there are some variables in the selection equation which are not contained in the outcome equation. A very detailed investigation into this issue has been undertaken by Leung and Yu (1996). The authors point out that the degree of collinearity between the $\mathbf{x}_1$ regressors and the inverse Mills ratio λ is the decisive criterion to judge the appropriateness of the LIML and FIML estimators in relation to the TPM (or subsample OLS). Collinearity also limits the power of the t-test for sample selectivity on the coefficient of the inverse Mills ratio. The lack of exclusion restrictions is one likely reason for collinearity problems. However, a small range of the argument $\mathbf{x}_2'\beta_2$ of the inverse Mills ratio as well as high degrees of censoring (*cf.* Manning, Duan, and Rogers, 1987; Zuehlke and Zeman, 1991) may also cause collinearity.

Leung and Yu (1996) therefore suggest to test for collinearity by calculating the condition number for the regressors in (7) (a *LIMDEP 7.0* programme is given in the appendix to this paper). If the condition number exceeds 20, the TPM (or subsample OLS) is more robust, otherwise, FIML (or LIML) is recommended.[3]

For the empirical analysis of, say, wage equations, the standard procedure to solve collinearity problems would be to find appropriate exclusion restrictions. That is to say, one has to find variables that determine the probability to work (selection equation), but not the wage rate (outcome equation) directly. Practical examples for such variables could be the income of the spouse, household wealth, non-labour household income, or children (especially when estimating female wage rates). However, these variables are not always available in practical situations. Rendtel's (1992) study shows that implementing exclusions restrictions without testing whether the variables that are only included in the selection equation are not also directly impacting on the outcome variable can be very harmful to the performance of the LIML and FIML estimators. If the empirical researcher is not able to solve the collinearity problem, the advice to be drawn from the studies surveyed here would be to use standard OLS to estimate empirical wage equations.

5. Conclusions

This paper has given a short overview of the usefulness of Heckman's (1976, 1979) two-step estimator for estimating selection models. Such models occur frequently in empirical work, especially in microeconometrics when estimating wage equations or consumer expenditures.

The general conclusions which may be drawn from the surveyed Monte Carlo studies as well as the theoretical considerations cast doubt on the omnipotence implicitly ascribed by many applied researchers to Heckman's (1976, 1979) two-step estimator. Indeed, Heckman himself is confirmed when he writes that the purpose of his estimator is only to 'provide good starting values for maximum likelihood estimation' and 'exploratory empirical work.' (Heckman, 1979, p. 160).

The cases where the need to correct for selectivity bias are largest are those with a high correlation between the error terms of the selection and the outcome equation, and those with a high degree of censoring. Unfortunately, though, as the Monte Carlo analyses show, in exactly those cases Heckman's estimator is particularly inefficient and subsample OLS may therefore be more robust. In addition, empirical researchers are often confronted with a high correlation between the exogenous variables in the selection and the- outcome equation. Because the inverse Mills ratio is approximately linear over wide ranges of its argument, such high correlation is likely to make Heckman's LIML, but also the FIML estimator very unrobust due to the collinearity between the inverse Mills ratio and the other regressors.

The practical advice one may draw from these results, for example for the estimation of empirical wage equations, is that the estimation method should be decided upon case by case. A first step should be to investigate whether there are collinearity problems in the data. This can be done by calculating R^2 of the regression of the inverse Mills ratio on the regressors of the outcome equation or by calculating the corresponding condition number (a short *LIMDEP 7.0* programme is given in the appendix). If collinearity problems are present, subsample OLS (or the Two-Part Model) may be the most robust and simple-to-calculate estimator. If there are no collinearity problems, Heckman's LIML estimator may be employed, but given the constant progress in computing power, the FIML estimator is recommended, as it is usually more efficient than the LIML estimator.

Acknowledgements

I thank Klaus F. Zimmermann, SELAPO, University of Munich, Viktor Steiner, ZEW, Mannheim, Michael Lechner, University of Mannheim, Colin J. Roberts, University of Edinburgh as well as an anonymous referee for helpful comments. All remaining errors are my own.

Notes

1. However, as Rendtel (1992) points out, the orthogonality of $\mathbf{x}_1$ to u_1 does not imply that $\mathbf{x}_1$ be orthogonal to ε_1.
2. Olsen (1982) proposes a residuals-based test to test for selectivity.
3. This differs from Belsley, Kuh, and Welsch's (1980, p. 105) suggestion of taking 30 as the critical value. Leung and Yu (1996, p. 224) believe that chosing 20 as the critical value gives fairly accurate results, as the standard error of the condition number is quite small relative to its mean.

References

Ahn, H. and Powell, J. L. (1993) Semiparametric Estimation of Censored Selection Models with a Non-Parametric Selection Mechanism. *Journal of Econometrics*, 458, 3–29.
Amemiya, T. (1985) *Advanced Econometrics*. Oxford: Basil Blackwell.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980) *Regression Diagnostics, Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
Chamberlain, G. (1986) Asymptotic Efficiency in Semi-Parametric Models with Censoring. *Journal of Econometrics*, 32, 89–218.
Cosslett, S. (1991) Semiparametric Estimation of a Regression Model with Sample Selectivity. In W. A. Barnett, J. Powell and G. Tauchen, G. (eds). *Nonparametric and Semiparametric Methods in Econometrics and Statistics* Cambridge University Press.
Cragg, J. G. (1971) Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, 39, 5, 829–844.
Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1983) A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics*, l, 2, 115–126.
Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1984) Choosing Between the Sample-Selection Model and the Multi-Part Model. *Journal of Business & Economic Statistics*, 2, 3, 283–289.
Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1984) Comments on Selectivity Bias. *Advances in Health Economics and Health Services Research*, 6, 19–24.
Duncan, G. M. (1986) A Semi-Parametric Censored Regression Estimator. *Journal of Econometrics*, 32, 5–34.
Goldberger, A. S. (1964) Econometric Theory. New York: John Wiley & Sons.
Hay, J. W., Leu, R., and Rohrer, P. (1987) Ordinary Least Squares and Sample-Selection Models of Health-Care Demand: Monte Carlo Comparison. *Journal of Business & Economic Statistics*, 5, 499–506.
Hay, J. W., Olsen, R. J. (1984) Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Medical Care. *Journal of Business & Economic Statistics*, 2, 3, 279–282.
Heckman, J. J. (1976) The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic Social Measurement*, 5, 4, 475–492.
Heckman, J. J. (1979) Sample Selection Bias as a Specification Error. *Econometrica*, 47, l, 53–161.
Ichimura, H., and Lee, L.-F. (1991) Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation. In W. A. Barnett, J. Powell, and G. Tauchen, G. (eds), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.
Lee, L.-F. (1982) Some Approaches to the Correction of Selectivity Bias. *Review of Economic Studies*, 49, 355–372.
Lee, L.-F. (1983) Generalized Econometric Models with Selectivity Bias. *Econometrica*, 51, 2, 507–512.
Lee, M.-J. (1996) Nonparametric Two-Stage Estimation of Simultaneous Equations with Limited Endogenous Regressors. *Econometric Theory*, 12, 305–330.
Leung, S., F., Yu, S. (1996) On the Choice Between Sample Selection and Two-Part Models. *Journal of Econometrics*, 72, 197–229.
Little, R. J. A. (1985) A Note About Models for Selectivity Bias. *Econometrica*, 53, 6, 1469–1474.
Little, R. J. A. and Rubin, D. B. (1987) Statistical Analysis with Missing Data. New York: John Wiley & Sons.
Maddala, G. S. (1985a) A Survey of the Literature on Selectivity Bias as it Pertains to Health Care Markets. *Advances in Health Economics and Health Services Research*, 6, 3–18.
Maddala, G. S. (1985b) Further Comments on Selectivity Bias. *Advances in Health Economics and Health Services Research*, 6, 25–26.
Manning, W. G., Duan, N., Rogers, W. H. (1987) Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models. *Journal of Econometrics*, 35, 59–82.

Melino, A. (1982) Testing for Sample Selection Bias. *Review of Economic Studies*, 49, 151–153.
Nawata, K. (1993) A Note on the Estimation of Models with Sample Selection Biases. *Economics Letters*, 42, 15–24.
Nawata, K. (1994) Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator. *Economics Letters*, 45, 33–40.
Nelson, F. D. (1984) Efficiency of the Two-Step Estimator for Models with Endogenous Sample Selection. *Journal of Econometrics*, 24, 181–196.
Newey, W. K., Powell, J. L., and Walker, J. R. (1990) Semiparametric Estimation of Selection Models: Some Empirical Results. *American Economic Review Papers and Proceedings*, 80, 324–328.
Olsen, R. J. (1980) A Least Squares Correction for Selectivity Bias. *Econometrica*, 48, 7, 1815–1820.
Olsen, R. J. (1982) Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator. *International Economic Review*, 23, 1, 223–240.
Paarsch, H. J. (1984) A Monte Carlo Comparison of Estimators for Censored Regression Models. *Journal of Econometrics*, 24, 197–213.
Powell, J. L. (1986) Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica*, 54, 6, 1435–1460.
Powell, J. L. (1992) Least Absolute Deviations Estimation for censored and truncated Regression Models, unpublished Ph.D. Dissertation, Stanford University, CA, U.S.A.
Rendtel, U. (1992) On the Choice of a Selection-Model when Estimating Regression Models with Selectivity, DIW-Discussion Paper, No.53, Berlin.
Robinson, P. M. (1988) Root-N-Consistent Semiparametric Regression. *Econometrica*, 56 931–954.
Stern, S. (1996) Semiparametric Estimates of the Supply and Demand Effects of Disability on Labor Force Participation. *Journal of Econometrics*, 71, 49–70.
Stolzenberg, R. M. and Relles, D. A. (1990) Theory Testing in a World of Constrained Research Design, The Significance of Heckmans' Censored Sampling Bias Correction for Nonexperimental Research. *Sociological Methods and Research*, 18, 4, 395–415.
White, H. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48, 4, 817–828.
Zuehlke, T. W. and Zeman, A. R. (1990) A Comparison of Two-Stage Estimators of Censored Regression Models. *The Review of Economics and Statistics*, 72, 185–188.

Appendix

A Limdep Programme Which Calculates a Condition Number

```
LOAD          ; FILE = c:\data\example.sav $
OPEN          ; OUTPUT = c:\out\cond#.out  $


?    List  of  variables  for  which  the
?  condition number is to be calculated
NAMELIST     ; X = var1, var2, var3 $


? Compute the normalised moment matrix
MATRIX       ; XX = X'X
             ; D = DIAG(XX); D = ISQR(D)
             ; XX = D * XX * D $


? Find the highest and lowest eigenvalues
MATRIX       ; E = ROOT(XX) $
CALCULAT     ; r = ROW(E) $
MATRIX       ; EH = PART(E,1,1) $
MATRIX        ; EL = PART(E, r ,r) $


?  Calculate and display the condition
? number
CALCULATE     ; Cond = EH/EL $
MATRIX        ; LIST ; Cond $
```

Note: I thank William Greene for help with the normalisation (conversation by electronic mail).