Rochester Center for Economic Research

Collinearity and Two-step Estimation of Sample Selection Models: Problems, Origins, and Remedies

Leung, Siu Fai and Shihti Yu

Working Paper No. 419
June 1996

University of
Rochester

Collinearity and Two-step Estimation of Sample Selection Models: Problems, Origins, and Remedies

Siu Fai Leung and Shihti Yu

Rochester Center for Economic Research
Working Paper No. 419

June 1996

Collinearity and Two-step Estimation of Sample Selection Models: Problems, Origins, and Remedies

Siu Fai Leung

University of Rochester

and

Shihti Yu

National Chung Cheng University

Abstract: This paper investigates the origins of the collinearity problems encountered in the two-step estimation method for sample selection models. The analysis reveals several critical misconceptions and deficiencies in the literature. Remedies to the collinearity problems are proposed and evaluated.

Keywords: sample selection model, Heckman's two-step method, collinearity problem, remedy, Monte Carlo experiment

J.E.L. Classification: C15, C24

May 1996

Correspondence To:
Siu Fai Leung
Department of Economics
University of Rochester
Rochester, NY 14627
U.S.A.

Phone: 716-275-5273
Fax: 716-256-2309
E-mail: SFLE@DB1.CC.ROCHESTER.EDU

We have benefited from discussions with Songnian Chen and Lung-Fei Lee. This research was supported by the National Institute of Alcohol Abuse and Alcoholism AA08393.

1. Introduction

The two-step method proposed by Heckman (1976, 1979) has extensively been employed to estimate a wide variety of sample selection models in economics and other related disciplines. In spite of, and perhaps because of, its impressive popularity, there is a steady stream of criticisms, coming from various economists and statisticians, against the two-step method (see, e.g., Little 1985, Little and Rubin 1987, Manning, Duan, and Rogers 1987). Issues such as identification and non-normality have been raised and discussed, some of which have stimulated subsequent developments in the semiparametric estimation of sample selection models.

A recent criticism which has attracted some attention is the problem of multicollinearity. Nelson (1984), Nawata (1993, 1994), and Leung and Yu (1996) utilize Monte Carlo experiments to analyze the sources as well as the seriousness of collinearity problems encountered in the two-step method. They demonstrate that the two-step estimation procedure contains a special feature that is prone to collinearity problems. In some of their simulations, the problems are so serious that the estimates become very unstable and unreliable because of large standard errors. The collinearity problems are not merely a theoretical possibility. There is some evidence that they have also infected empirical work (e.g., Bockstael et al. 1990, Leung and Yu 1996). Despite the emergence of these simulation and empirical evidence, several aspects concerning collinearity problems and the two-step method are not yet well understood. The objectives of this paper are to further examine the sources of the collinearity problems and evaluate the effectiveness of some proposed remedies or solutions to the problems.

Before elaborating the objectives of this paper, it is necessary to discuss the value of the two-step approach to the estimation of sample selection models. Given the limitations of the two-step method, it seems puzzling that practitioners continue to utilize it despite the fact that there is a better alternative, maximum likelihood, that is more efficient and not as vulnerable. In fact, many have recommended maximum likelihood over the two-step method, For example, Maddala (1983, pp. 222-223) states that

"In the case of the tobit model, because the likelihood function is well-behaved and the computation of the ML [maximum likelihood] estimates is easy, there is no need for the two-stage method. In more complicated models in which ML methods are computationally burdensome the two-stage methods are worthwhile."

Similarly, Davidson and MacKinnon (1993, p.545) suggest that

"Although the two-step method for dealing with sample selectivity is widely used, our recommendation would be to use regression ... only as a procedure for testing the null hypothesis that selectivity bias is not present. When that hypothesis is rejected, ML estimation ... should probably be used in preference to the two-step method, unless it is computationally prohibitive."

The popular view holds that ease of computation is the only advantage of the two-step method over maximum likelihood. As a result of the recent explosive advance in computer technology and the availability of accessible econometric program packages like LIMDEP, some allege that the edge of the two-step method has rapidly disappeared because there is little difference in computational burden between maximum likelihood and the two-step method for sample selection models. Furthermore, many believe that maximum likelihood is not susceptible to the kind of collinearity problems encountered in the two-step procedure (Nelson 1984, Nawata 1994).

We believe that the prevailing views are questionable. Maximum likelihood is not immune to collinearity problems. The simulation results in Nelson (1984) and Leung and Yu (1996) indicate that maximum likelihood also experiences some degree of collinearity problems (e.g. bigger standard errors of the parameter estimates) that are noticeable though not as serious as those of the two-step method. Furthermore, the two-step method is still valuable and important for at least three reasons.

First, maximum likelihood is still computationally costlier than the two-step method especially when the sample size is big and the number of parameters is large. Even if the two estimation methods have the same costs of computation, the loglikelihood of a typical sample selection model is not necessarily concave except in some simple models. Hence, there is no guarantee that the root of the likelihood equation is unique (Nawata 1994). In addition, the maximum likelihood estimates can be very sensitive to the choice of the starting values for the parameters. The two-step estimates have been shown to be dependable and effective starting values for maximum likelihood.

Second, the dominance in efficiency of maximum likelihood over the two-step method is an asymptotic result. There is some evidence that the two-step method can be more efficient than maximum likelihood in small samples especially for certain parameter values of the models (Nelson 1984, Nawata 1994, Leung and Yu 1996). In addition, the two-step estimates also tend to have a smaller parameter bias than the maximum likelihood estimates in small samples (Leung and Yu 1996).

Third, the two-step method is more robust than maximum likelihood. Stapleton and Young (1984) show that, if there are measurement errors in the dependent variable of the outcome equation, the maximum likelihood estimators will no longer be consistent because the likelihood function is misspecified. In contrast, the two-step estimators remain consistent because the measurement errors are absorbed into the disturbance term of the outcome equation. Besides consistency, Stapleton and Young (1984) discover from their numerical experiment another superiority of the two-step method. They find that the two-step estimators are in general substantially more efficient than the maximum likelihood estimators when the dependent variable is measured with error. Since measurement errors are not uncommon in economic and socio-demographic data, the merits of the two-step method go far beyond mere computational considerations.

Because of these favorable comparisons with maximum likelihood, the two-step method is still irreplaceable in the parametric estimation of sample selection models. It is therefore worthwhile to further improve the two-step method by studying whether and how the collinearity problems can be reduced or removed. Section 2 examines the origins of the collinearity problems and points out some inadequacies and misleading claims in the literature. Without correcting the deficiencies, the collinearity problems will never be fully understood and effective remedies or solutions cannot be constructed. Section 3 assesses the efficacy of a number of traditional remedies to alleviate the collinearity problems. Some new remedies are proposed and evaluated. Section 4 concludes the paper.

2. Origins of the Collinearity Problems in the Two-step Method

There are many different types of sample selection models in the literature. We first consider the model:

$$y_{1i} = w_i'\alpha + u_{1i} \qquad i = 1,2,\ldots,N \tag{1}$$

$$d_i = I(y_{1i} > 0) \tag{2}$$

$$y_{2i} = x_i'\beta + u_{2i} \qquad \text{if } d_i = 1 \tag{3}$$

where $w_i' = (1,w_{1i},\ldots,w_{J-1,i})$, $x_i' = (1,x_{1i},\ldots,x_{K-1,i})$, $\alpha = (\alpha_0,\alpha_1,\ldots,\alpha_{J-1})'$, $\beta = (\beta_0,\beta_1,\ldots,\beta_{K-1})'$, and I(A) denotes an indicator function which equals 1 if the event A is true, 0 otherwise. For each i, w_i, x_i, and d_i are observed. If $d_i = 1$, y_{2i} is also observed. The error terms u_{1i} and u_{2i} are assumed to be independent and identically distributed bivariate normal random variables, with $E(u_{1i}) = E(u_{2i}) = 0$, $var(u_{1i}) = 1$, $var(u_{2i}) = \sigma^2$, $cov(u_{1i},u_{2i}) = \sigma_{12}$. It is well known that

$$\begin{aligned} E(y_{2i}|d_i = 1) &= x_i'\beta + E(u_{2i}|d_i = 1) \\ &= x_i'\beta + \sigma_{12}\lambda(w_i'\alpha), \end{aligned} \tag{4}$$

where $\lambda(z) = \phi(z)/\Phi(z)$ is the inverse Mills' ratio, $\phi(.)$ and $\Phi(.)$ are the probability density function and the cumulative distribution function of the standard normal distribution, respectively. Equations (1) and (3) are typically known as the selection equation and the outcome equation, respectively.

Given these parametric and distributional assumptions, the most efficient estimation method is a full scale (full information) maximum likelihood. As an alternative, Heckman (1976, 1979) proposes a simpler two-step method to estimate the model. The first step estimates α from the probit model (1) and (2) by means of maximum likelihood. Under very mild conditions, the maximum likelihood estimate $\hat{\alpha}$ is unique by virtue of the strict concavity of the probit likelihood function (Amemiya 1985, pp. 270-274). The second step estimates β and σ_{12} by ordinary least squares on the regression

$$y_{2i} = x_i'\beta + \sigma_{12}\lambda(w_i'\hat{\alpha}) + v_i \qquad \text{for } d_i = 1, \tag{5}$$

where $\lambda(w_i'\hat{\alpha})$ is treated as a regressor and v_i is an error term.

The curve $\lambda(z)$ has the properties that $\lambda'(z) < 0$, $\lambda''(z) > 0$, $\lim_{z\to\infty} \lambda(z) = 0$, $\lim_{z\to-\infty} \lambda(z) = \infty$, $\lim_{z\to\infty} \lambda'(z) = 0$, and $\lim_{z\to-\infty} \lambda'(z) = -1$. In particular, $\lambda(z)$ is almost linear in z over a long range of z. Figure 1 graphs $\lambda(z)$ for $z \in [-10,10]$. Treating B as a kink, $\lambda(z)$ can be divided into two main parts, AB and BC. The part AB can be approximated very well by a linear function of z. For example, if we take 111 points of z evenly spaced on [-10,1], then the regression $\hat{\lambda}(z) = 0.6634 - 0.9252z$ will fit $\lambda(z)$ on [-10,1] with $R^2 = 0.9975$. Such an excellent fit indicates that AB is essentially a linear line.

Although the part BC looks flat in Figure 1, it is actually hardly linear at all. If one fits a straight line to $\lambda(z)$ on [1,10] using 91 evenly spaced points of z, the R^2 of the regression will only be 0.3182. The illusive flatness of BC is due to a scale problem. Most of the values of $\lambda(z)$ on [1,10] are so small that the graph cannot accurately reflect the highly nonlinear shape of BC.[1] However, if one makes the transformation $\lambda^*(z) = \log_e\lambda(z)$ and takes 91 evenly spaced points of z on [1,10], then the straight line $\hat{\lambda}^*(z) = 10.8011 - 5.5063z$ will fit $\lambda^*(z)$ very well ($R^2 = 0.9569$). As depicted in Figure 2, $\log_e\lambda(z)$ declines approximately linearly with z on [1,10]. The near linearity of $\log_e\lambda(z)$ signifies how nonlinear $\lambda(z)$ is on [1,10].[2]

The shape of $\lambda(z)$ and the presence of $x_i'\beta$ and $\sigma_{12}\lambda(w_i'\hat{\alpha})$ in (5) are responsible for the collinearity problems that are peculiar to the two-step method.[3] Referring to $\lambda(.)$ and the kink B in Figure 1, if

(C1) all (or almost all) the data points $(w_i'\hat{\alpha},\lambda(w_i'\hat{\alpha}))$ for $d_i = 1$ fall on the left side of the kink, and

(C2) x_i and $w_i'\hat{\alpha}$ are highly collinear,

---

[1] For example, $\lambda(2) = 5.52\text{x}10^{-2}$, $\lambda(4) = 1.34\text{x}10^{-4}$, $\lambda(6) = 6.08\text{x}10^{-9}$, $\lambda(8) = 5.05\text{x}10^{-15}$, and $\lambda(10) = 7.69\text{x}10^{-23}$. In contrast, $\lambda(-2) = 2.37$, $\lambda(-4) = 4.23$, $\lambda(-6) = 6.16$, $\lambda(-8) = 8.12$, and $\lambda(-10) = 10.0981$.

[2] We mistakenly treated $\lambda(z)$ as "essentially flat for $z \geq 3$" in our earlier work and failed to recognize the nonlinearity of $\lambda(z)$ on BC (Leung and Yu 1996, p.211). As will be shown below, the nonlinearity of $\lambda(z)$ on BC plays an important role in the analysis of the collinearity problems.

[3] We assume that the regressors in x_i are not collinear so that we can focus on the collinearity problems that arise from the special structure of the two-step method.

then x_i and $\lambda(w_i'\hat{\alpha})$ will be highly collinear, and collinearity problems will likely appear in (5).[4] When condition (C1) is satisfied, $\lambda(w_i'\hat{\alpha})$ will almost be linear in $w_i'\hat{\alpha}$, hence x_i and $\lambda(w_i'\hat{\alpha})$ will be highly collinear if condition (C2) is satisfied. Neither (C1) nor (C2) is a sufficient condition for collinearity problems. For example, even if x_i and $w_i'\hat{\alpha}$ are highly collinear, there may not be any collinearity problems in (5) because it depends upon where the values of $\lambda(w_i'\hat{\alpha})$ are scattered. If the data points $(w_i'\hat{\alpha},\lambda(w_i'\hat{\alpha}))$ for $d_i = 1$ are scattered on the right side of the kink or on both sides of the kink, then $\lambda(w_i'\hat{\alpha})$ will not be linear in $w_i'\hat{\alpha}$ for all $d_i = 1$, hence collinearity problems will not appear. This point has often been ignored in the literature. For instance, Nawata (1993, p.24) asserts that

> "Heckman's two-step estimator is a reasonably good estimator when the degree of multicollinearity of x_i and $w_i'\hat{\alpha}$ is low, but performs poorly when the degree of multicollinearity is high."

The problem with this assertion is that a high degree of collinearity between x_i and $w_i'\hat{\alpha}$ is not a sufficient cause of collinearity problems. The two-step estimator can perform <u>very well</u> even when there is a high degree of multicollinearity between x_i and $w_i'\hat{\alpha}$. The flaw arises because Nawata (1993) examines $\lambda(z)$ only for $z \in [-3,3]$ and ignores the shape of $\lambda(z)$ for $z > 3$. Hence, he recognizes only the linearity of $\lambda(z)$ to the left of the kink, and misses the highly nonlinear part of $\lambda(z)$ to the right of the kink. As a consequence, a high degree of collinearity between x_i and $w_i'\hat{\alpha}$ will give rise to the collinearity problems that Nawata (1993) observes in his simulations.[5] Thus, his analysis is incomplete and misleading because there are no a priori reasons for confining the range of z to [-3,3]. As y_{1i} is a latent unobserved variable, it is possible that $w_i'\alpha$ lies outside [-3,3] for a significant number of observations.

---

[4] As is well known, there is not yet a commonly recognized and adopted measure of collinearity. Even if there is a consensus on the measure of collinearity, there is still the question of what is the threshold value of the measure for serious collinearity problems. Based on our experience, we prefer to use the condition number and we believe that the threshold condition number for moderate collinearity problems is somewhere between 20 and 30, depending on the type of models (Belsley, Kuh, and Welsch 1980, Leung and Yu 1996).

[5] We made a similar criticism in our earlier work, but we did not prove our claim (Leung and Yu 1996). A convincing proof will be given below.

Similar deficiencies are very common in the literature. In those studies that mention the collinearity problems in the two-step method, it is typically expressed either explicitly or implicitly that x_i and $\lambda(w_i'\hat{\alpha})$ will <u>tend</u> to be highly collinear if x_i and $w_i'\hat{\alpha}$ are highly collinear.[6] While this statement is not necessarily false (as it does not claim that x_i and $\lambda(w_i'\hat{\alpha})$ will be highly collinear if x_i and $w_i'\hat{\alpha}$ are highly collinear), it has never been substantiated. In particular, none of the studies proceed further to investigate: (i) <u>why</u> there is a tendency for x_i and $\lambda(w_i'\hat{\alpha})$ to be highly collinear if x_i and $w_i'\hat{\alpha}$ are highly collinear, and (ii) <u>what</u> extra conditions are needed for x_i and $\lambda(w_i'\hat{\alpha})$ to be highly collinear. Without addressing these two questions, the existing studies in the literature have misled many to believe that a high collinearity between x_i and $w_i'\hat{\alpha}$ is a sufficient condition for x_i and $\lambda(w_i'\hat{\alpha})$ to be highly collinear. Our theory shows that conditions (C1) <u>and</u> (C2) are sufficient for x_i and $\lambda(w_i'\hat{\alpha})$ to be highly collinear. Neither (C1) nor (C2) alone is sufficient.[7]

To prove the validity of our theory, we employ Nawata's (1993) own Monte Carlo experiments to show that, even when x_i and $w_i'\hat{\alpha}$ are perfectly collinear, the two-step method does not necessarily suffer any collinearity problems. Instead, the two-step estimators can perform very well. The Monte Carlo results clearly substantiate our theory on the sufficiency of conditions (C1) and (C2).

Since Nawata's (1993, 1994) experimental design is adapted from Paarsch (1984), we will label it the Paarsch-Nawata design. The Paarsch-Nawata design of the experiments is as follows. Let

$$y_{1i} = \alpha_0 + \alpha_1 w_i + u_{1i} \qquad i = 1,2,\ldots,N \tag{8}$$

$$d_i = I(y_{1i} > 0) \tag{9}$$

$$y_{2i} = \beta_0 + \beta_1 x_i + u_{2i} \qquad \text{if } d_i = 1 \tag{10}$$

---

[6] See, e.g., Heckman (1976, p.483), Maddala (1983, p.252), Nelson (1984, p.195), Bockstael et al. (1990, p.43), and Schmertmann (1994, p.115).

[7] Are (C1) and (C2) also necessary conditions for the collinearity problems? The answer is no because it is possible to give counterexamples in which $w_i'\hat{\alpha}$ has so little variations that $\lambda(w_i'\hat{\alpha})$ becomes like a constant, causing a high collinearity between $\lambda(w_i'\hat{\alpha})$ and the intercept term in x_i.

where x_i and w_i are scalars. The error terms u_{1i} and u_{2i} are drawn randomly from a bivariate normal distribution with $E(u_{1i}) = E(u_{2i}) = 0$, $var(u_{1i}) = 1$, $var(u_{2i}) = 100$, and $\rho = 0, 0.2, 0.4, 0.6, 0.8, 1.0$, where ρ denotes the correlation coefficient of u_{1i} and u_{2i}. To make the clearest and strongest contrast between the prevailing theory and our theory, we assume that x_i and w_i are identical, i.e., $x_i = w_i$. According to the prevailing theory, the two-step method is expected to perform poorly in this case since x_i and w_i are perfectly correlated.

Following Paarsch (1984) and Nawata (1993), we set $\beta_0 = -10$ and $\beta_1 = 1$. The regressor x_i is drawn randomly from U(0,20), where U(a,b) denotes a uniform distribution with range [a,b]. There are three designs in our experiments, which differ only in the values of α_0 and α_1:

Design I : $\alpha_0 = -1$ and $\alpha_1 = 0.1$

Design II : $\alpha_0 = -1$ and $\alpha_1 = 1$

Design III: $\alpha_0 = -10$ and $\alpha_1 = 1$

For each experiment, N = 200 and the number of replications is 500. Design I is exactly identical to the one in Nawata (1993, p.17) in which the correlation between x_i and w_i is 1. We employ the same design so that we can replicate Nawata's (1993) findings and use them as a basis of comparison. Since β_0 and β_1 are the parameters of interest in this type of models, we will just report these estimates along with the two-step estimate of σ_{12}. As in Nawata (1993), we compare the two-step estimates with the ordinary least squares (OLS) estimates of equation (10) which are obtained from a simple OLS regression of y_{2i} on an intercept and x_i for the uncensored sample of observations ($d_i = 1$). Tables 1, 2, and 3 present the simulation results for Designs I, II, and III, respectively.

Table 1 essentially replicates Nawata's (1993) findings. All the numbers are very close to those in Nawata's (1993) Table 7; the minor discrepancies are due to differences in the random number generators and the seeds. Compared with the OLS estimates, the two-step estimates do perform poorly when x_i and w_i are perfectly correlated. The standard deviations, as well as the first and third quartiles, of the two-step estimates are all considerably larger than those of the OLS estimates. In some cases, the two-step estimate of β_1 is even negative (as can be seen from the negative first quartiles). These results indicate that the two-step method is less stable and less reliable than ordinary least squares. Despite this poor performance, the means of the two-step estimates of β_0 and β_1 are all close to the true values (-10 and 1) especially for β_1. Such small mean biases (mean bias = mean of the estimates - true value) are expected because the two-step estimates remain consistent even when there are severe collinearity problems. On the other hand, the mean biases of the OLS estimates are very large except in the case where $\rho = 0$. In many cases, the OLS estimate of β_0 even takes a positive value. The mean biases of the OLS estimates increase with ρ because the omitted term in the OLS regression, $\sigma_{12}\lambda(\alpha_0+\alpha_1 w_i)$, increases with ρ (as $\sigma_{12} = 10\rho$).

Table 2 gives a remarkably different picture. Design II differs from Design I only in the value of α_1, yet the two-step method performs very well. The standard errors of the two-step estimates of β_0 and β_1 are small and are very close to those of the OLS estimates. The first quartile, the median, and the third quartiles show that the two-step estimates are well behaved and there are no signs of collinearity problems. The means of the two-step estimates of β_0 and β_1 are very close to the true values and the mean biases are smaller than those in Table 1. The OLS estimates also improve too. Although there are persistent mean biases in the OLS estimates of β_0 and β_1, the magnitude of the biases is much smaller than that in Table 1.

Table 3 presents another different picture. Design III differs from Design II only in the value of α_0. The two-step method continues to produce excellent estimates for β_0 and β_1 with small mean biases and low standard errors. However, the OLS estimates deteriorate substantially. The mean biases of the OLS estimates of β_0 and β_1 are much larger than those in Table 2, and some of the mean biases are comparable to those in Table 1 especially for β_1. The only problem with the two-step method is that the mean estimate of σ_{12} is abnormal in two cases ($\rho = 0.2$ and 0.8) due to the presence of two hugh outliers.

Nevertheless, the medians of the estimates of σ_{12} in these two cases, 2.39 and 9.51, are still reasonably close to the true values of 2 and 8. Consistent with the findings in the literature, Tables 1 - 3 show that the two-step method provides better estimates for β_0 and β_1 than σ_{12}.

Our theory can offer a complete account for the simulation results in Tables 1 - 3. In all three designs, condition (C2) is satisfied. Since $w_i \sim U(0,20)$, $\alpha_0 + \alpha_1 w_i \sim U(\alpha_0, \alpha_0+20\alpha_1)$. Thus, the ranges of $\alpha_0 + \alpha_1 w_i$ are [-1, 1], [-1,19], and [-10,10] in Designs I, II, and III, respectively. The narrow range of $\alpha_0 + \alpha_1 w_i$ in Design I implies that most of the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ will likely appear to the left side of the kink of $\lambda(.)$ (Figure 1). Therefore, both conditions (C1) and (C2) are satisfied in Design I. On the other hand, the wider ranges of $\alpha_0 + \alpha_1 w_i$ in Designs II and III suggest that the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ will likely scatter on both sides of the kink, hence condition (C1) will not be satisfied. One can get an idea of the severity of the collinearity problems by calculating for each design the correlation coefficient between $\alpha_0 + \alpha_1 w_i$ and $\lambda(\alpha_0 + \alpha_1 w_i)$ for those observations in which $\alpha_0 + \alpha_1 w_i + u_{1i} > 0$. The correlation coefficients are -0.9977, -0.5032, and -0.5699 for Designs I, II, and III, respectively. Therefore, our theory predicts that, despite the perfect correlation between x_i and w_i, the two-step method will suffer collinearity problems in Design I but not in Designs II and III.

Tables 4 and 5 verify our theory. Table 4 reports the summary statistics of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ (the first-step estimates of the two-step method). The maximum and the minimum values in Table 4 show that the ranges of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ in Designs I, II, and III are approximately [-0.95,1], [-0.7,22], and [-1,12], respectively. These ranges confirm that the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ fall on only the left side of the kink in Design I, but scatter on both sides of the kink in Designs II and III.[8]

---

[8] The range of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ is different from that of $\alpha_0 + \alpha_1 w_i$ because, apart from the obvious reason that $\hat{\alpha}_j \neq \alpha_j$ (j = 0,1), the range of $\alpha_0 + \alpha_1 w_i$ is calculated from the entire sample (i = 1,2,...,N) whereas the range of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ is computed from the uncensored sample ($d_i = 1$). Since an observation will be censored if $\alpha_0 + \alpha_1 w_i \leq u_{1i}$, the uncensored sample will tend to contain observations with larger and positive values of $\alpha_0 + \alpha_1 w_i$. Thus, the lower end points of the ranges of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ are greater than the corresponding ones of $\alpha_0 + \alpha_1 w_i$, whereas the upper end points are roughly the same.

Table 5 contains the summary statistics of the condition number. The large condition numbers in all the cases in Design I confirm that there are serious collinearity problems in the two-step method. Even the minimum condition numbers in Design I are substantially higher than 30 (the threshold for collinearity problems suggested by Belsley, Kuh, and Welsch 1980). On the other hand, the condition numbers in Designs II and III are all lower than 20, which are in accord with the lack of collinearity problems in the experiments for these two designs. The wider range of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ in Design II implies that a larger number of observations will fall on the highly nonlinear region of $\lambda(.)$ on the right side of the kink, which explains why the condition numbers in Design II are not only much smaller than those in Design I but also considerably smaller than those in Design III. In sum, Tables 1 - 5 clearly support our claim that a high collinearity between x_i and $w_i'\hat{\alpha}$ does not necessarily impair the two-step method.

Our theory can also account for the variations in the mean biases of the OLS estimates in Tables 1 - 3. The key lies in the degree (probability) of censoring, which is given by $P(d_i = 0) = P(\alpha_0 + \alpha_1 w_i + u_{1i} \leq 0)$. Since $w_i \sim U(0,20)$ and $u_{1i} \sim N(0,1)$, $P(\alpha_0 + \alpha_1 w_i + u_{1i} \leq 0) = -\alpha_0/(20\alpha_1)$. Hence, the degree of censoring is 50 percent in Design I and 5 percent in Design II. Since only 5 percent of the observations are censored in Design II, there is little information loss in the sample. In contrast, there is much more information loss in Design I because 50 percent of the data are unobserved due to the censoring. Thus, the mean biases of the OLS estimates in Table 2 are smaller than those in Table 1. To see this in a slightly more rigorous way, notice that $E(u_{2i}) = P(d_i=1)E(u_{2i}|d_i=1) + P(d_i=0)E(u_{2i}|d_i=0) = 0$. If $P(d_i=0) \approx 0$, then $E(u_{2i}) \approx E(u_{2i}|d_i=1) \approx 0$, so the bias of the OLS estimates, which depends on the size of $E(u_{2i}|d_i=1)$, will likely be small. In other words, the bias of the OLS estimates will tend to increase with the degree of censoring. As for Design III, the degree of censoring is 50 percent. Hence, the mean biases of the OLS estimates in Table 3 are larger than those in Table 2. The same degree of censoring in Designs I and III explains why the mean biases of the OLS estimates in Tables 1 and 3 are comparable.

3. Remedies to Collinearity Problems

Given the potentiality and the severity of the collinearity problems in the two-step method, it is surprising to find that there have been very few systematic investigations into the remedies of the problems. Most studies simply recommend giving up the two-step method and using ordinary least squares regression or maximum likelihood as the alternative (Nelson 1984, Nawata 1994). Because of the merits of the two-step method (as outlined in the Introduction), it would be useful to study whether it is possible to cure the collinearity problems without renouncing the two-step method.

We first examine two remedies frequently mentioned in econometrics textbooks. The first one is ridge regression, which is a mechanical way to reduce the variations of OLS estimates (e.g. Greene 1993). In Monte Carlo results not reported here, we do find that a ridge regression in the second step of the two-step method can considerably stabilize the two-step estimates in the experiments in Design I. The stabilizing effect is expected as the ridge estimators are designed for this purpose. However, as is well known, the ridge estimators are biased and standard statistical inference on the estimators is invalid. Given the emphasis on unbiasedness, consistency, and hypothesis testing in applied economic work, the limitations of ridge regression have severely diminished its appeal.

The second remedy often mentioned is to get more data (e.g. Davidson and MacKinnon 1993). One way to get more data is to increase the sample size. As the number of observations in our simulations is quite small (N equals 200 only), we examine whether the collinearity problems will be less serious when the sample size increases. Table 6 presents how the condition number varies with the sample size in the experiments. For brevity, we only report the results for Design I with $\rho = 1$. There are two rows, "uncensored" (i.e. $d_i = 1$) and "all" (i.e. $i = 1,2,\ldots,N$), for each sample size N. Here we will just focus on the "uncensored" row, the "all" row will be explained and discussed later. Several results are prominent. First, the mean of the condition numbers drops very little, from 97.23 to 95.77, when the sample size increases from 500 to 10,000. The fall in the standard deviation (from 18.36 to 3.74), the skewness (from 0.92 to 0.24), as well as the kurtosis (from 4.61 to 3.09) is much more significant. Second, the narrowing difference between the maximum and the minimum values suggests that the condition number is converging to some value around 95 as N gets indefinitely large. Third, relative to the threshold 30 suggested by Belsley, Kuh, and Welsch (1980), the condition numbers are still very large even when the sample size is 10,000. Hence, getting more data in this case does not help much in alleviating the collinearity problems.

Next we evaluate several possible remedies which are specific to the two-step estimation method. Two remedies follow immediately from our theory. In order to alleviate the collinearity problems, conditions (C1) or (C2), or both, must be weakened. To neutralize (C1), the regressors w_i have to be chosen such that there is a significant number of observations for which the values of $w_i'\alpha$ for $d_i = 1$ fall on the right side of the kink. Because of the high nonlinearity of $\lambda(.)$ on the right side of the kink, these observations will appreciably lower the collinearity between x_i and $\lambda(w_i'\alpha)$. The problem with this remedy is that, given the nature of the non-experimental data in economics, there may be very little choice in the regressors w_i and the range of w_i.

To weaken (C2), the compositions of w_i and x_i have to be altered so as to reduce the overlap, and hence the degree of collinearity, between w_i and x_i. This remedy has often been mentioned in the literature, partly because it also eases the identification problems between x_i and $\lambda(w_i'\alpha)$ (e.g. Little and Rubin 1987, p. 230). While this strategy is widely adopted by practitioners, the main problem is that economic theory seldom predicts which variables should be included in w_i but not in x_i. Consequently, the exclusion restrictions on the regressors in w_i and x_i may be arbitrary.

A third remedy which we propose is to incorporate additional information. For instance, one can include the sample of censored observations in the two-step method if the data are available. Suppose, in addition to (1), (2), (3), we have

$$y_{2i} = 0 \qquad \text{if } d_i = 0, \tag{11}$$

and we observe w_i, x_i, d_i, y_{2i} for each $i = 1,2,\ldots,N$. There are many examples of this type of sample selection models and it has been known as the Type 2 Tobit model according to Amemiya's (1985) classification system. Wales and Woodland (1980) show that the additional information in (11) can be incorporated into the two-step method. Since $E(y_{2i}) = \Pr(d_i = 1)E(y_{2i}|d_i = 1) + \Pr(d_i = 0)E(y_{2i}|d_i = 0)$, therefore

$$E(y_{2i}) = \Phi(w_i'\alpha)[x_i'\beta + \sigma_{12}\lambda(w_i'\alpha)], \qquad i = 1,2,\ldots,N. \tag{12}$$

Wales and Woodland (1980) suggest a Heckman-type two-step method based on (12). The first step, which is the same as before, obtains maximum likelihood estimates $\hat{\alpha}$ from (1) and (2). The second step estimates β and σ_{12} from the linear regression

$$y_{2i} = \Phi(w_i'\hat{\alpha})[x_i'\beta + \sigma_{12}\lambda(w_i'\hat{\alpha})] + \epsilon_i, \quad i = 1,2,\ldots,N \tag{13}$$

where $\Phi(w_i'\hat{\alpha})x_i$ and $\Phi(w_i'\hat{\alpha})\lambda(w_i'\hat{\alpha})$ are treated as regressors, and ϵ_i is an error term. Since this two-step method incorporates more information than the one based on (5) and the sample size is also bigger, it may be a useful remedy for the collinearity problems in (5).[9]

To evaluate this suggestion we conduct some Monte Carlo experiments using a design similar to Design I except that the parameter values $\alpha_1 = 0.1$, $\rho = 1$, and $\alpha_0 = -0.2, -0.5, -1, -1.5$ are employed. We choose several values of α_0 in order to study whether the results are sensitive to the degree of censoring.[10] Table 7 reports the condition numbers of the regressions (5) and (13). The results are disappointing because they show that (13) does not help alleviate, but instead aggravate, the collinearity problems. In every case studied, the mean condition number of the regression using all the observations is larger than the one using the uncensored observations.

---

[9] As this two-step method makes use of more information from the data, intuition suggests that it should be more efficient than the one based on (5). However, Wales and Woodland (1980) and Amemiya (1985) show that this is not necessarily the case.

[10] The degree of censoring determines the sizes of the censored and the uncensored samples. For $\alpha_0 = -0.2, -0.5, -1, -1.5$, the degree of censoring $P(\alpha_0 + \alpha_1 w_i + u_{1i} \leq 0)$ is 10, 25, 50, and 75 percent, respectively.

To check that this poor performance is not caused by a small sample size, we employ Design I with $\rho = 1$ and calculate the condition number of (13) for larger sample sizes (N ranging from 500 to 10,000). Each "all" (for all observations) row in Table 6 presents the summary statistics of the condition number of (13) for each N. First, the results indicate that the condition number of (13) changes very little (from 110.17 to 108.53) as the sample size increases from 500 to 10,000. Second, for each N, the mean condition number of (13) is always greater than that of (5). Hence, the poor performance of (13) is not related to the degree of censoring and the sample size N.

Our theory can offer an explanation for this puzzle. Let us first rewrite (13) as

$$y_{2i} = \Phi(w_i'\hat{\alpha})\{[x_i'\beta + \sigma_{12}\lambda(w_i'\hat{\alpha})]d_i + [x_i'\beta + \sigma_{12}\lambda(w_i'\hat{\alpha})](1-d_i)\} + \epsilon_i, \quad i = 1,2,\ldots,N. \qquad (14)$$

Since $\Phi(w_i'\hat{\alpha})$ in (14) can be regarded as a scale factor, it has no bearing on the collinearity problems. Recall that the collinearity problems in design I arise because all the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$ fall on only the left side of the kink of $\lambda(.)$ in Figure 1. Thus, whether the two-step method based on (14) can reduce the collinearity problems hinges on the collinearity between x_i and $\lambda(\hat{\alpha}_0 + \hat{\alpha}_1 w_i)$ for $d_i = 0$. Since $u_{1i} \sim U(0,1)$ and $d_i = 0$ if $\alpha_0 + \alpha_1 w_i' \leq -u_{1i}$, the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 0$ will likely be small and negative. In other words, most of the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 0$ will also tend to fall on only the left side of the kink. As a consequence, the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for both $d_i = 0$ and $d_i = 1$ will scatter mostly on the left side of the kink, hence $\lambda(\hat{\alpha}_0 + \hat{\alpha}_1 w_i)$ will remain linear in $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$. This explains why the two-step method based on (13) suffers even more serious collinearity problems than the one based on (5).

Table 8 confirms that our explanation is correct. For each α_0, the mean of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ of the censored observations ($d_i = 0$) is significantly smaller than that of the uncensored observations ($d_i = 1$). Although the minimum and the maximum values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ of the censored sample are similar to those of the uncensored sample, the number of cases in which $\hat{\alpha}_0 + \hat{\alpha}_1 w_i > 0$ for $d_i = 0$ is notably smaller than that for $d_i = 1$ especially for smaller values of α_0. These results verify that the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$

of the censored observations tend to fall even further to the left of those of the uncensored observations.

The above discussion shows that, if (5) suffers collinearity problems, then (13) will definitely be even worse. On the other hand, it is possible for (13) to experience some collinearity problems even if there are no collinearity problems in (5). For instance, if the number of censored observations is considerably larger than that of the uncensored observations and if all the values of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ of the censored observations are scattered on the left side of the kink, then (15) may encounter some collinearity problems because the censored sample raises the collinearity between x_i and $\lambda(w_i'\hat{\alpha})$. For these reasons it is not advisable to apply the two-step method to (13).

Although the previous example shows that combining the censored and the uncensored observations is not fruitful in lessening the collinearity problems, there are examples in which the remedy works very well. For example, consider (1), (2), and

$$y_{2i} = x_i'\beta + \gamma d_i + u_{2i}, \qquad i = 1,2,\ldots,N \tag{15}$$

where w_i, x_i, d_i, and y_{2i} are observed for each $i = 1,2,\ldots,N$. The differences between (15) and (3) are that (15) is not restricted to $d_i = 1$ only (as in (3)) and that the dummy variable d_i also appears as a regressor in (15). This type of endogenous dummy variable model, which has been applied to study education, job training, and program evaluation, is frequently found in the sample selection literature (Maddala 1983, Greene 1993). Since

$$\begin{aligned} E(y_{2i}|d_i = 1) &= x_i'\beta + \gamma + E(u_{2i}|d_i = 1) \\ &= x_i'\beta + \gamma + \sigma_{12}\lambda(w_i'\alpha), \end{aligned} \tag{16}$$

the OLS regression in the second step of the two-step method is given by

$$y_{2i} = x_i'\beta + \gamma + \sigma_{12}\lambda(w_i'\hat{\alpha}) + \epsilon_i \qquad \text{for } d_i = 1, \tag{17}$$

where ϵ_i is an error term. The parameter γ in (17) will be automatically absorbed into the intercept term in x_i. Similarly,

$$E(y_{2i}|d_i = 0) = x_i'\beta + E(u_{2i}|d_i = 0)$$

$$= x_i'\beta - \sigma_{12}\lambda(-w_i'\alpha), \quad (18)$$

hence an alternative second step can be based on the regression

$$y_{2i} = x_i'\beta - \sigma_{12}\lambda(-w_i'\hat{\alpha}) + \epsilon_i \qquad \text{for } d_i = 0. \quad (19)$$

One critical shortcoming in both regressions (17) and (19) is that they do not produce an estimate for γ, which is a key parameter in this type of models. Barnow, Cain, and Goldberger (1981) suggest a third way to estimate the model. Combining (17) and (19), y_{2i} can be expressed as

$$y_{2i} = x_i'\beta + \gamma d_i + \sigma_{12}[d_i\lambda(w_i'\hat{\alpha}) - (1-d_i)\lambda(-w_i'\hat{\alpha})] + \epsilon_i, \quad i = 1,2,\ldots,N \quad (20)$$

which can also be treated as the regression in the second step of the two-step method.[11] One obvious advantage of (20) over (17) and (19) is that the regression produces an estimate for γ.[12] Our concern is whether (20) is less vulnerable to collinearity problems than (17) and (19).

Now suppose the worst scenario in which there are serious collinearity problems in both (17) and (19). All the values of $w_i'\hat{\alpha}$ for $d_i = 1$ and $-w_i'\hat{\alpha}$ for $d_i = 0$ fall on the left side of the kink of $\lambda(.)$, so that x_i is highly collinear not only with $\lambda(w_i'\hat{\alpha})$ for $d_i = 1$ but also with $\lambda(-w_i'\hat{\alpha})$ for $d_i = 0$. Despite this high collinearity, the combined regressor $d_i\lambda(w_i'\hat{\alpha}) - (1-d_i)\lambda(-w_i'\hat{\alpha})$ in (20) will not be highly collinear with x_i because $\lambda(w_i'\hat{\alpha})$ and $-\lambda(-w_i'\hat{\alpha})$ have opposite signs.[13] Therefore, combining the uncensored and the censored observations will help reduce the potentiality as well as the seriousness of collinearity problems in this type of sample selection models.

---

[11] Clearly, the error terms in (13), (17), (19), and (20) are all dissimilar, but we use the same symbol ϵ_i for notational simplicity. We will continue to adopt this convention for the rest of the paper.

[12] This sample selection model can be conveniently estimated by a program in LIMDEP 6.0 (Greene 1992, pp. 605-610). Users can choose (17), (19), or (20) to estimate the parameters.

[13] To see this, suppose that, for the purpose of illustration, $\lambda(z)$ is a linear function of z, i.e., $\lambda(z) = a_0 + a_1 z$, where a_0 is a positive constant and a_1 is a negative constant. Suppose there are two columns of data Z_1 (dimension N_1x1) and Z_2 (dimension N_2x1), then Z_1 will be perfectly correlated with $\lambda(Z_1)$, and Z_2 will be perfectly correlated with $\lambda(-Z_2)$. However, the column of data $(Z_1',Z_2')'$ is not perfectly collinear with the vector $(\lambda(Z_1)',-\lambda(-Z_2)')'$ because the latter is equal to $((a_0 e_1 + a_1 Z_1)',(-a_0 e_2 + a_1 Z_2)')'$, which cannot be expressed as $m_0(e_1',e_2')' + m_1(Z_1',Z_2')'$ for some constants m_0 and m_1, where e_1 is an N_1x1 unit vector and e_2 an N_2x1 unit vector.

Finally, we comment on the Type 1 Tobit model (Amemiya 1985):

$$y_i = (x_i'\beta + u_i)I(x_i'\beta + u_i > 0), \qquad i = 1,2,\ldots,N \tag{21}$$

where x_i and y_i are observed for all i. The Type 1 Tobit model is a special case of the Type 2 Tobit model because it can be obtained from (1), (2), and (11) by imposing the restrictions $w_i = x_i$, $\alpha = \beta$, and $u_{1i} = u_{2i}$. The Type 1 Tobit is an example of a sample selection model in which there are completely no exclusion restrictions on w_i and x_i (hence $w_i = x_i$). On the other hand, there is a complete restriction on α and β ($\alpha = \beta$). Let $u_i \sim N(0,\sigma^2)$, then

$$E(y_i|d_i = 1) = x_i'\beta + \sigma\lambda(x_i'\beta/\sigma). \tag{22}$$

In this case, the first step of the two-step method for (21) is to obtain an estimate of β/σ from the probit model $d_i = I(x_i'\beta + u_i > 0)$. Denoting the maximum likelihood estimate of β/σ by $\hat{\delta}$, the second step estimates β and σ from the ordinary least squares regression

$$y_i = x_i'\beta + \sigma\lambda(x_i'\hat{\delta}) + \epsilon_i \qquad \text{for } d_i = 1, \tag{23}$$

where $\lambda(x_i'\hat{\delta})$ is treated as a regressor. It is clear that (23) will likely suffer collinearity problems if most of the values of $x_i'\hat{\delta}$ fall on the left side of the kink of $\lambda(.)$. We propose to remedy the collinearity problems in the following way. Since (22) can be written as $E(y_i|d_i = 1) = \sigma[x_i'\beta/\sigma + \lambda(x_i'\beta/\sigma)]$, the regression

$$y_i = \sigma[x_i'\hat{\delta} + \lambda(x_i'\hat{\delta})] + \epsilon_i \qquad \text{for } d_i = 1 \tag{24}$$

can be used as the second step of the two-step method. This regression treats $x_i'\hat{\delta} + \lambda(x_i'\hat{\delta})$ as a single regressor and produces an OLS estimate for σ, which we denote by $\hat{\sigma}$. A consistent and asymptotically normal estimate of β can then be obtained by multiplying $\hat{\sigma}$ by $\hat{\delta}$ (since $\beta = \sigma(\beta/\sigma)$). Even if there is perfect collinearity between $x_i'\hat{\delta}$ and $\lambda(x_i'\hat{\delta})$, one can still estimate σ from (24) very well. As there is only one regressor, $x_i'\hat{\delta} + \lambda(x_i'\hat{\delta})$, in (24), there will never be any collinearity problems in (24). Unfortunately, this remedy cannot be extended to other types of sample selection models because it hinges crucially on the restrictions $w_i = x_i$, $\alpha = \beta$, and $u_{1i} = u_{2i}$.

4. Conclusion

This paper examines the origins of, and the remedies to, the collinearity problems encountered in the two-step estimation method for sample selection models. We show that the linearity of the inverse Mills' ratio $\lambda(z)$ over a certain range of z and the collinearity between the regressors of the outcome equation and the selection equation can give rise to serious collinearity problems in the second step of the two-step method. A major new finding in the paper is that the high nonlinearity of $\lambda(z)$ over another range of z can substantially reduce the potentiality as well as the severity of the collinearity problems.

In practice, whether the two-step method is vulnerable to collinearity problems depends largely on the range of $w_i'\hat{\alpha}$ in empirical work. All those studies that attack the two-step method for the collinearity problems assume either explicitly or implicitly that $w_i'\hat{\alpha}$ lies in a certain range. These studies, however, have never provided any justifications for the assumption of a specific range for $w_i'\hat{\alpha}$. Thus, the criticisms are unbalanced and misleading because there are no a priori reasons for restricting $w_i'\hat{\alpha}$ to a specific range. As shown in equation (1), $w_i'\alpha$ is the systematic part of the latent unobserved variable y_{1i} which is constructed to explain the selection outcome. The range of $w_i'\hat{\alpha}$ depends on, among others, the distribution of w_i, the sign as well as the magnitude of the elements in w_i and $\hat{\alpha}$. Hence, $w_i'\hat{\alpha}$ can take any value on the real line. The higher the incidence of large positive values of $w_i'\hat{\alpha}$ (by "large" we mean $w_i'\hat{\alpha}$ greater than 3 only), the less likely will there be collinearity problems in the two-step method because of the powerful nonlinearity of $\lambda(z)$ for larger values of z.

In addition to reviewing the effectiveness of a number of oft suggested cures to collinearity problems, we also propose some remedies that are specific to the two-step method. By exploiting the structures of the sample selection models and the distinctive features of the two-step method, we show that some of our remedies can help reduce the collinearity problems. Unfortunately, the remedies still lack generality as they are applicable to certain types of sample selection models only. The limited success of the remedies indicates that further investigations into the problems are warranted.

Figure 1
The Inverse Mills' Ratio
$\lambda(z) = \phi(z)/\Phi(z)$

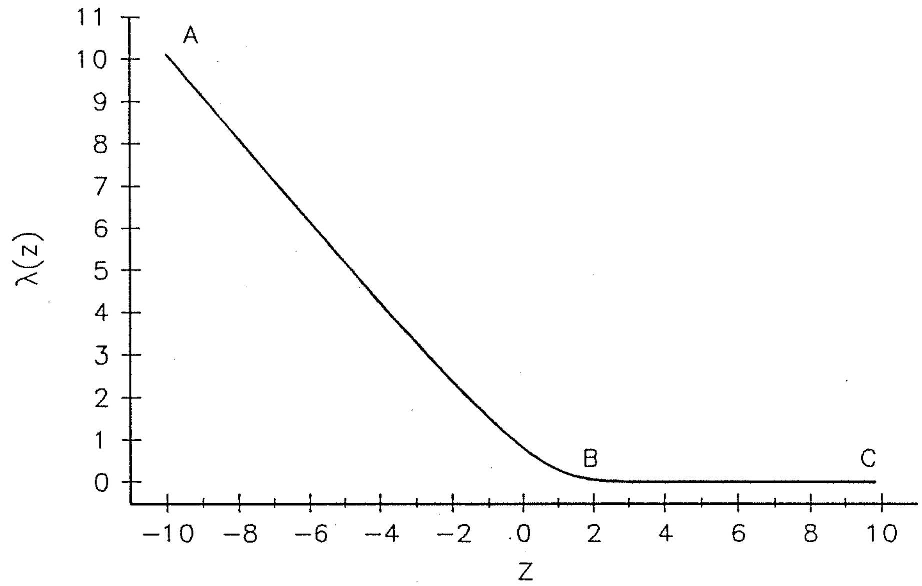

Figure 2
$\lambda(z)$ and $\log_e \lambda(z)$

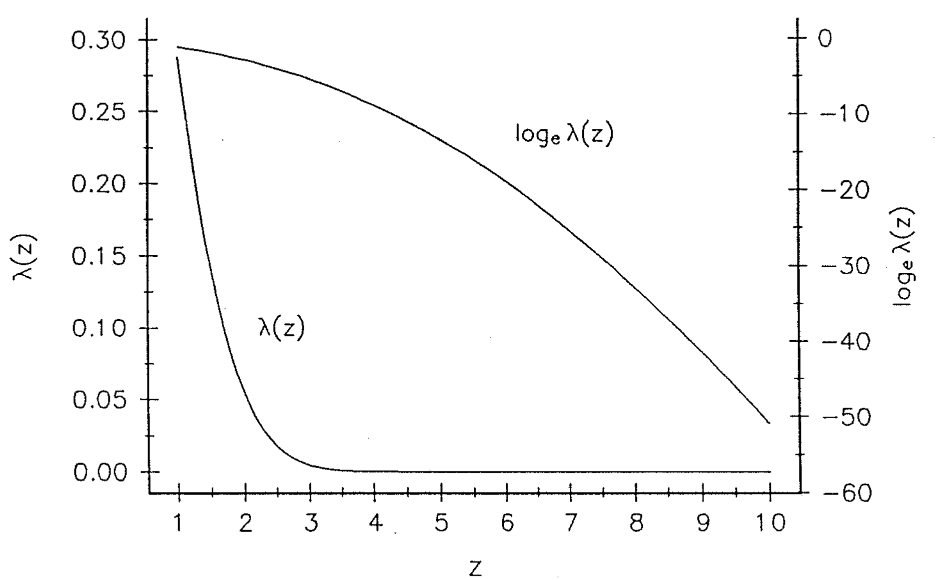

Table 1
Simulation Results for Design I
$\beta_0 = -10$, $\beta_1 = 1$, $\sigma_{12} = 10\rho$

ρ	Estimation Method	Parameter	Mean	Standard Deviation	First Quartile	Median	Third Quartile
0	Heckman	β_0	-11.72	52.35	-46.06	-13.94	17.66
		β_1	1.07	2.11	-0.22	1.18	2.46
		σ_{12}	1.42	38.01	-20.10	1.94	24.38
	OLS	β_0	-9.96	2.61	-11.73	-10.19	-8.17
		β_1	1.00	0.19	0.87	1.00	1.14
0.2	Heckman	β_0	-12.12	52.52	-43.65	-9.34	19.20
		β_1	1.07	2.12	-0.33	0.91	2.43
		σ_{12}	4.37	38.01	-17.05	3.02	25.19
	OLS	β_0	-6.61	2.66	-8.43	-6.66	-4.82
		β_1	0.86	0.20	0.73	0.86	1.00
0.4	Heckman	β_0	-13.44	51.12	-44.12	-13.57	15.06
		β_1	1.12	2.04	-0.10	1.17	2.39
		σ_{12}	8.35	37.28	-12.16	7.46	29.65
	OLS	β_0	-2.10	2.40	-3.79	-2.02	-0.62
		β_1	0.67	0.18	0.55	0.67	0.79
0.6	Heckman	β_0	-12.44	38.58	-34.20	-9.69	14.15
		β_1	1.08	1.58	-0.05	0.98	2.02
		σ_{12}	10.16	27.59	-7.74	7.83	26.06
	OLS	β_0	1.95	2.00	0.61	1.92	3.36
		β_1	0.49	0.15	0.39	0.50	0.59
0.8	Heckman	β_0	-12.53	36.06	-33.92	-12.67	9.46
		β_1	1.09	1.48	0.16	1.14	2.01
		σ_{12}	11.63	25.82	-3.34	10.48	25.93
	OLS	β_0	3.88	1.48	2.93	3.91	4.77
		β_1	0.41	0.12	0.33	0.41	0.49
1.0	Heckman	β_0	-11.41	33.48	-28.17	-11.38	8.53
		β_1	1.04	1.36	0.20	1.08	1.80
		σ_{12}	11.20	24.02	-2.57	10.89	22.13
	OLS	β_0	4.34	1.44	3.34	4.35	5.30
		β_1	0.39	0.12	0.32	0.40	0.47

Table 2
Simulation Results for Design II
$\beta_0 = -10$, $\beta_1 = 1$, $\sigma_{12} = 10\rho$

ρ	Estimation Method	Parameter	Mean	Standard Deviation	First Quartile	Median	Third Quartile
0	Heckman	β_0	-9.90	1.88	-11.00	-9.89	-8.72
		β_1	0.99	0.15	0.90	0.99	1.07
		σ_{12}	-0.06	5.31	-3.69	0.19	3.49
	OLS	β_0	-9.90	1.62	-10.95	-9.95	-8.79
		β_1	0.99	0.13	0.90	1.00	1.08
0.2	Heckman	β_0	-10.01	1.86	-11.22	-9.95	-8.80
		β_1	1.00	0.15	0.90	1.00	1.09
		σ_{12}	2.53	5.34	-0.84	2.45	6.06
	OLS	β_0	-9.55	1.63	-10.62	-9.52	-8.35
		β_1	0.97	0.14	0.87	0.97	1.05
0.4	Heckman	β_0	-9.84	1.83	-11.01	-9.74	-8.59
		β_1	0.99	0.15	0.89	0.98	1.08
		σ_{12}	5.33	4.83	2.12	4.92	8.61
	OLS	β_0	-8.86	1.56	-9.88	-8.80	-7.84
		β_1	0.92	0.13	0.82	0.92	1.00
0.6	Heckman	β_0	-9.98	1.64	-11.03	-9.99	-8.91
		β_1	1.00	0.13	0.90	1.00	1.08
		σ_{12}	8.61	4.55	5.32	8.36	11.54
	OLS	β_0	-8.36	1.36	-9.20	-8.34	-7.48
		β_1	0.88	0.12	0.80	0.88	0.95
0.8	Heckman	β_0	-9.91	1.59	-10.91	-9.87	-8.92
		β_1	0.99	0.13	0.90	0.99	1.08
		σ_{12}	9.85	3.91	7.10	9.53	12.27
	OLS	β_0	-8.08	1.32	-8.91	-8.12	-7.20
		β_1	0.86	0.12	0.78	0.86	0.93
1.0	Heckman	β_0	-9.96	1.71	-11.13	-10.02	-8.86
		β_1	1.00	0.14	0.90	1.00	1.09
		σ_{12}	10.11	3.96	7.24	9.66	12.81
	OLS	β_0	-8.06	1.39	-9.04	-8.05	-7.16
		β_1	0.86	0.12	0.78	0.86	0.93

Table 3
Simulation Results for Design III
$\beta_0 = -10,\ \beta_1 = 1,\ \sigma_{12} = 10\rho$

ρ	Estimation Method	Parameter	Mean	Standard Deviation	First Quartile	Median	Third Quartile
0	Heckman	β_0	-9.97	6.77	-14.66	-10.10	-5.04
		β_1	1.00	0.42	0.71	1.00	1.28
		σ_{12}	-0.24	4.36	-2.92	-0.38	2.68
	OLS	β_0	-10.17	5.42	-14.08	-10.16	-6.73
		β_1	1.01	0.35	0.78	1.00	1.26
0.2	Heckman	β_0	-10.14	6.53	-14.22	-10.05	-6.03
		β_1	1.01	0.42	0.74	0.99	1.27
		σ_{12}	-382.56	8609.55	-0.22	2.39	5.21
	OLS	β_0	-7.81	5.41	-11.61	-7.97	-4.23
		β_1	0.87	0.35	0.62	0.89	1.11
0.4	Heckman	β_0	-10.09	7.09	-14.88	-9.81	-5.23
		β_1	1.01	0.45	0.68	0.99	1.30
		σ_{12}	5.54	4.01	2.89	5.71	8.13
	OLS	β_0	-4.84	5.71	-9.03	-4.89	-0.89
		β_1	0.69	0.38	0.43	0.69	0.97
0.6	Heckman	β_0	-10.01	6.75	-14.48	-10.05	-5.55
		β_1	1.00	0.42	0.72	0.99	1.28
		σ_{12}	8.34	3.80	6.02	8.46	10.62
	OLS	β_0	-2.15	5.18	-5.60	-1.95	1.36
		β_1	0.53	0.34	0.30	0.52	0.76
0.8	Heckman	β_0	-9.95	6.79	-14.14	-9.90	-5.22
		β_1	1.00	0.43	0.70	0.99	1.28
		σ_{12}	-1646.38	37027.74	7.01	9.51	11.94
	OLS	β_0	-0.81	4.96	-4.13	-0.62	2.36
		β_1	0.45	0.33	0.24	0.44	0.67
1.0	Heckman	β_0	-9.84	6.42	-13.94	-9.95	-5.56
		β_1	0.99	0.41	0.71	0.97	1.26
		σ_{12}	9.78	3.60	7.51	9.69	12.09
	OLS	β_0	-0.62	4.91	-3.72	-0.39	2.64
		β_1	0.44	0.32	0.23	0.44	0.63

Table 4
Summary Statistics of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ for $d_i = 1$

Design	ρ	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
I	0	0.2624	0.5273	-0.5060	2.1872	-0.9471	1.0112
	0.2	0.2653	0.5289	-0.5062	2.1915	-0.9513	1.0170
	0.4	0.2543	0.5260	-0.5053	2.1865	-0.9513	1.0006
	0.6	0.2612	0.5307	-0.5131	2.2053	-0.9589	1.0135
	0.8	0.2584	0.5296	-0.5087	2.1938	-0.9573	1.0103
	1.0	0.2575	0.5322	-0.5146	2.2023	-0.9690	1.0109
II	0	11.5651	6.6512	-0.0440	1.7975	-0.7020	22.8881
	0.2	11.2934	6.4926	-0.0445	1.7980	-0.6877	22.3453
	0.4	11.7758	6.7729	-0.0437	1.7968	-0.6916	23.3072
	0.6	11.0624	6.3558	-0.0448	1.7983	-0.6557	21.8809
	0.8	11.4384	6.5801	-0.0442	1.7977	-0.6988	22.6398
	1.0	11.2061	6.4422	-0.0448	1.7984	-0.6954	22.1717
III	0	6.1994	3.4584	-0.2153	2.0345	-1.2049	12.0147
	0.2	6.8139	3.7753	-0.2136	2.0293	-1.1748	13.1726
	0.4	6.2067	3.4613	-0.2130	2.0298	-1.1997	12.0277
	0.6	6.2037	3.4637	-0.2131	2.0282	-1.1749	12.0276
	0.8	6.4328	3.5816	-0.2161	2.0369	-1.1892	12.4609
	1.0	6.1974	3.4565	-0.2127	2.0294	-1.1965	12.0108

Table 5
Summary Statistics of Condition Number

Design	ρ	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
I	0	103.25	30.87	1.74	10.29	48.72	330.06
	0.2	101.79	27.94	1.40	8.18	48.72	284.24
	0.4	105.58	35.15	1.91	9.76	49.71	330.06
	0.6	101.48	28.01	1.42	7.97	48.72	272.80
	0.8	102.96	29.97	1.27	6.34	49.75	284.24
	1.0	102.39	30.67	1.74	8.46	49.71	284.24
II	0	4.79	0.26	0.68	4.34	4.09	6.08
	0.2	4.81	0.28	0.66	3.68	4.18	6.08
	0.4	4.79	0.27	0.34	3.07	4.09	5.74
	0.6	4.83	0.28	0.70	3.91	4.19	6.08
	0.8	4.80	0.26	0.57	4.07	4.18	5.86
	1.0	4.82	0.27	0.45	2.93	4.19	5.74
III	0	13.69	0.92	0.47	3.09	11.32	16.62
	0.2	13.61	0.90	0.37	2.88	11.39	16.58
	0.4	13.63	0.91	0.44	2.98	11.51	17.01
	0.6	13.64	0.92	0.37	2.88	11.22	16.67
	0.8	13.68	0.93	0.30	2.96	11.56	17.01
	1.0	13.63	0.90	0.32	2.67	11.56	16.51

Table 6
Condition Number by Observations and Sample Size
Design I, $\rho = 1$

Sample Size (N)	Observations	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
500	uncensored	97.23	18.36	0.92	4.61	57.12	175.44
	all	110.17	18.74	0.84	4.62	66.59	190.42
1000	uncensored	97.62	12.64	0.66	4.33	67.12	165.11
	all	110.79	12.96	0.53	4.07	79.61	177.26
2000	uncensored	97.52	8.65	0.55	3.94	74.39	135.69
	all	109.14	8.80	0.48	3.69	84.87	145.61
3000	uncensored	96.86	7.21	0.36	3.78	78.16	124.83
	all	108.92	7.40	0.35	3.72	89.32	137.00
4000	uncensored	97.46	6.03	0.28	2.99	83.22	127.72
	all	109.76	6.23	0.24	2.85	95.56	128.57
5000	uncensored	96.37	5.48	0.19	2.74	81.17	115.20
	all	109.02	5.65	0.13	2.65	94.98	128.06
10000	uncensored	95.77	3.74	0.24	3.09	86.28	109.02
	all	108.53	3.81	0.21	3.12	97.58	121.29

Table 7
Condition Number by Observations
$\alpha_1 = 0.1, \rho = 1$

α_0	Observations	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
-0.2	uncensored	45.04	14.92	1.22	4.96	19.62	107.62
	all	46.93	15.08	1.19	4.80	20.38	109.38
-0.5	uncensored	59.78	17.96	1.29	6.01	28.83	163.49
	all	63.92	18.37	1.21	5.49	31.98	163.80
-1.0	uncensored	102.39	30.67	1.74	8.46	49.71	284.24
	all	115.78	31.01	1.65	8.56	54.96	314.60
-1.5	uncensored	184.99	58.01	2.78	20.28	93.36	725.60
	all	223.50	56.93	2.41	17.56	126.38	736.17

Table 8
Summary Statistics of $\hat{\alpha}_0 + \hat{\alpha}_1 w_i$ by Observations
$\alpha_1 = 0.1, \rho = 1$

α_0	Observations	$\hat{\alpha}_0 + \hat{\alpha}_1 w_i$				Proportion of cases in which $\hat{\alpha}_0 + \hat{\alpha}_1 w_i > 0$		
		Mean	Standard Deviation	Minimum	Maximum	Mean	Minimum	Maximum
-0.2	uncensored	0.94	0.56	-0.19	1.82	0.93	0.85	1.00
	censored	0.41	0.47	-0.19	1.62	0.77	0.42	1.00
-0.5	uncensored	0.68	0.55	-0.48	1.52	0.86	0.74	1.00
	censored	0.16	0.50	-0.50	1.39	0.57	0.34	1.00
-1.0	uncensored	0.26	0.53	-0.97	1.01	0.67	0.47	0.81
	censored	-0.25	0.53	-1.02	0.95	0.30	0.19	0.51
-1.5	uncensored	-0.15	0.50	-1.41	0.50	0.45	0.00	0.69
	censored	-0.67	0.55	-1.52	0.48	0.15	0.00	0.25

References

Amemiya, Takeshi. Advanced Econometrics. Cambridge, MA: Harvard University Press, 1985.

Barnow, Burt; Cain, Glen; and Goldberger, Arthur. "Issues in the Analysis of Selectivity Bias." In Evaluation Studies Review Annual, Vol. 5, edited by E. Stromsdorfer and G. Farkas. Beverly Hills, CA: Sage, 1981.

Belsley, David A.; Kuh, Edwin; and Welsch, Roy E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York, NY: John Wiley, 1980.

Bockstael, Nancy; Strand, Jr., Ivar; McConnell, Kenneth; and Arsanjani, Firuzeh. "Sample Selection Bias in the Estimation of Recreation Demand Functions: An Application to Sportfishing." Land Economics 66 (1990): 40-49.

Davidson, Russell, and MacKinnon, James G. Estimation and Inference in Econometrics. Oxford: Oxford University Press, 1993.

Greene, William. LIMDEP 6.0: User's Manual and Reference Guide. New York, NY: Econometric Software, 1992.

Greene, William. Econometric Analysis. New York, NY: Macmillan, 1993.

Heckman, James. "Sample Selection Bias as a Specification Error." Econometrica 47 (1979), 153-161.

Heckman, James. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and Social Measurement 5 (1976): 475-592.

Leung, Siu Fai, and Yu, Shihti. "On the Choice Between Sample Selection and Two-part Models." Journal of Econometrics 72 (1996): 197-229.

Little, Roderick. "A Note about Models for Selectivity Bias." Econometrica 53 (1985): 1469-1474.

Little, Roderick, and Rubin, Donald. Statistical Analysis and Missing Data. New York, NY: John Wiley, 1987.

Maddala, G. S. Limited-dependent and Qualitative Variables in Econometrics. New York, NY: Cambridge University Press, 1983.

Manning, W.; Duan, N.; and Rogers, W. "Monte Carlo Evidence on the Choice between Sample Selection and Two-part Models." Journal of Econometrics 35 (1987): 59-82.

Nawata, Kazumitsu. "A Note On the Estimation of Models with Sample-selection Biases." Economics Letters 42 (1993): 15-24.

Nawata, Kazumitsu. "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-step Estimator." Economics Letters 45 (1994): 33-40.

Nelson, Forrest. "Efficiency of the Two-step Estimator for Models with Endogenous Sample Selection." Journal of Econometrics 24 (1984): 181-196.

Paarsch, Harry. "A Monte Carlo Comparison of Estimators for Censored Regression Models." Journal of Econometrics 24 (1984): 197-213.

Schmertmann, Carl P. "Selectivity Bias Correction Methods in Polychotomous Sample Selection Models." Journal of Econometrics 60 (1994): 101-132.

Stapleton, David, and Young, Douglas. "Censored Normal Regression with Measurement Error on the Dependent Variable." Econometrica 52 (1984): 737-760.

Wales, T. J., and Woodland, A. D. "Sample Selectivity and the Estimation of Labor Supply Functions." International Economic Review 21 (1980): 437-468.