Misclassification in Binary Choice Models with Sample Selection

Maria Felice Arezzo* Giuseppina Guagnano †

Abstract

In parametric models, there are several reasons why the estimators can be biased and inconsistent. We focus on two sources of bias, namely measurement error in the dependent variable and unrepresentative samples, and we propose an estimation method that simultaneously corrects for this double source of bias. Most empirical work is based on observational data that are unrepresentative of the population of interest. Sample selection models attempt to correct for non-randomly selected data in a two-model hierarchy where, on the first level, a binary selection equation determines if a particular observation will be available for the second level (outcome equation). In the case of binary choice models, we assume that also the dependent variable of the outcome equation is binary. The likelihood function takes into account the selection mechanism and allows for unbiased parameters estimation. We extend this framework to the situation of a measurement error in the dependent variable of the outcome equation. We use a parametric approach to the estimation of the probabilities of misclassification by incorporating them in the likelihood of a binary choice model with sample selection.

Key Words: Misclassified dependent variable, Sample selection bias

1. Introduction

In parametric models, there are several reasons why the estimators can be biased and inconsistent. A long lasting stream of literature has focused in reducing (or eliminating) the bias of estimators, in order to alleviate the problems that it can cause in inference. In this work, we focus on two sources of bias, namely measurement error in the dependent variable and unrepresentative samples. We propose an estimation method that simultaneously corrects for this double source of bias.

Most empirical work in the social sciences is based on observational data that are incomplete and therefore unrepresentative of the population of interest. There are many types of selection mechanisms that result in a non-random sample. Some of them are due to sample design, while others depend on the behavior of the units being sampled, other than non-response or attrition. In the first case, data are usually missing on all the variables of interest; for example, in estimating a saving function for all the families of a given country, a bias would arise if only families whose household head shows certain characteristics were sampled. However, when causes of missingness are appropriately exogenous, using a sub-sample has no serious consequences.

In the second case, instead, there is a self-selection of the sample units and data availability on a key variable depends on the behavior of the units about another variable. The classical example is that of the linear wage equation where we want to estimate the expected wage of an individual using a set of exogenous characteristics (gender, age, education etc). The key problem is that in regressing wages on the characteristics of *employed* individuals, we are not making inferences for the population as a whole. In fact those in employment are a selected sample of the population and their wages are higher than those not in the labor force would have. Hence the results will tend to be biased (sample selection bias). In

*Sapienza University of Rome, Via del Castro 9, 00161 Rome
†Sapienza University of Rome, Via del Castro 9, 00161 Rome

order to avoid the bias, we need to take into account the selection mechanism by which an individual decides to take a job and then receives a wage.

As it is well known, Heckman (1979) proposed a useful framework for handling estimation when the sample is subject to a selection mechanism. In the original framework, the dependent variable in the outcome equation (the wage equation in the above example) is continuous and can be explained by a linear regression model with a normal random component; in addition to the output equation a selection equation describes the selection rule by means of a binary choice model (probit).

The original Heckman's model was extended in many directions and a survey would be beyond the scope of this paper, but the interested reader can refer to Vella (1998) and Lee (2003). To our purposes the relevant framework is the one where both the output and the selection equations are defined as a binary choice model (Dubin and Rivers, 1989).

Sample selection models attempt to correct for non-randomly selected data in a two-model hierarchy where, on the first level, a binary selection equation determines whether a particular observation will be available for the second level (outcome equation). In the case of binary choice models, we assume that also the dependent variable of the outcome equation is binary. The likelihood function takes into account the selection mechanism and allows for unbiased estimates of the parameters of interest (i.e. the coefficients of the selection and the outcome equations and the correlation coefficient of the two processes).

We extend this framework to the situation of a measurement error in the dependent (binary) variable of the outcome equation. Misclassification of a binary variable means that an observation with a true value of 0 is observed as 1 or an observation that is truly a 1 is observed as a 0. This mistake are very common in applied context. It could easily happen, for example, during an interview if the respondent misunderstands the question or the interviewer simply checks the wrong box.

When traditional estimation methods (like logit or probit) are used in binary choice models with a misclassified dependent variable, the resulting estimates are inconsistent.

Previous work on misclassified dependent variables in discrete choice models follows two approaches. In the first, supplemental data are used to verify the accuracy of responses. In Chua and Fuller (1987) a parametric model that incorporates all possible $J(J-1)$ misclassification of a J-level outcome variable is developed. This approach has been seldom used because it is very data demanding, as a minimum of three independent sets of survey responses obtained by re-interviewing the original respondents are required. A similar approach, based on a conditional logit procedure, was proposed in Poterba and Summers (1995). It also incorporates all possible misclassification and the estimation of the misclassification probabilities is done by analyzing the divergences between interview and reinterview outcomes.

Other authors have taken a different path to deal with misclassification. In particular Hausman et al. (1998) and Abrevaya and Hausman (1999) incorporates the probability of misclassification directly into the estimation procedure. In particular, they consider a parametric model for a binary response variable with two types of misclassification; these unknown misclassification probabilities are estimated parametrically simultaneously with the usual coefficients of the binary choice model.

We extend the work of Hausman (1998) on misclassified dependent variable in binary choice models incorporating the other source of bias coming from sample selection. We use a parametric approach to simultaneously estimate the parameters of the selection and of the outcome equation, the correlation between them and the probabilities of misclassification.

2. The model

Let's first introduce some notations and briefly illustrate the sample selection framework with a binary choice model for both the selection and the output equations (Dubin and Rivers, 1989).

Let Y^* and S^* be two latent (unobservable) variables characterizing the output and the selection equations respectively. The model, in its general form, is:

$$Y_i^* = \boldsymbol{X_{1i}}\beta + \epsilon_{1i} \tag{1a}$$
$$S_i^* = \boldsymbol{X_{2i}}\theta + \epsilon_{2i} \tag{1b}$$

where $\boldsymbol{X_i} = (\boldsymbol{X_{1i}}, \boldsymbol{X_{2i}})$ is a vector of exogenous variables (namely, $\boldsymbol{X_{1i}}$ for Y_i^* and $\boldsymbol{X_{2i}}$ for S_i^*), containing all the relevant covariates, and β and θ are the vectors of regression coefficients. Note that, for model (1a-1b) to be identified, it is necessary that $\boldsymbol{X_{1i}}$ and $\boldsymbol{X_{2i}}$ do not fully overlap; that is the covariates of the selection and outcome equations must differ for at least one variable. Let's define Y_i and S_i as two observable variables such that:

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2} \qquad S_i = \begin{cases} 1 & \text{if } S_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The p.d.f. of Y_i and S_i are Bernoulli, with probability of success depending on the parameters β and θ respectively.

Model (3) defines the mechanism which governs the censoring process: we can observe Y_i if and only if $S_i = 1$. On the contrary, if $S_i = 0$, Y_i will be missing.

In the general case, if we estimated the parameters of equation (1a) without considering the selection process (1b), a bias would arise. This is because the processes represented by the two equations are related, i.e. $corr(\epsilon_1, \epsilon_2) = \rho$ is not null (see for example Cameron and Trivedi, 2005, for further details).

The likelihood function for model (1a-1b) is:

$$\begin{aligned} L(\eta) &= \prod_{i=1}^{n} \Big[Pr(S_i^* < 0) \Big]^{1-S_i} \cdot \Big[Pr(Y_i = y_i | S_i^* > 0) \cdot Pr(S_i^* > 0) \Big]^{S_i} = \\ &= \prod_{i=1}^{n} \Big[1 - {}_S\pi(\boldsymbol{X_i}) \Big]^{1-S_i} \cdot \Big[Pr(Y_i = y_i | S_i = 1) \cdot {}_S\pi(\boldsymbol{X_i}) \Big]^{S_i} \end{aligned} \tag{4}$$

where $\eta = (\beta, \theta, \rho)$ is the vector of parameters to be estimated, $y_i = 0, 1$ and the function ${}_S\pi(\cdot)$ gives the probability that an observation is uncensored.

Now, if we assume that:

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim NID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right] \tag{5}$$

we can compute the probabilities $P(S_i = 0) = 1 - {}_S\pi(\boldsymbol{X_2})$ and $P(Y_i = i) = {}_Y\pi(\boldsymbol{X_1})$ in (4) as follows:

$$Pr(S_i = 0) = 1 - {}_S\pi(\boldsymbol{X_{2i}}) = \Phi\left(-\theta' \boldsymbol{X_{2i}}\right) \tag{6}$$

$$\begin{aligned} Pr(Y_i = 1, S_i = 1) &= Pr(Y_i = 1 | S_i = 1) \cdot Pr(S_i = 1) = {}_Y\pi(\boldsymbol{X_{1i}}) \cdot {}_S\pi(\boldsymbol{X_{2i}}) \\ &= \Phi_2\left(\beta' \boldsymbol{X_{1i}}, \theta' \boldsymbol{X_{2i}}, \rho\right) \end{aligned} \tag{7}$$

$$\begin{aligned} Pr(Y_i = 0, S_i = 1) &= Pr(Y_i = 0 | S_i = 1) \cdot Pr(S_i = 1) = (1 - {}_Y\pi(\boldsymbol{X_{1i}})) \cdot {}_S\pi(\boldsymbol{X_2}) \\ &= \Phi_2\left(\beta' \boldsymbol{X_{1i}}, -\theta' \boldsymbol{X_{2i}}, -\rho\right) \end{aligned} \tag{8}$$

where Φ and Φ_2 are c.d.f. of univariate and bivariate normal respectively.

Now, let us suppose that Y_i can be misclassified, that is some true 1 are observed as 0, and some true 0 are observed as 1. It follows that what we observe can differ from the true response variable of the outcome equation. Let's denote ${}_{obs}Y_i$ the observed binary variable affected by error, and ${}_{T}Y_i$ the true response variable of equation 2. Following Hausman et al. (1998), we assume that the probability of misclassification depends on the value of ${}_{T}Y_i$, but is otherwise independent of the covariates $\boldsymbol{X_1}$. To be more specific, we set the following misclassification probabilities:

$$\alpha_0 = \Pr\left({}_{obs}Y_i = 1 | {}_{T}Y_i = 0\right) \tag{9}$$
$$\alpha_1 = \Pr\left({}_{obs}Y_i = 0 | {}_{T}Y_i = 1\right) \tag{10}$$

The probability that a true zero is misclassified as a one is given by α_0; the probability that a true one is misclassified as a zero is given by α_1. The stochastic mechanism that determines the values of the observed dependent variable ${}_{obs}Y$ becomes:

$$\begin{aligned}
&\Pr\left({}_{obs}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right) = \\
&\Pr\left({}_{obs}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}, {}_{T}Y_i = 1\right) \Pr\left({}_{T}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right) + \\
&+ \Pr\left({}_{obs}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}, {}_{T}Y_i = 0\right) \Pr\left({}_{T}Y_i = 0 | S_i = 1, \boldsymbol{X_{1i}}\right) \\
&= (1 - \alpha_0 - \alpha_1) \Pr\left({}_{T}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right) + \alpha_0
\end{aligned} \tag{11}$$

where $\Pr\left({}_{T}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right) = {}_{TY}\pi(\boldsymbol{X_{1i}})$ is the homologous of ${}_{Y}\pi(\boldsymbol{X_{1i}})$ in equations (7-8).

Obviously we can put:

$$\begin{aligned}
&\Pr\left({}_{obs}Y_i = 0 | S_i = 1, \boldsymbol{X_{1i}}\right) = 1 - \Pr\left({}_{obs}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right) \\
&= 1 - \alpha_0 - (1 - \alpha_0 - \alpha_1) \Pr\left({}_{T}Y_i = 1 | S_i = 1, \boldsymbol{X_{1i}}\right)
\end{aligned} \tag{12}$$

To estimate the vector of parameters, $\gamma = (\theta, \beta, \alpha_0, \alpha_1, \rho)$, we have to extend the likelihood function (4) bearing in mind that the *observed* values of the dependent variable in the outcome equation are misclassified. Rewriting the likelihood function by plugging in (11) and (12) into (4) and considering assumption (5), we get the following log-likelihood function:

$$\begin{aligned}
\log L(\gamma) = \sum_{i=1}^{n} &(1 - S_i) \cdot \log \Phi\left(-\theta' \boldsymbol{X_{2i}}\right) + \\
&S_i \cdot \log \left[\alpha_0 \Phi\left(\theta' \boldsymbol{X_{2i}}\right) + (1 - \alpha_0 - \alpha_1) \Phi_2\left(\beta' \boldsymbol{X_{1i}}, \theta' \boldsymbol{X_{2i}}, \rho\right) \right]^{{}_{obs}Y_i} + \\
&S_i \cdot \log \left[(1 - \alpha_0) \Phi\left(\theta' \boldsymbol{X_{2i}}\right) - (1 - \alpha_0 - \alpha_1) \Phi_2\left(\beta' \boldsymbol{X_{1i}}, \theta' \boldsymbol{X_{2i}}, \rho\right) \right]^{1 - {}_{obs}Y_i}
\end{aligned}$$

3. Simulation results

In this section, we present Monte Carlo simulations performed to evaluate finite sample performance of the proposed model. We consider the following generating model:

$$Y_i^* = -1 + 0.2X_{11i} + 1.5X_{12i} - 0.6X_{13i} + \epsilon_{1i}$$
$$S_i^* = 0.5 + 0.8X_{21i} - 0.5X_{22i} + \epsilon_{2i}$$

The outcome equation is the same as Hausman et al. (1998), where X_{11} is drawn from a lognormal, X_{12} is a dummy variable equal to one with probability $1/3$ and X_{13} is a uniform $(0, 1)$. For the selection equation, we drawn both X_{21} and X_{22} from a standard normal distribution. The choice of $\theta_0 = 0.5$ is to ensure a medium amount of censored data (approximately 30%).

We performed 200 replications with samples of size $n = 5,000$. We choose $\rho \in \{-0.8, -0.4; -0.2; 0; 0.2; 0.4; 0.8\}$ and the following pairs of misclassification probabilities: $(\alpha_0 = 0.02, \alpha_1 = 0.02)$, $(\alpha_0 = 0.05, \alpha_1 = 0.05)$, $(\alpha_0 = 0.05, \alpha_1 = 0.2)$ and $(\alpha_0 = 0.2, \alpha_1 = 0.02)$.

We compared 4 models: the simple probit, a model that corrects for sample selection only (named SS in the following), a model that corrects for misclassification only (MIS) and a model that corrects for sample selection *and* misclassification (MIS-SS).

The results, reported in tables 1 and 2[1], clearly show that the MIS-SS dominates all others.

With regards to probit estimates, consistently with Hausman et al. (1998), we found biased estimates. Depending on the coefficient, the bias spans from 6 to 30% even with a small amount of misclassification. In our simulations, unlike theirs, we don't find the problem to worsen as the α's increase.

When correcting for misclassification (MIS), the bias of $\hat{\beta}$s considerably reduces (around 5%) only if ρ is moderate. However, as expected, when the correlation between the outcome and the selection equation errors increases, the bias rises to 15-20%.

Correcting for sample selection (SS) induces the bias under 10%, no matter the value of ρ. As anticipated, the best results are obtained when correcting for both sample selection and misclassification

4. Conclusions

We proposed an estimation method of the parameters of a binary response model affected by a censoring mechanism, which makes unobservable a relevant part of units, and by a measurement error in the dependent variable.

We derived a likelihood function analytically, obtaining the maximum likelihood estimator for the parameter vector. The results obtained in a simulation study are very promising: actually our estimator considerably reduces the bias compared to alternative estimators that do not simultaneously account for the two sources of bias (probit, Hausman and sample selection).

Further work will be oriented to allow the probability of misclassification to vary according to some known covariates.

[1]To spare space, we provide only partial results but the interest reader can obtain them all by writing to the authors

Table 1: Monte Carlo simulation results ($n = 5,000$, $\alpha_0 = \alpha_1 = 0.02$, $\rho \in \{\pm 0.2; \pm 0.8\}$)

		Probit								MIS							
		ρ								ρ							
		-0.8		-0.2		0.2		0.8		-0.8		-0.2		0.2		0.8	
		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
α_0	0.02									0.0436	0.0146	0.0596	0.0274	0.0616	0.0332	0.0748	0.0482
α_1	0.02									0.0484	0.0266	0.0496	0.0266	0.0524	0.0272	0.0496	0.0212
β_0	-1	-0.6335	0.0560	-0.6335	0.0560	-0.6335	0.0560	-0.6335	0.0560	-1.5144	0.1460	-1.1477	0.1542	-0.9734	0.1589	-0.8110	0.1694
β_1	0.2	0.1729	0.0152	0.1729	0.0152	0.1729	0.0152	0.1729	0.0152	0.2290	0.0284	0.2099	0.0299	0.2112	0.0415	0.2365	0.0402
β_2	1.5	1.4259	0.0533	1.4259	0.0533	1.4259	0.0533	1.4259	0.0533	1.7109	0.1523	1.5612	0.1588	1.5589	0.1696	1.7318	0.1917
β_3	-0.6	-0.5284	0.0855	-0.5284	0.0855	-0.5284	0.0855	-0.5284	0.0855	-0.6729	0.1329	-0.6141	0.1315	-0.6154	0.1441	-0.6901	0.1546

		SS								MIS-SS							
		ρ								ρ							
		-0.8		-0.2		0.2		0.8		-0.8		-0.2		0.2		0.8	
		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
α_0	0.02									0.0215	0.0086	0.0299	0.0212	0.0311	0.0279	0.0134	0.0136
α_1	0.02									0.0455	0.0329	0.0439	0.0328	0.0360	0.0307	0.0114	0.0088
ρ		-0.8300	0.0822	-0.2046	0.0751	0.1659	0.0768	0.8577	0.0950	-0.7927	0.1674	-0.2226	0.1002	0.2131	0.1014	0.7708	0.1362
θ_0	0.5	0.4968	0.0214	0.5004	0.0216	0.4991	0.0217	0.4952	0.0212	0.5017	0.0308	0.5018	0.0299	0.5025	0.0282	0.5016	0.0215
θ_1	0.8	0.7967	0.0250	0.7987	0.0255	0.8008	0.0255	0.7975	0.0248	0.8062	0.0345	0.8035	0.0333	0.8041	0.0321	0.8025	0.0251
θ_2	-0.5	-0.5027	0.0220	-0.5062	0.0227	-0.5042	0.0227	-0.5088	0.0218	-0.5029	0.0314	-0.5034	0.0306	-0.5027	0.0292	-0.5033	0.0220
β_0	-1	-0.9852	0.0690	-0.9253	0.0680	-0.9120	0.0637	-0.9137	0.0532	-1.0237	0.1003	-1.0397	0.1385	-1.0432	0.1461	-0.9854	0.0725
β_1	0.2	0.1876	0.0139	0.1814	0.0146	0.1833	0.0148	0.1851	0.0144	0.2071	0.0288	0.2122	0.0346	0.2104	0.0341	0.1954	0.0176
β_2	1.5	1.4131	0.0551	1.3827	0.0529	1.3991	0.0527	1.4306	0.0529	1.5395	0.0954	1.5767	0.1419	1.5608	0.1443	1.4748	0.0715
β_3	-0.6	-0.5490	0.0874	-0.5380	0.0885	-0.5771	0.0870	-0.5853	0.0787	-0.5745	0.1150	-0.6110	0.1344	-0.6168	0.1322	-0.5772	0.0890

Table 2: Monte Carlo simulation results ($n = 5,000$, $\alpha_0 = 0.05; \alpha_1 = 0.2$, $\rho \in \{\pm 0.2; \pm 0.8\}$)

		Probit								MIS							
		ρ								ρ							
		-0.8		-0.2		0.2		0.8		-0.8		-0.2		0.2		0.8	
		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
α_0	0.05									0.0936	0.0222	0.1002	0.0371	0.1028	0.0465	0.1053	0.0829
α_1	0.2									0.2559	0.0820	0.2624	0.0696	0.2705	0.0653	0.2723	0.0172
β_0	-1	-0.6335	0.0560	-0.6335	0.0560	-0.6335	0.0560	-0.6335	0.0560	-1.5166	0.2317	-1.1438	0.2305	-0.9747	0.2323	-0.7965	0.2804
β_1	0.2	0.1729	0.0152	0.1729	0.0152	0.1729	0.0152	0.1729	0.0152	0.2287	0.0507	0.2106	0.0516	0.2148	0.0544	0.2352	0.0494
β_2	1.5	1.4259	0.0533	1.4259	0.0533	1.4259	0.0533	1.4259	0.0533	1.7111	0.2912	1.5689	0.2961	1.5874	0.3126	1.7236	0.2754
β_3	-0.6	-0.5284	0.0855	-0.5284	0.0855	-0.5284	0.0855	-0.5284	0.0855	-0.6707	0.2013	-0.6210	0.2071	-0.6415	0.2200	-0.6842	0.1975

		SS								MIS-SS							
		ρ								ρ							
		-0.8		-0.2		0.2		0.8		-0.8		-0.2		0.2		0.8	
		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
α_0	0.05									0.0428	0.0111	0.0469	0.0259	0.0363	0.0331	0.0373	0.0173
α_1	0.2									0.1950	0.0594	0.2293	0.0745	0.1726	0.0636	0.1731	0.0376
ρ		-0.6320	0.0752	-0.1261	0.0732	0.1106	0.0738	0.5551	0.0798	-0.7287	0.2886	-0.2359	1.1019	0.1758	0.1200	0.7140	0.2725
θ_0	0.5	0.4979	0.0213	0.5019	0.0217	0.4987	0.0216	0.4987	0.0215	0.5014	0.0551	0.5014	0.0365	0.5024	0.0217	0.5015	0.0939
θ_1	0.8	0.7952	0.0251	0.8031	0.0255	0.7974	0.0254	0.8008	0.0252	0.8048	0.0568	0.8031	0.0401	0.8029	0.0255	0.8019	0.0913
θ_2	-0.5	-0.4997	0.0221	-0.5040	0.0227	-0.5055	0.0227	-0.5066	0.0223	-0.5019	0.0540	-0.5024	0.0374	-0.5030	0.0227	-0.5037	0.0891
β_0	-1	-0.9235	0.0671	-0.9392	0.0659	-0.8862	0.0624	-0.8653	0.0550	-1.0351	0.1631	-0.9610	0.1979	-0.9673	0.1870	-0.9702	0.1710
β_1	0.2	0.1238	0.0118	0.1187	0.0122	0.1168	0.0123	0.1112	0.0118	0.1979	0.0598	0.2088	0.0586	0.1889	0.0429	0.1925	0.0911
β_2	1.5	1.0940	0.0518	1.0850	0.0510	1.0725	0.0503	1.0763	0.0492	1.4925	0.1698	1.5749	0.2556	1.4777	0.2330	1.4779	0.1957
β_3	-0.6	-0.4103	0.0856	-0.3632	0.0866	-0.4302	0.0849	-0.4477	0.0797	-0.5269	0.1715	-0.6077	0.2069	-0.5616	0.1800	-0.5737	0.1887

REFERENCES

Abrevaya J. and Hausman J.A. (1999), "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Panel Data on Employment Spells", *Annales D'Economie et de Statistique*, 56, 243–75.

Cameron, A.C. and Trivedi, P.K. (2005), *Microeconometrics: Methods and Applications* , New York, USA: Cambridge University Press.

Chua, T.C. and Fuller, W.A. (1987), "A model for multinomial response error applied to labor flows", *Journal of the American Statistical Association*, 82, 46–51.

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

Hausman, J.A., Jason Abrevaya, J. and Scott-Morton, F.M. (1998), Misclassification of the dependent variable in a discrete-response setting, *Journal of Econometrics*, 87, 239–269.

Heckman, J.J. (1979), "Sample selection bias as a specification error," *Econometrica*, 47, 153–162.

Lee, L.F. (2003), "Self-selection", in *A Companion to Theoretical Econometrics*, eds. B. H. Baltagi, Malden, USA: Blackwell Publishing, pp. 383-409.

Poterba, J.M. and Summers, L.H. (1995), "Unemployment benefits and labor market transitions: a multinomial logit model with errors in classification", *Review of Economics and Statistics*, 77, 207–2016.

Vella, F. (1998), "Estimating models with sample selection bias: a survey", *The Journal of Human Resources*, 33, 127–169.